MASTER

Improving an External Last Mile Delivery Process Using Carrier Event Log Data

Ploegmakers, D.J.Z.

Award date:
2020

November 2020

Improving an External Last Mile Delivery Process Using Carrier Event Log Data

Master Thesis by

Dalí Ploegmakers
Student ID: 1263390

Supervisor Eindhoven University of Technology

Dr S.E.C. (Sarah) Gelper

Second Supervisor Eindhoven University of Technology

Dr N.R. (Nevin) Mutlu

Supervisor Hilti AG

J. (Johan) De Smedt

Department of Industrial Engineering and Innovation Sciences
Sub-department of Innovation Technology Entrepreneurship & Marketing
Eindhoven University of Technology

TU/e School of Industrial Engineering

Series Master Thesis Innovation Management

Keywords: Process Mining, Outsourced last mile delivery, B2B, Discovery, Conformance checking, Enhancement, Prediction, Bottleneck analysis, event log data

Preface

This Master Thesis project was started during my internship at Hilti AG in Liechtenstein and was completed in the months following this internship. I want to thank my Hilti colleagues for their openness to new ideas and welcoming me to the Hilti family from day one. In particular, I would like to thank my supervisor Johan De Smedt for all his flexibility, his genuine interest in pushing the limits further and gently pushing me to showcase the value I created.

Furthermore, I would like to thank my mentor Sarah Gelper for her continuous feedback and thinking with me to create a coherent story that gets the message across. I would also like to thank my second supervisor Nevin Mutlu for her flexibility and providing me with feedback from a new angle.

One of my goals going into the project was learning how to code and knowing how things are done in a professional setting. I am happy to say that I learned loads on both these aspects.

This thesis project has been a long journey with many ups and downs. A special thanks to my friends Rijk, Julia, Lucas and Klaas, and the support of my family, celebrating the ups and supporting me at rough times. A special thanks to Babette and Tajana, without whom this thesis writing process would not have been the same.

Abstract

As an increasing percentage of shippers outsource their last mile delivery process, more shippers see a part of their overall service quality depend on a process that is largely outside of their control. This thesis reviews if shippers can improve the external last mile delivery process using carrier event log data. The following main Research Question is addressed; “How can carrier event log data be used to improve an external last mile delivery process?”

To answer the main Research Questions, a shipper’s last mile management team was interviewed to find Opportunities to improve the external last mile. Successively, a literature review was performed to determine which insights could be gained by analysing last mile event logs using Process Mining. Finally, experiments with a shipper´s dataset were conducted to see how Process Mining methods could be applied to the transformed event log and realize the Opportunities.

Three Process Mining methods; Discovery, Conformance checking and Enhancement, were found to be able to understand, compare and improve processes using event logs respectively. Six Opportunities were found to improve the external last mile. Of these, three Opportunities; improving data quality, creating an overview of (live) deliveries and increasing transparency of the carrier’s process, are based on Discovery and Conformance checking and were shown to be achievable in the experiments. These three Opportunities plus a notification system, mainly support the shipper in their current responsibilities. The notification system, bottleneck analysis and optimised parcel loading are Opportunities based on Enhancement which, as far as the researcher is aware, have not yet been used to improve an external last mile delivery process. The notification system may directly enable shippers to have a direct impact on service quality by informing customers of late deliveries, where bottleneck analysis and optimised parcel loading might even directly raise on-time delivery rates. Although these Opportunities have yet to be executed, it was seen how the carrier’s dataset can be used to realize them.

Shippers are recommended to start with reviewing the event log available to them in order to understand if and how this dataset represents the carrier’s process, before continuing to explore the other opportunities. This to close the gap between the process captured in the event log and the real-life process. Furthermore, it is suggested to collect the original event log of the carrier for reference, select a carrier that tracks many accurate data points in its last mile and understand the carrier’s process through cooperation. Finally, it is suggested that the findings of this thesis can be strengthened by understanding the relationship between shippers and carrier through future (literature) research.

Keywords: Process Mining, Outsourced last mile delivery, B2B, Discovery, Conformance checking, Enhancement, Prediction, Bottleneck analysis, event log data

Summary

Over the last 15 years, shippers have increasingly outsourced their last mile delivery process to external carriers. Outsourcing this process frees the shipper from the majority of the physical responsibilities such as driving to customers and maintaining a transportation network. However, the shipper is still held responsible for the quality of the overall delivery process by the customer. Failing to provide the customer with an adequate service may negatively impact the relationship between shipper and customer, especially in a business-to-business (B2B) market. Yet, a side-effect of outsourcing the last mile is reducing the shipper's abilities to influence the last mile process. A second trend in the last mile delivery sector is the increased availability of data describing the last mile process. Despite the increased availability, capitalizing on these data has been challenging. This thesis reviews if the increasingly available last mile data (i.e., event logs) can be used to enable the shipper to improve the external last mile delivery process. As such, the main Research Question is defined as:

How can carrier event log data be used to improve an external last mile delivery process?

This question is addressed by conducting experiments with the data available to shipper *Hilti AG*, and consist of three Research Questions (RQ's) depicted in Figure 1.

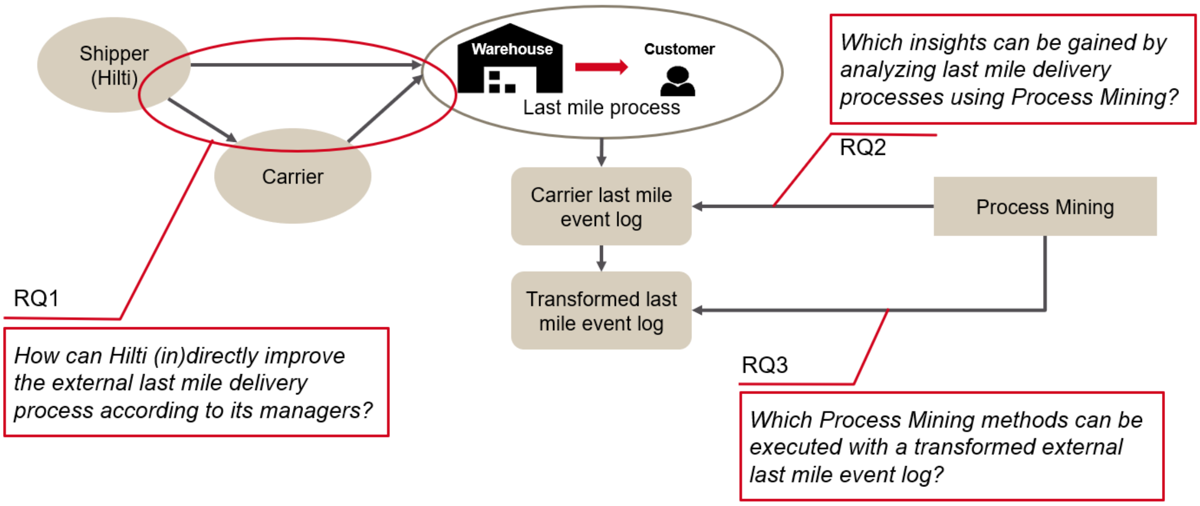

Figure 1. Research Questions

Findings

In answer to Research Question 1, Chapter 3 determined six relevant Opportunities to improve the last mile through interviews with the shipper's last mile management.

<u>Direct Opportunities</u>
Opportunity 1: Improve the data quality to structure the shipper´s last mile management
Opportunity 2: Use data to construct a (live) overview of parcels in the external process
Opportunity 3: Notify customers of a (potentially) late delivery based on live carrier data
Opportunity 4: Change the order of parcel in the carrier's truck to reach high-risk areas

<u>Indirect Opportunities</u> (depending on collaboration with the carrier to a large extend)
Opportunity 5: By creating transparency within the carrier network, the carrier improves the process
Opportunity 6: Find bottlenecks within the carrier network to avoid or solve them with the carrier

Chapter 4 illustrates how the dataset available to the carrier is created. The event log is generated during the delivery process, after which the data are transformed into a universal format and sent to the shipper. Here, the shipper enriches the transformed event log with internal data. This transformed and enriched dataset is the data available for analysis.

Chapter 5 addressed Research Question 2 through a literature review. It was found that, generally, three Process Mining methods are used to analyse event logs and are applied in the following order; First Discovery, then Conformance checking and finally Enhancement. Discovery commonly aims to extract a process described in an event log and allows the analyst to understand this process and all its variants. Conformance checking is commonly applied to compare a discovered process to a model of that process, pointing out similarities and differences. Finally, Enhancement aims to improve the process. Furthermore, it was found that Process Mining is applied to some logistic domains but has scarcely been used to analyse last mile event log data.

Chapter 6 aims to answer Research Question 3 by conducting experiments with the shipper's data. First, it is determined that the data hold the right information to be used for the Process Mining methods. A subset of the available data is selected to which Discovery is applied. It is found that the first model created is rather difficult to interpret due to the high number of process variants. However, the outline of the most common process can be extracted, detailing the steps (i.e., activities) of the process, total lead-time, time intervals between the activities and loops within the process. A second experiment is performed using fewer specific activities to reduce the number of variants, making it easier to assess if the discover model is logical. This is done using Conformance checking, comparing the process extracted from the dataset to the shipper´s understanding of the carrier´s last mile process. This assessment revealed that the process in the dataset, in principle, corresponds to Hilti's understanding of the carrier's process. It also showed that some parts of the process were most-likely misrepresented in the dataset, but the origin of this error was difficult to determine from the data alone. A third experiment consisted of a small case study in which Conformance checking was applied to a part of the process that was estimated to be accurate, aiming to see if original estimations of which parcel arrived on-time were correct. The case study discovered a deviating pattern which, if having occurred in reality, would indicate that a small number of parcels was wrongfully marked as having arrived on-time. Even though the carrier indicated that this issue was most probably caused by re-using packaging material of Hilti parcels, the example does demonstrate the potential of Conformance checking to discover important deviances in the external process. The experiments showed that using Discovery and Conformance checking techniques could be utilized when working towards realizing Opportunities 1, 2 and 5.

Unlike Discovery and Conformance checking, no practical experiments were executed with Enhancement. Rather, it was assessed how the dataset could be used to facilitate Enhancement analysis in future experiments. First, it was estimated that Opportunities 3, 4 and 6 may be realized using two Enhancement methods: Bottleneck analysis and Prediction. Furthermore, it was shown that the last mile can be tracked by eight data points, indicating departure and arrival times at physical locations throughout the last mile. In the reviewed dataset, two data points were missing, three were unclear or in-accurate and three were found to be accurate. It is expected that Enhancement's Bottleneck analysis and Prediction are most likely to be achieved with many accurate data points. Finally, it was shown that area's which have a low on-time-delivery rate can be determined from the dataset, which is a pre-requisite for Opportunity 4.

Conclusion

In answer to the main Research Question, six Opportunities were identified that can help improve the external process using the transformed event log data. Using Process Mining Discovery and Conformance checking, it was shown that the data quality can be increased (Opportunity 1), the carrier's process can be made more transparent (Opportunity 5) and that a (live) overview (Opportunity 2) can, in principle, be obtained. These three Opportunities and Opportunity 3 allow the shipper to improve the external last mile by empowering the shipper in its current last mile responsibilities.

It is speculated that Process Mining Enhancement's Bottleneck analysis and Prediction can notify the shipper or the customer of (potentially) late deliveries (Opportunity 3), increase on-time delivery rates by changing the order in which parcels are loaded (Opportunity 4) and find bottlenecks in the carrier network (Opportunity 6). Opportunity 3 has the potential to improve the customer's perception of service quality; Even if a parcel arrives late, the notification system avoids having them wait in vain. Moreover, Opportunities 4 and 6 may improve the on-time delivery rate itself by reaching high-risk areas on-time using altered truck loading (Opportunity 4) or by identifying and avoiding bottlenecks (Opportunity 6). Such concrete opportunities to improve the service quality and/or the on-time delivery rate using event log data and Process Mining Enhancement are, to the researcher's knowledge, new and potentially significant ways for shippers to improve their externally executed last mile delivery process.

Implications, Limitations and Recommendations

The experiments with Discovery and Conformance checking showed the impact of working with an event log that is transformed when intending to improve an external process. This since both the dataset and the understanding of the carrier´s process may deviate from the real-life process. As such, it is difficult to assess if a discrepancy between a discovered process and the process model is caused by an error in the data or the understanding of the carrier´s process. It is therefore recommended that either the transformed event log or the understanding of the carrier's process is verified as a first step, to facilitate quality insights using Process Mining. Verification of the transformed dataset could be achieved by collecting the original event log of the carrier, and a good understanding of the carrier process can be achieved by asking the carrier to verify if discovered processes are accurate.

After having verified the transformed dataset and/or verified the model of the carrier's process, it is recommended to continue with Opportunities 2 and 5, using Discovery and Conformance checking, before attempting Opportunities 3, 4 and 6 using Enhancement. This since both the experiments and the literature review showed that Process Mining Enhancement depends on a good understanding of the data and process, which can be obtained with applying Discovery and Conformance checking while attempting Opportunities 2 and 5.

Since experiments have only been executed with a transformed event log of one carrier in one region, more research should be done to better understand the effect transforming the event log and the effect of data quality on Process Mining. Furthermore, it is suggested that more research should be conducted on how the relationship between the shipper and the carrier relates to the findings of this research. Especially Opportunities 5 and 6, creating transparency and analysing bottlenecks, respectively, are largely dependent on the carrier to succeed. When attempting Opportunities 3, 4 and 6, which are enabled by Enhancement, it is recommended to select a carrier whose dataset holds a large number of accurate data points to make Bottleneck analysis and Prediction more precise.

Table of Contents

Acronyms

B2B	Business to Business
B2C	Business to Consumer
CRM	Customer Relationship Management
ERP	Enterprise Resource Planning
GPM	Global Process Manager
HS	Harmonized status
HU	Handling Unit
ICT	Information and Communications Technology
KPI	Key Performance Indicators
MO	Market Organisation
PAIS	process-aware information systems
RFID	Radio-frequency identification
RHT	Regional Heads of Transport
RTE	Regional Transport Expert
RQ	Research Question
SCM	Supply Chain Management
SQL	Structured Query Language
WFM	Workflow Management

List of Figures

List of Tables

Chapter 1. Introduction

1.1 Topic relevance

The percentage of shippers that outsource their last mile delivery to external carriers has steadily increased over the last one and a half decade and is expected to do so in the foreseeable future (Langley & Capgemini, 2012; Langley & Infosys, 2019). Outsourcing last mile delivery reduces the shippers´ influence on the delivery process. Nevertheless, the outsourced (last mile) delivery process still affects how customers perceive the service quality of the shipper (Vakulenko, Shams, Hellström & Hjort, 2019). As service quality of the delivery process is shown to be a key-factor in Business-to-Business relationships (Sohn, Woo & Kim, 2017), it is in the shippers' interest to find ways to improve the customer´s experience of last mile delivery. This can be challenging since the process is largely executed by the external carrier, over which the shipper has limited control. Developments in data-sharing and IT-capabilities (Sohn et al., 2017) may present an opportunity to overcome this challenge by giving the shipper more insight into the process. Yet, companies generally find it difficult to use available data to obtain the desired insights (Mikalef, Boura, Lekakos & Krogstie, 2019). In an attempt to contribute to overcoming these challenges, this thesis explores how data detailing the last mile can be used to improve the external last mile delivery process from a shipper´s perspective. This exploration is executed within the context of a large tool manufacturer, Hilti AG.

1.2 Company description

1.2.1 Hilti AG

Hilti AG develops, manufactures and markets products and services for professional construction. With a strong emphasis on innovation and quality, Hilti aims to serve the high-end segment of the construction world and almost exclusively offers its products and services on the B2B market.

One of the pillars of Hilti's strategy is direct customer contact. Instead of selling products through a retailer, Hilti representatives go to construction sites in person and directly assess the needs of customers. The effect of this can be seen in Hilti's distribution network where most products are directly delivered to construction sites or offices rather than to personal addresses or via retailers.

1.2.2 Last mile delivery at Hilti

Langley and Infosys (2019) state that more than 50% of companies, among which many manufacturers, outsource their logistical processes. Similarly, Hilti's last mile delivery process is largely outsourced. For every region, multiple external logistics companies, i.e., carriers, take products from Hilti warehouses to the customer. Here, the last mile is defined as all logistic movement between Hilti's warehouses and the customer. Through its carriers, Hilti offers different delivery services depending on region, carrier, and the product being delivered. Common services are express delivery, early morning delivery (before 10 am or before noon), standard delivery, delivery by appointment or special job site services. These many different and sometimes complex delivery services make the Hilti's last mile delivery of products diverse and complex.

1.3 Problem context

1.3.1 Need for measuring and improving last mile delivery

For Hilti, improving the last mile delivery process is in line with its long-term strategy to serve the high-end sector of the construction world. Key elements of this strategy are

- having high service standards to outperform the competition in service quality, and
- being a fast follower in using new techniques to increase service levels.

An important metric in measuring service quality is on-time delivery (Milgate, 2001; Karim et al., 2010; Nakandala, Samaranayake, & Lau, 2013). The optimal on-time delivery rate depends on many factors (Burns, Hall, Blumenfeld, & Daganzo, 1985; Boyaci & Ray, 2006; Forslund & Jonsson, 2010). No optimal rate has been determined for Hilti, but following their strategy, they aim to be at least as good as their competitors. However, estimating if this is the case is difficult, as the current on-time delivery rate is unclear. Carrier reports state that Hilti's on-time delivery rate is 97.7%, while estimates based on process data range around 92%. Such discrepancy between two measures of on-time delivery is not uncommon, as 84% of carriers and shippers have a non-standardized definition of on-time delivery (Forslund & Jonsson, 2010). However, the discrepancy does make it difficult to estimate if Hilti is ahead of the competition in service quality. Moreover, an internal multi-year customer survey (see Appendix A) showed that perceived delivery service levels have been reduced, followed by stagnation. This to Hilti's strong disliking. As such, knowing the true on-time delivery rate is important for Hilti management. Furthermore, as Hilti aims to be a fast follower in implementing new techniques to increase the overall service quality for customers, it is in line with Hilti strategy to see which (technological) *Opportunities* are available in order to improve on-time delivery and/or perceived service levels.

1.3.2 Ad-hoc last mile management

The last mile is executed and managed by a mix of Hilti and carrier personnel. The delivery itself is processed by the carrier, while Hilti employees handle supportive processes.

Until recently, Hilti employees were mostly unaware of what occurred during the physical delivery process. Although process data have been retrieved for over a decade, these data were mostly unstandardized and unstructured. Once the parcel had been picked up by the carrier, Hilti did not monitor its status until delivery at the customer or when a customer called to inquire about the status of the parcel. As such, the next update Hilti received after a pickup from the carrier was either an on-time delivery confirmation, a failed delivery or a last-minute error that had to be resolved immediately. As a result, managers were largely unaware of the delivery process, did not have real-time insight in the whereabouts of the delivery and could only act responsively to occurring problems. Combined, these elements describe a responsive management style by which Hilti's current last mile management can be characterised. Not only does a lack of insight into the physical process force the last mile team to solve problems ad-hoc, but it also makes improving the last mile process a challenge.

1.3.3 Use of transformed carrier data

In recent years, Hilti has increased its efforts in standardizing and structuring last mile data from its carriers to gain insight into the full process. Specifically, Hilti now collects the so-called *event logs* of almost every delivery in Great-Britain, Spain, Italy, France and the German-speaking area. An example of an event log is depicted in Figure 2.

	A	B	C	D	E	F	G
1	CaseID	Timestamp	Medium	Status	Service Line	Urgency	
2	case9700	20.8.09 11:46	Phone	Registered	1st line	0	
3	case9700	20.8.09 11:50	Phone	Completed	1st line	0	
4	case9701	23.9.09 12:23	Phone	Registered	1st line	0	
5	case9701	23.9.09 12:27	Phone	Completed	1st line	0	
6	case9705	20.10.09 14:21	Phone	Registered	Specialist	2	
7	case9705	20.10.09 16:48	Phone	At specialist	Specialist	2	
8	case9705	19.11.09 10:31	Phone	In progress	Specialist	2	
9	case9705	19.11.09 10:32	Phone	Completed	Specialist	2	
10	case3939	15.10.09 11:48	Mail	Registered	Specialist	2	
11	case3939	15.10.09 11:48	Mail	Offered	Specialist	2	
12	case3939	20.10.09 17:18	Mail	In progress	Specialist	2	
13	case3939	20.10.09 17:19	Mail	At specialist	Specialist	2	

Figure 2. Example of an event log. Adapted from *Data Requirements*, by Fluxicon, n.d., retrieved from https://fluxicon.com/book/read/dataext/ Copyright 2020 Fluxicon

As each carrier has its own data system, comparing and analysing the collected data from different carriers is a challenge. As a first step, Hilti has developed a code transformation system that aims to unify these data streams. A transformation-map was created for every carrier, converting the respective carrier codes to a standard list of Hilti codes. Every code the carrier sends is transformed into a Hilti code, aiming for one data-system where similar events from different carriers are recorded by the same, universal Hilti code.

The available data enabled Hilti to start and understand the last mile process. Currently, the data are used to calculate Key Performance Indicators (KPIs), which allow Hilti to check the success rates of the carrier. However, given that the data describe an externally managed process and almost all carriers have a different coding system, analysing the data any further has not been straight forward. More importantly, Hilti is currently unaware of how these data can be used to improve the delivery process itself.

In summary, Hilti aims to provide customers with a high-standard last mile delivery process, fitting its overall service level. Yet, current on-time delivery rates are unclear and may be a reason for a reduction in perceived service levels. Furthermore, the last mile management team indicates to be spending a disproportionate part of their time on "firefighting", ad-hoc solving of urgent problems leaving them with little time and resources for structural improvement. Lastly, the last mile process and dataset being in control of an external carrier make improving the process more challenging. Analysing how Hilti can improve the last mile process by using the last mile data may function as an example of how other companies can improve their externally executed last mile delivery process.

1.4 Research Questions

1.4.1 Problem statement

As described in the previous section, Hilti is actively collecting and analysing carrier data in an attempt to improve the external last mile delivery. However, using these data has proven to be challenging. Therefore, the problem statement of this thesis is:

How can carrier event log data be used to improve an external last mile delivery process?

To find an answer, three Research Questions are answered. An overview of how these Research Questions relate to each other is presented in Figure 3.

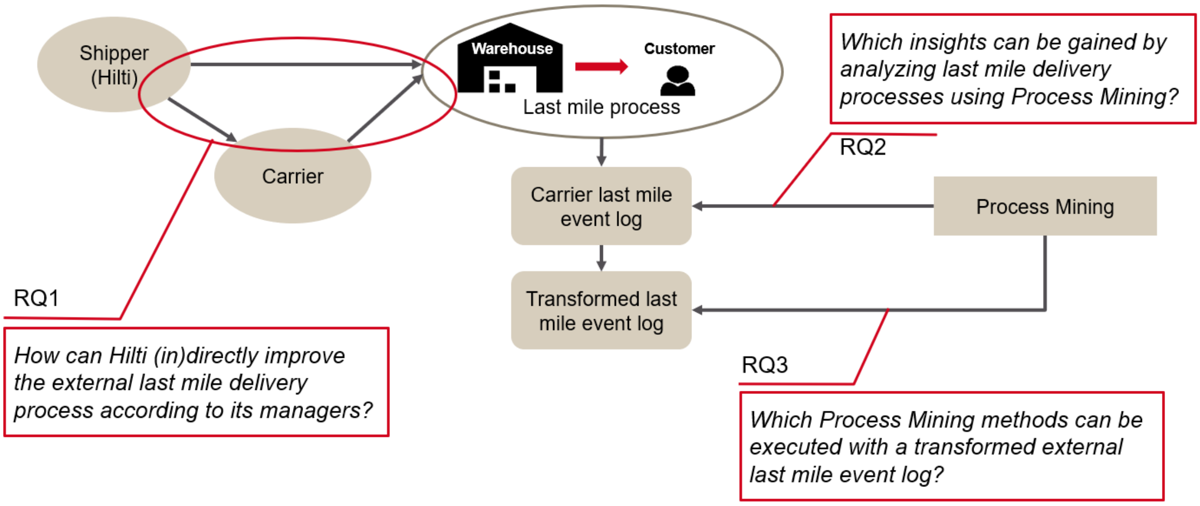

Figure 3. Research Questions

1.4.2 Research Question 1

The last mile delivery process is largely executed by the carrier. This has an impact on Hilti's ability to improve the last mile delivery process since Hilti has a less direct impact on the process. This context is important since possible insights that can be gained from the data are only relevant if Hilti can actively use these insights to improve the external process.

Therefore, it is first explored how Hilti can improve the last mile process, either directly or through the carrier. Explorative research was conducted by asking Hilti's last mile management team how they believe Hilti can improve the last mile. This is addressed by the first Research Question:

How can Hilti (in)directly improve the external last mile delivery process according to its managers?

The answer to this question poses as a starting point of the thesis, as it exposes those points where Hilti can, directly or through the carrier, improve the last mile process.

1.4.3 Research Question 2

Hilti is not fully aware of what potential lies within the data and how to extract value from it. To address this, literature was consulted to determine which insights can be gained by applying *Process Mining*, the branch of data science dedicated to extracting value from event logs, to last mile event logs with the second Research Question:

Which insights can be gained by analyzing last mile delivery processes using Process Mining?

1.4.4 Research Question 3

Research question 2 addresses how event log data detailing the last mile process can be analysed using Process Mining. However, the fact that the dataset is created by, and describes the process of, an external carrier, is not incorporated in Research Question 2. As the transformation of the external data may have a significant effect on the analysis in practice, an experiment was performed to see which of the Process Mining methods can be executed with a transformed external dataset. The third Research question addresses this by answering:

Which Process Mining methods can be executed with a transformed external last mile event log?

Conclusions from this question are Hilti specific since the dataset used may be different from those of other carriers or shippers. Findings may serve as an indication of how other external last mile event logs can be analysed.

1.5 Thesis structure

The structure of the remainder of this thesis is as follows. Chapter 2 contains a description of the last mile process. The chapter includes details of the role division between Hilti and the carrier, which serves as a context to understand the dynamic between the two parties. Chapter 3 addresses the first Research Question using qualitative analysis. This by conducting an explorative interview with Hilti management aiming to discover ways Hilti management can improve the last mile. Chapter 4 describes the properties of the data available to Hilti in preparation of the literature review. In Chapter 5, a literature review provides details on how Process Mining can analyze (last mile) event log data in answer to Research Question 2. In Chapter 6, Research Question 3 is addressed, reviewing which analysis are possible for Hilti´s transformed dataset through experiments with the dataset. Using the outcomes of the Research Questions, Chapter 7 discusses the findings, their implications and recommendation of this thesis. Here, attention is given to the following:

- Which of the Opportunities identified by Hilti management can (potentially) be achieved with Hilti's transformed event log.
- What general Opportunities are present for shippers wanting to improve the last mile delivery process using an external event log.

Finally, Chapter 8 concludes the thesis, addressing the problem statement.

Chapter 2. Description of last mile delivery process

This thesis was conducted in the context of Hilti's last mile. To understand how Hilti can improve the last mile process, it is crucial to understand the setup of Hilti's last mile delivery process, which will be done in this chapter. This to assist the reader in understanding the context within which this thesis took place. The second part of Chapter 2 provides a specification of the role division between Hilti and the carrier within the last mile.

2.1 Defining the last mile

The term "last mile", in a logistical context, is used in different ways throughout academic literature. A systematic literature review by Olsson, Hellström & Pålsson (2019) confirms the ambiguity of the term. The definition used in the remainder of this thesis should therefore be specified.

Many articles (Reyes, Savelsbergh & Toriello, 2017; Akeb, Moncef, Durand, 2018; Lim, Jin & Srai, 2018) exclusively apply the term last mile to a Business-to-Consumer (B2C) context. However, others replace the term consumer with "customer", a term that also refers to business customers (Olsson, Hellström & Pålsson, 2019). As such, business to customer last mile logistics includes business-to-business (B2B) processes (Saenz, Figliozzi & Faulin 2016; Hübner, Kuhn, Wollenburg & Trautrims 2018). Examples of the term last mile being used in B2B context are delivery to brick-and-mortar stores (Kin, Ambra, Verlinde & Macharis, 2018) and hospitality services (Fancello, Paddeu & Fadda, 2017). As Hilti almost exclusively has business customers, the last mile will be viewed in a B2B context.

The end of the last mile process is similar in almost all definitions and can be called the destination point. As this is often the point selected by the consignee (a home, office, reception box or pickup location) this point is also called the consignee preferred pickup point.

However, the start of the last mile differs between B2C and B2B contexts. In a B2C context, the last mile only entails the very last part of the process, between the final distribution hub and the customer (Gevaers, Van de Voorde & Vanelslander, 2009). The full outbound delivery process entails many steps to get a parcel from the business to the customer. In a B2C context, the term last mile is reserved for the very last step, often including delivery by small vans in an urban area. In a B2B context, the last mile process varies more since business customers are frequently located in non-urban areas as well. The process as such does not have the characteristics of a typical B2C last mile such as urban areas or the use of small vans.

This variety includes different starting points. The term to indicate the varying starting points is introduced by (Fernie and Sparks, 2009), which they call the "order penetration point". This point refers to an inventory location (e.g., fulfilment centre, manufacturer site or warehouse) where a fulfilment process is set in motion after an order from a customer.

Finally, some definitions of the last mile include the process within the warehouse. However, this thesis focuses only on the transportation between the starting and ending point in a B2B setting and excludes order picking and other process steps in the starting location, similar to the definition Gevaers, Van de Voorde and Vanelslander (2009).

As a result, for this thesis, the last mile is defined as:

> "The last mile is the stretch of parcel delivery service that takes place from the order penetration point to the final consignee´s preferred destination point" based on the definition by Lim, Jin and Srai (2018)

2.2 General last mile process

Hilti's last mile process is depicted in Figure 4. The process starts at the Hilti warehouse, where Hilti products are stored until shipped. With a few exceptions, Hilti holds stock of all its products in its warehouses. After an order has been placed by the customer, the order is prepared for shipment by Hilti employees. After the needed parcels are picked from the warehouse, Hilti sends a digital request for a pickup to the carrier including delivery details. In return, the carrier sends a label which Hilti attaches to the parcel. The label enables the carrier to keep track of the parcel during delivery and send information on the parcel to Hilti.

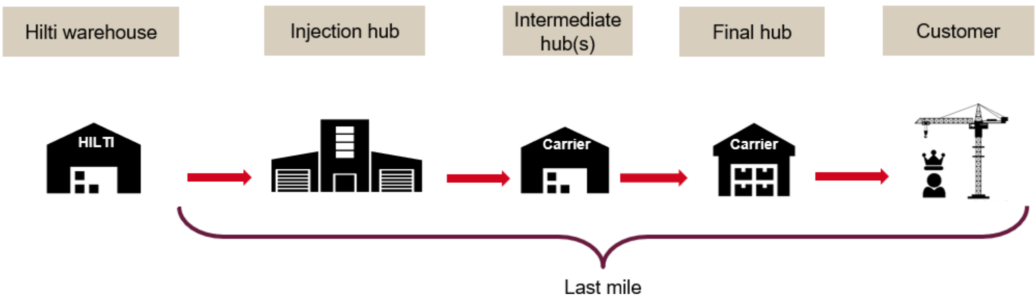

Figure 4. Schematic overview of Hilti's last mile

The parcel is picked up by the carrier at the Hilti warehouse. Although it varies per warehouse, there are usually one or two trucks per carrier per day, collecting goods. The parcels are loaded into the truck by Hilti personnel. After pickup, the parcel remains in the carrier network. The next steps the parcel goes through vary per route and service level (the time within the parcel has to reach its destination). Often, the parcel is moved to the closest carrier location, called the injection hub. Depending on the route, the parcel is then moved to intermediate hubs, small buildings where parcels are unloaded from one truck and moved to the next one, to be rerouted to a new location. Finally, the parcel arrives at the final hub. Here it is moved onto the final delivery van or truck, which takes the parcel to the delivery address. It occurs that the delivery does not succeed in one go. If this is the case, the parcel is returned to the final hub, and a new attempt is made on the following business day. After delivery, a confirmation of delivery, including a signature of acceptance, is collected by the carrier and send to Hilti. The timestamp of this confirmation is regarded as the time of delivery and determines if the delivery should be considered on-time or late.

2.3 Division of responsibilities

As has become clear, the last mile is an interaction between Hilti and the Carrier. From the moment the parcel is picked up from the Hilti Warehouse, it is the carrier's responsibility to transport the parcel to the customer. However, other responsibilities are to be considered.

2.3.1 Carrier responsibilities

The carrier executes the physical part of the delivery. The carrier is free to determine how it gets the parcel from the Hilti warehouse to the customer at the delivery address. It is however the carriers' responsibility to do this within the timeframe associated with the chosen service level (standard delivery, express delivery, etcetera). It is also the carrier's responsibility to forward the confirmation of delivery to Hilti after it has been signed by the customer.

In recent years, Hilti has requested to receive internal status i.e., event codes, detailing each delivery. The contract between Hilti and the carriers specifies that Hilti wants to receive these codes and asks the carrier for collaboration.

Lastly, the carrier is also responsible for reporting the transport reliability i.e., on-time delivery rate, to Hilti. This is a monthly report the carrier sends to Hilti which details how many deliveries were delivered at the customer within the agreed-upon time limit. The carrier usually includes the reasons why the delivery reached the customer outside the time-window.

2.3.2 Hilti's responsibilities

The information regarding the parcel is Hilti's responsibility. Hilti needs to provide the carrier with information on the physical parcel to be delivered (e.g., size, weight, etcetera.), as well as details on the final point of delivery. If it turns out that the address of the customer was incorrect or incomplete during the process, Hilti management will be asked to provide a solution. Similarly, Hilti is responsible for safety information regarding the parcels shipped. As several Hilti products contain chemical substances, for example, these products are classified as dangerous goods and should be handled and registered accordingly. If during transportation any ambiguity arises regarding these dangerous goods, Hilti is expected to assist in solving the situation.

Hilti also handles the preparation of the goods before shipment and loads the carriers' trucks in the Hilti warehouse. Furthermore, the parcel is packaged and wrapped by Hilti before departure.

The third responsibility of Hilti is communication with the customer. As mentioned in Chapter 1, part of Hilti's core strategy is a strong relationship with the customer. Hilti aims to communicate and inform the customer with one voice, as best as it can. Before shipment, Hilti currently informs the customer when the delivery has been collected, when the delivery is ready for shipment and within how many days the customer can expect the delivery. The carrier has a phone number to be used during delivery but does not provide the customer with status updates while the parcel is in the carrier network. Thus, after the carrier has picked up the goods at the Hilti Warehouse, the customer no longer receives updates until the parcel has reached the customer. An overview of the division of actions and communication required by all parties, in the standard situation, is depicted in Figure 5.

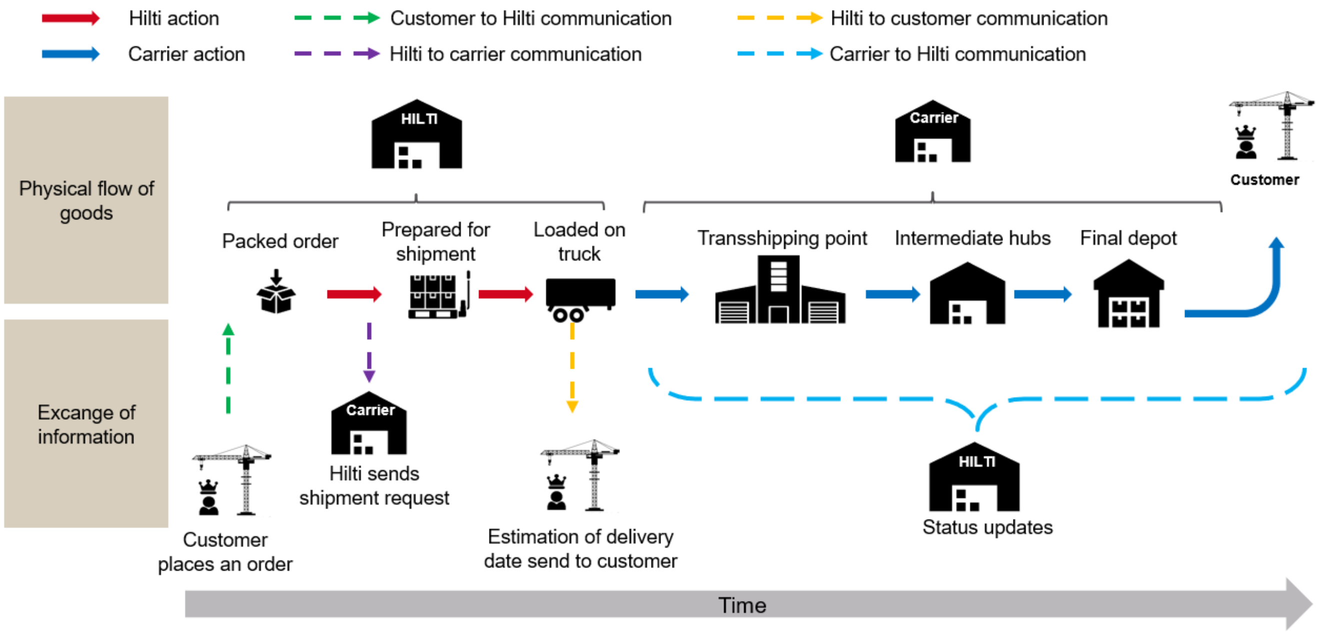

Figure 5. Division of actions and communications in the last mile

Chapter 3. Research Question 1

For an outsider, it is difficult to assess how Hilti can improve the external last mile process, since the daily operation and interaction between the carrier and Hilti is complex, changes per region/carrier and is often very specific.

Exploring how the data can be used to improve the process was therefore approached from Hilti's perspective by interviewing Hilti's last mile management, in answer to Research Question 1:

How can Hilti (in)directly improve the external last mile delivery process according to its managers?

Here, an improvement is considered direct if its success is not dependent on the carrier. Indirect improvements are those which depend on the carrier to a large extend.

3.1 Approach

3.1.1 Participants

There are three main management groups within Hilti's last mile management:

1. Global Process Managers (GPMs)
2. Regional Heads of Transport (RHTs)
3. Regional Transport Experts (RTEs)

Besides the main management groups, a fourth group, Business Analysts, can be distinguished. To collect improvements from different perspectives, members of all four groups were asked to participate in the interview.

<u>Group 1: Global Process Managers</u> (3 participants)
GPMs launch new initiatives, set Global standards and facilitate interregional collaboration. GPMs have a 'bird-eye' overview of the processes in different Market Organizations (MOs). Three GPMs that were (in)directly related to last mile management or the carrier agreed to participate in the interview.

<u>Group 2: Business Analyst</u> (1 participant)
Business Analyst is a function within Hilti that assists many Global management areas from an IT perspective, including last mile data. One business analyst that has worked with the carrier data before agreed to participate.

<u>Group 3: Regional Heads of Transport</u> (3 participants)
RHTs are responsible to oversee all the incoming and outgoing deliveries within their region, manage daily operations and collaborate with GPMs to align overall strategies and address specific topics. RHTs from the MOs France, Germany and Italy agreed to participate. Generally speaking, RHTs have a good idea of what data are recorded in the event log but have limited hands-on experience analysing the data.

<u>Group 4: Regional Transport Experts</u> (4 participants)
RTEs handle the daily delivery process and support Heads of Transport in managing their respective regions. RTEs from the MOs: France, Germany, Italy and Spain participated in the interview. Transport Experts are increasingly working with the carrier data to understand the external last mile. Their experience with the data is more extensive than those of the RHTs.

3.1.2 Interview structure

Lune & Berg (2017) describe that an unstandardized interview is appropriate when:

- identical questions are likely to be interpreted differently by different respondents,
- it is uncertain which direction the conversation will go, and
- a topic needs to be explored and understood rather than confirming/disconfirming hypothesis.

For this interview, the group of respondents is considered heterogenous, considering their different functions within the last mile, the different MOs for which they are responsible and their varying levels of experience with the data. It was therefore expected that questions would be interpreted in different ways and that the interview should be able to go in various direction. As such, an unstandardized interview approach was selected. This way, respondents were free to show their vision of how the data can be used to improve the last mile. Furthermore, the interview was rather an exploration effort than an exhaustive analysis, corresponding with the goal of an unstandardized interview.

The unstandardized interview is structured by guidelines, which are based on conceptual areas (i.e., topics) that need to be covered during the interview (Lune & Berg, 2017). The following topics are relevant for this interview:

1. points of interaction/influence with the carrier
2. current issues or opportunities in the last mile (general)
3. potential opportunities to use carrier data

These topics needed to be addressed during the interview. Furthermore, the tone of the interview was casual and based on interaction, as is common for unstandardized interview. Finally, the interviewer aimed to introduce the topic of using carrier data towards the end of the interview, to leave room for opportunities that may be indirectly related to the carrier data.

3.1.3 Processing steps

During the interviews, all topics discussed were noted down in the form of statements attributed to the respondents. These statements were analysed to select those that answer Research Question 1.

With this goal in mind, Opportunities were formed out of the statements and evaluated, as suggested by Stuckey (2015). The process consisted of three phases:

<u>Phase 1: Cleaning the list of responses</u>

First, only those statements that refer to the topics described in section 3.1.2 were retained. All other statements were dropped (i.e., unrelated topics, function descriptions of the respondent, etc). Some of the statements were slightly re-phrased to have meaning outside the context of the interview. This means that:

- abbreviations were written out,
- partial sentences were completed, and
- small additions to the statements were made to clarify their meaning.

<u>Phase 2: Finding common themes within statements</u>

After the statements were cleaned, Phase 2 aimed to find common themes among them. An approach to creating themes in qualitative data is described by Lune & Berg (2017). They state that by scanning the data, i.e., statements, and tracking the individual themes that the statements refer to, common themes can be identified. Since the dataset consists of individual statements (as opposed to whole interviews), it was relatively straightforward to scan through the list of statements and note their respective themes. After doing so, the statements with common themes were grouped.

Phase 3: Selecting Opportunities

There is no clear-cut way of analysing qualitative results. Doing so involves creativity as much as structure (Lune & Berg, 2017, p.90) and should be approached from the perspective of the goal of the analysis (Stuckey, 2015, p.8). The goal of this analysis, as stated in Research Question 1, was to present opportunities that have the potential to improve the last mile by using the event log data. Furthermore, it is important to note that this assessment aimed to explore opportunities as Hilti management sees them, rather than exhausting the total pool of opportunities. As such, a value judgement was made of the remaining statements aided with input from Hilti. The assessment entailed four steps:

Step 1: A format was needed to make the statements more comprehensible and comparable. Following the goal of Research Question 1, all statements were transformed into the Opportunities they present to improve the last mile. Furthermore, similar Opportunities were merged. The names of the respondents making the original Opportunities were added to the merged Opportunity, to keep track of which respondents contributed to which Opportunities.

Step 2: After the statements were transformed into Opportunities, those most relevant were selected. This was done based on two attributes: which Opportunities relate to using event log data and which Opportunities are deemed most relevant by Hilti management?.

Step 3: Having a list of relevant Opportunities, it was seen if multiple Opportunities can be merged into an “overarching” Opportunity, better representing the idea that the individual Opportunities aimed to communicate.

Step 4: In this final step, the goal was to estimate the potential impact and implementation effort of the remaining Opportunities in collaboration with Hilti management. This entailed presenting the remaining Opportunities to Hilti’s last mile managements, discussing the estimated impact and effort of each Opportunity. Based on this discussion, the researched positioned the Opportunities on a matrix, marking the estimation of their relative impact and effort. Finally, this overview was presented to management asking if they agreed with the relative effort and impact of the Opportunities. Finally, it was estimated per Opportunity if it aims to improve the external last mile directly or indirectly through the carrier.

3.2 Processing interview outcomes

Phase 1 reviewed the 59 statements that came out of the interviews on relevance. All irrelevant statements were dropped, which resulted in 34 relevant statements. The statements were then adjusted if needed so that their meaning was clear outside the context of the interview. The 34 relevant and re-phrased statements are listed in Appendix B.

In Phase 2, these 34 statements were reviewed to find common themes. The identified themes and their respective number of statements are listed below.

- Human resources (1 statement)
- Customer interaction (2 statements)
- Data quality (3 statements)
- Live overview (2 statements)
- Notification (8 statements)
- Data analysis (6 statements)
- Motivating the carrier (7 statements)
- Collaborating with the carrier (5 statements)

Phase 3's goal is to select the relevant Opportunities and consisted of four steps.

Step 1 transformed the statements into Opportunities. Furthermore, similar Opportunities were merged, which resulted in 24 distinct Opportunities listed in Appendix C.

Step 2 aimed to select only the most relevant Opportunities, based on if they can be obtained by using the event log data and if Hilti deems them relevant. Examples of dropped Opportunities are those that were too general (i.e., "There are lots of unexploited opportunities with power BI: apps, AI, Microsoft flow"), had no clear goal (i.e., "It could be interesting to see which project businesses are delivered on-time") or those that have already been implemented (i.e., "Cases of discrepancies between the data and the carrier report are discussed with the carrier every 1-2 months. This keeps carriers sharp"). The 11 Opportunities which are deemed relevant are highlighted green in Appendix C.

Step 3 reviewed the Opportunities once again, to see if any multiple of Opportunities could be merged into an "overarching" Opportunity. This was done for Opportunities in the categories "Data quality", "Notifications" and "Data analysis", transforming the 11 relevant Opportunities into 6 final Opportunities. An example of the theme "*Notification"* is illustrated below.

- Opportunity 1: When an HS-code (Harmonized status code) is generated, there could be an automated action following, helping the region manager to take proactive measures (Marco Battista, Johan De-Smedt, Olivier Haas)
- Opportunity 2: Use the data to check if everything leaves on-time from the depots [warehouse and hubs] and receive a notification otherwise. (Thomas Krohn)
- Opportunity 3: Logistical data could be used to see if a parcel is able to make it to the customer and Hilti could receive an update if the parcel will be late. (Marco Battista, Thomas Krohn, Oliver Weich, Oliver Haas)
- **Overarching Opportunity**: Create a notification system that automatically provides Hilti Managers with relevant notifications based on live carrier data, such as timely departure, error events or prediction of a timely arrival. (Marco Battista, Johan de-Smedt, Thomas Krohn, Oliver Weich, Oliver Haas)

Step 4 entailed collaborating with Hilti management to estimate the execution effort and potential impact of the remaining Opportunities. Two managers, a GPM and RTE, agreed to assist in the estimation. Both managers had experience with the dataset and had worked closely together with the other management levels involved in the interview, making them best suited for estimating how the opportunities could unfold. It should be noted that these managers also participated in the interview, and are thus not fully independent. The six remaining opportunities were presented to the managers, and the estimated impact and effort of each opportunity was discussed. No scoring method was used. Based on the discussions, the researcher created Figure 6 which was then verified by the managers.

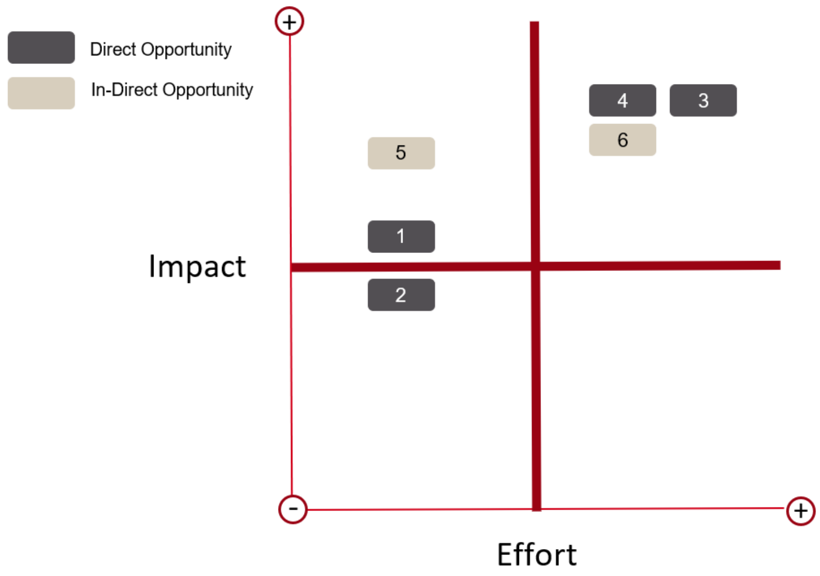

Figure 6. Impact/Effort matrix of selected Opportunities

Furthermore, it was determined that four out of the six remaining Opportunities are direct means to improve the external last mile and two Opportunities involve the carrier, making them indirect Opportunities. The six remaining Opportunities are described below, and the estimated impact and effort required is discussed in section 3.3 in answer to Research Question 1.

<u>Direct Opportunities</u>

Opportunity 1: Data quality

By improving the data quality, Hilti will be able to base its management decisions on a systematic representation of the carrier network rather than on manual moments of contact with the carrier. (Oliver Haas, Johan de Smedt)

Opportunity 2: Live Overview

Having a live overview of parcels in the carrier network would make managing more effective. (Thomas Krohn, Luca Mandelli)

Opportunity 3: Notification system

Creating a notification system would automatically provide Hilti Managers with relevant updates based on live carrier data, such as timely departure, error events or prediction of a timely arrival. (Marco Battista, Thomas Krohn, Oliver Weich, Oliver Haas)

Opportunity 4: Optimized parcel loading

The order by which the parcels enter the carrier's truck is determined by Hilti. If it can be shown that certain areas are tricky to reach, they could be put in the truck in such a way so they are sorted first in the depot [hub] and have a higher chance of reaching the customer. (Johan de Smedt)

Indirect Opportunities

Opportunity 5: Transparency

The more transparency we can create in the carrier's process by using the data, the more of a priority Hilti becomes for that carrier.
(Gerrit Budde, Johan de Smedt)

Opportunity 6: Bottleneck analysis

Analyse carrier data to find root causes/bottlenecks in the carrier system, such as problematic carrier hubs or postal codes with an increased risk of receiving deliveries late.
(Ralf Eggenberger, Johan De Smedt, Thomas Krohn)

3.3 Answer to Research Question 1

In answer to Research question 1: "*How can Hilti (in)directly improve the external last mile delivery process according to its managers?"* six Opportunities are presented, including their estimated impact and implementation effort. These Opportunities are further discussed below.

Opportunity 1 and 5 were categorized as having the most optimal effort/impact ratio. Of these, Opportunity 1 describes improving the data quality to make it usable. This could prove to be critical in enabling other Opportunities, and in using the event log data in general. Opportunity 5 suggests that any method of analysing the event log data that provides Hilti with a better understanding of the carrier process can bring about a better delivery performance in practice. This sets the bar conveniently low for suggesting analysis methods and can perhaps be seen as making the first steps in using the carrier data valuable. This Opportunity does depend on the carrier´s response to the increased transparency, making Opportunity 5 indirect.

Opportunity 2 was estimated to have the least impact, but also costing the least effort to implements. Currently, Hilti's ERP system can show received information for each parcel. Those managers that know from experience which carrier sends updates shortly after the event, may know how reliable the current overview is. Furthermore, it is technically possible to receive the data in a timely fashion from almost all carrier with an acceptable amount of effort. However, the currently used ERP system has limited functionality to show what occurred to the parcel before the last event, making a new overview add value. The exact impact this improvement has on Hilti's capability to manage the last mile is therefore uncertain. Hilti does however indicate to be interested in researching the impact of such an improvement.

Finally, Opportunities 3, 4 and 6 were estimated to have a high potential impact combined with high implementation effort. These three Opportunities are a combination of a more sophisticated way of using the data, combined with a currently unknown implementation. Opportunity 4 and 6 have in common that they could increase the on-time delivery rate of Hilti parcels at the customer. Opportunity 6 largely depends on the cooperation of the carrier since the carrier is expected to solve bottlenecks if identified, making Opportunity 6 indirect. Opportunity 3 could allow Hilti to improve the customers' experience in case a parcel seems to be arriving late. Hilti indicates that such capability to receive important notifications of failing deliveries is as important as preventing failures, making Opportunity 3 as relevant as Opportunities 4 and 6.

Chapter 4. Data description

Having explored how Hilti believes they can improve the last mile in Chapter 3, this chapter describes the process of how the dataset was generated and made available to Hilti in preparation of reviewing the literature in Chapter 5. This process is important context for the remainder of the thesis since a fundamental part of the final dataset is generated and communicated by the external carrier and is not Hilti's own. As this fact can have a considerable impact on current Opportunities within the data, a description of the dataset creation process is needed.

The process from the generation of the data until the final dataset available to Hilti can be divided into three phases:

Phase 1. Generation of the data by the carrier
Phase 2. Retrieval of carrier data and transformation into Hilti's code system (HS-codes)
Phase 3. Enrichment of the dataset

These phases are represented in Figure 7.

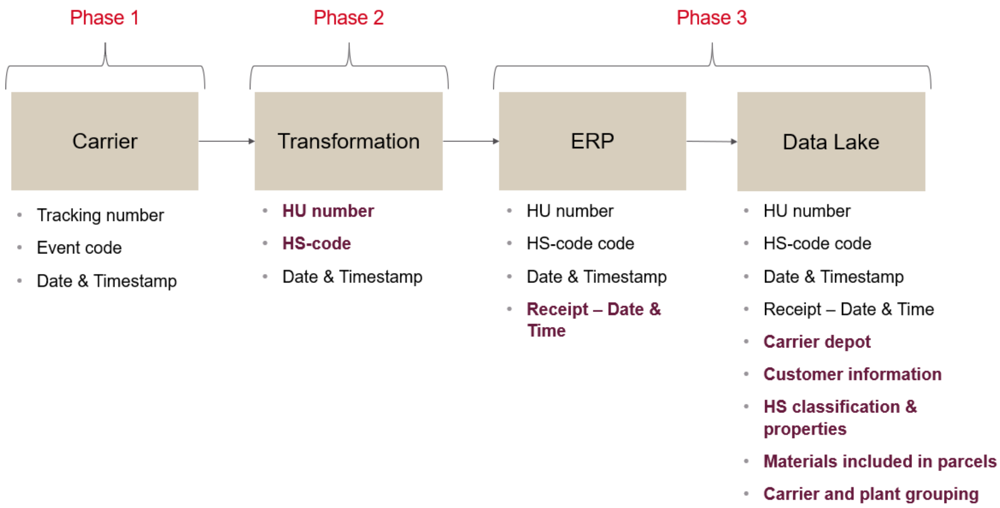

Figure 7. Transformation of event log data

4.1 Phase 1: Generation

Carriers that register and share information of events that occur to Hilti's parcels are called integrated carriers, which will be considered for the analysis. Several integrated carriers, among which a few of the large carriers operating in Europe, have become integrated carriers in the last few years.

Registration of events lies at the core of the event log. What information the integrated carriers register can be explained as follows: the carrier has a list of possible events (i.e., arrival at place X, delivery at the customer, etcetera.) which all have a unique event code attached to them. Every parcel has a unique tracking number, which is physically attached to the parcel as a barcode label. When the carrier scans the label at a specific location, an event code is generated and stored in the carrier´s data system linked to that tracking number, together with the date and time of the event code. As a result, each event is described by three data points:

Data point 1. the tracking number of the parcel
Data point 2. an event code
Data point 3. the timestamp of the event

A collection of multiple parcels and their respective events is called an event log. Every carrier has a unique system of tracking parcels and their events. The systems have large differences in the number of events registered, the level of detail of events and the codes used to represent the occurrence of an event for example. At a minimum, an integrated carrier will register an event when:

- he leaves the hub to deliver the parcel to the customer,
- the parcel was delivered to the customer, and
- a delivery failed, including information on who was responsible for the failure.

Most integrated carriers also register when a parcel has left an interim hub. Other carriers also register so-called 'error codes' when an unexpected event occurred.

4.2 Phase 2: Transformation

After the event has occurred and is registered, the event log is made available for extraction. How much time after the event the data are available differs per carrier and ranges between mere minutes or hours. Hilti uses an external party to extract the event log from the carrier, apply a transformation and send the result to Hilti. This transformation takes the event codes from the carrier and turns them into HS-codes, a coding standard defined by Hilti themselves. Furthermore, the tracking numbers are transformed into Handling Unit (HU) numbers, which correspond to the unique numbers Hilti has assigned to their parcels internally. These HS-codes and HU numbers are then forwarded to Hilti together with their relative timestamps, where they are stored in the ERP system.

The reason for transforming the carriers event codes to HS-codes is to unify the different coding systems used by the carries into one system, where one code corresponds to the same event for all carriers. Without such a system, each carrier should be analysed separately, which is to be avoided for practical reasons. For example, carrier GLS may use event code X to signal the arrival of the parcel at the final hub, and carrier TNT may use event code Y. The transformation system maps both codes to HS-code A. Whenever an analysis is performed, HS-code A indicates the same event for both carriers, saving Hilti the time they would spend on re-doing every analysis for each carrier. Furthermore, it makes the event-logs of the different carriers comparable.

4.3 Phase 3: Enrichment and storage

After the event logs have been transformed to HS-codes and sent to Hilti, they are stored in Hilti's ERP system. Once per day, the carrier data are taken from this ERP system and transferred to the Data lake, a database that combines a multitude of Hilti's datasets. From the Data lake, data can be combined, grouped, filtered and extracted using Structured Query Language (SQL) statements. The SQL statements can also be used to make computations while extracting the data.

As said, the carrier data detail three things: which event occurred, at what time, to which parcel. The Data lake holds many other tables of information, ranging from details on the customer to specifications of materials included in the delivery. By adding the carrier data to the Data lake, the carrier data are enriched by internal Hilti data.

Chapter 5. Analyzing last mile event logs using Process Mining: a literature review

As was mentioned, the branch of Data Science called "Process Mining" is dedicates to extract insights from event logs. Van Beest et al. describes this in his definition of Process Mining: "Process Mining is a family of methods to extract insights from logs of business process executions" (van Beest et al., 2010, p. 386). To understand how the last mile can be improved using the event log it produces, this chapter will review academic literature to see which insights can be gained by using Process Mining methods on last mile event log data, as captured by Research Question 2:

Which insights can be gained by analyzing last mile delivery processes using Process Mining?

5.1 Scope and definitions

To determine the scope of the review, an exploration of event log analysis and Process Mining literature was performed. This exploration showed that no articles describe event log analysis of the last mile process using Process Mining. To address this possible gap in the literature and to answer Research Question 2, this review must therefore first summarise which insights can be obtained from an event log with Process Mining in general, followed by an assessment of how these methods can be used in the context of the last mile. To ensure the quality of the literature in the remainder of this literature review, the following inclusion and exclusion criteria were applied;

Inclusion criteria:

- academic articles listed on Scopus and/or Web of Science,
- articles using event logs in their analysis, and
- articles referring to Process Mining.

Exclusion criteria:

- articles before the year 1990,
- articles with less than one citation (except for articles dating from 2019 or 2020),
- articles to which the researcher had no access,
- articles only mentioning event log analysis in the introduction, and
- articles having any of the concepts as their main topic: mining iron, metal mining, IT-security and Microsoft Windows software

Before continuing, the interpretation of the terms "event log" and "last mile" are clarified; The term "event log" has long been used in the Information Technology (IT) world, predominately in the security system and operating system domain. However, as it became more common to use IT systems to monitor other processes within the firm, event logs now also describe events in business processes. This new definition of the event log in a Business Process Management context is defined by Castellanos et al.:

> "An event log records business events from process-aware information systems (PAIS) such as WFM (Workflow Management), ERP (Enterprise Resource Planning), SCM (Supply Chain Management) and CRM (Customer Relationship Management) systems. Typically, event logs contain information about the start and completion of activities, their ordering, resources which executed them and the process instance they belong to." (Castellanos et al., 2009, p. 488)

For this literature, the definition of the ”last mile” will be taken as described in section 2.1.

> “The last mile is the stretch of parcel delivery service that takes place from the order penetration point to the final consignee´s preferred destination point” based on the definition by Lim, Jin, and Srai (2018).

5.2 General Process Mining insights

5.2.1 Description

Over the last two decades, the branch of data science called Process Mining has been developing rapidly (Berti, van Zelst, & van der Aalst, 2019). Process Mining aims to understand processes based on event data, also called event logs. One of the prime scientific contributors to the field of Process Mining, Wil van der Aalst, provides relevant context in his description of Process Mining “Process mining is a relatively young research discipline that sits between machine learning and data mining on the one hand and process modelling and analysis on the other hand. The idea of Process Mining is to discover, monitor and improve real processes (i.e., not assumed processes) by extracting knowledge from event logs readily available in today’s systems.” (van der Aalst, 2016, p.31). What drives the advantage of Process Mining over pre-Process Mining techniques is that it lets the data speak for itself (van der Aalst, 2012). Rather than using a dataset to test a pre-determined hypothesis, Process Mining takes the data as the point of departure and uses machine learning and data mining methods to describe and improve the process.

5.2.2 Types of Process Mining and insights

The three main types of Process Mining are *Discovery*, *Conformance checking* and *Enhancement* (van der Aalst, 2011), which are discussed below. An overview of these analysis methods and the insights they can bring forth is presented in Figure 8.

<u>Discovery</u>

Discovery takes the event log and distils the as-is process from it. This is useful to get a first overview of the dimensions of the process. An overview can help to clean the dataset from redundant data. If these first steps are taken, Discovery can be used to understand how the different events occur in relation to each other. An understanding of the dynamic between sequential events is not easily obtained without using Process Mining techniques (De Medeiros & van der Aalst, 2008). Furthermore, event logs can contain more information, such as documents, resources used, or people associated with an event. If such information is included in the event log, this information can be combined with event information to track supporting processes.

<u>Conformance checking</u>

Conformance checking needs both an event log and a model of the process that is analysed, as input. Models of the process may be formally documented or exist in the form of an understanding of the process by the people in the organisation. Conformance checking compares such a model/understanding of a process to a process obtained from an event log, showing similarities and differences between the model and reality (van der Aalst, 2011).

Enhancement

The goal of the third type, Enhancement, can be defined as follows: “the extension or improvement of an existing process model using information about the actual process recorded in some event log” (van der Aalst, 2016, p.33). Since Enhancement can improve actual business processes, its potential impact on business performance is regarded as the highest of the three Process Mining categories. Enhancement applies Machine Learning and data mining techniques to the data to show advanced insights and potential improvements.

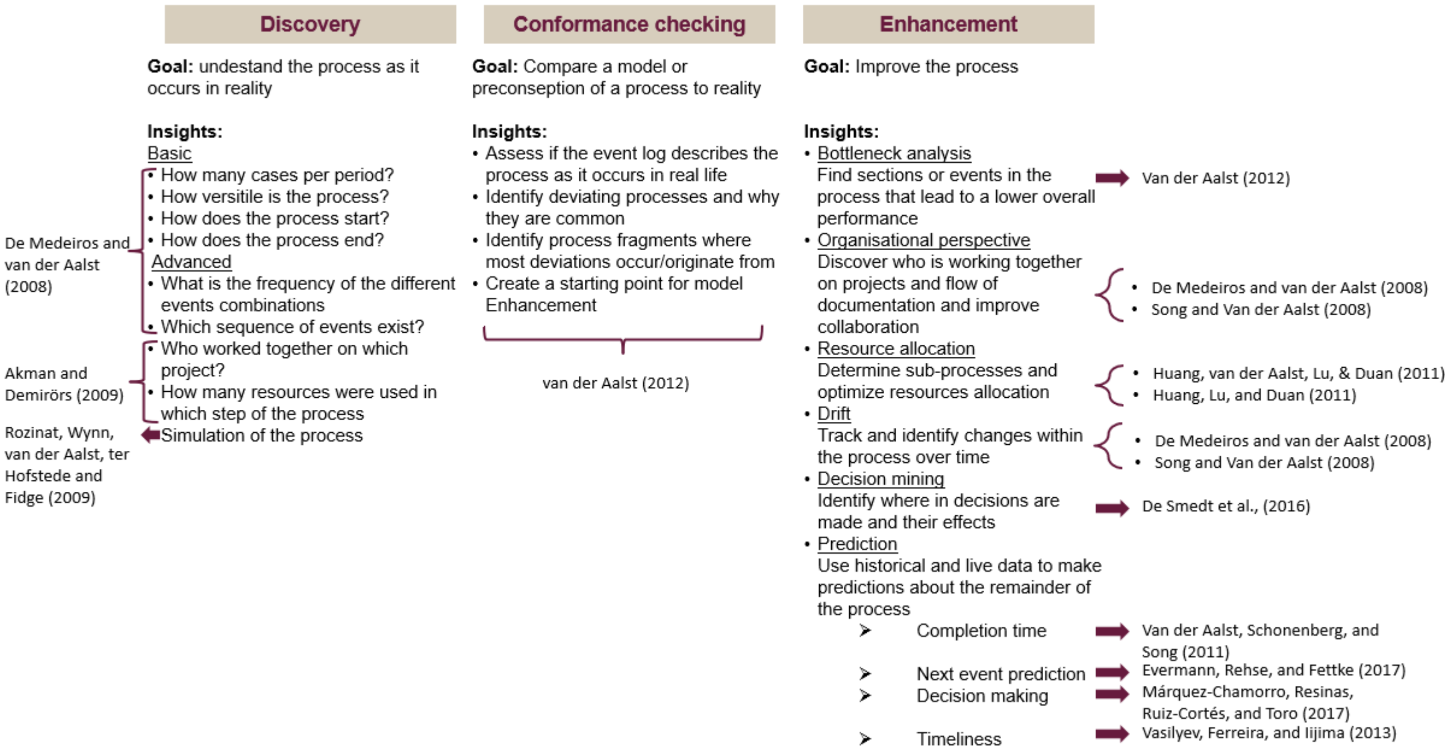

Figure 8. Process Mining techniques and insights

5.3 Process Mining insights in logistics

5.3.1 The relevance of Logistic Management literature

Having determined the general event log analysis methods and insights described by Process Mining, this chapter aims to see how Process Mining methods can be applied to last mile event log data from a literature perspective. No literature was found that describes a clear case of analysing last mile event logs with Process Mining. This raised the suspicion that reviewing literature that specifically describes last mile event logs may be too narrow of a scope. To broaden the scope, the overarching literature domain of *Logistic Management*, of which last mile management is a sub-process, was selected.

Logistic Management is defined as:

> “that part of Supply Chain Management that plans, implements, and controls the efficient, effective forward and reverse flow and storage of goods, services, and related information between the point of origin and the point of consumption in order to meet customer requirements.” (Council of Supply Chain Management Professionals, 2013).

The logistic process described in this definition includes the last mile process as defined in section 2.1. Since the logistic domain seems to be an extension of the last mile domain, it seems reasonable to assume that knowing which insights have been gained by analysing logistic event logs can be used as an approximation of what insight may be gained by analysing last mile event logs. Therefore, this chapter reviews literature on how the Process Mining methods discussed in section 5.2 have been applied to the logistic domain.

5.3.2 Applied methods and insights

According to dos Santos Garcia et al. (2019), logistics is the smallest but yet a distinguished category in Process Mining literature. 27 articles were identified that describe Process Methods being applied to logistic processes. These articles were classified on the Process Mining categories discussed in section 5.2.2. This classification is listed in Table 1. A selection of the articles that are most relevant to the last mile event log and last mile process are discussed below.

Discovery

Gerke et al. (2009) aim to make logistical datasets available for Process Mining by transforming Radio-frequency identification (RFID) data into a Process-Mining-ready dataset. RFID technology is one of the ways in which parcels in the logistic process can be tracked. Their work is continued by and Repta and Stanescu (2017) who describe the challenges of using the RFID data and succeed in creating an overview of a small logistic system in a case study.

Becker and Intoyoad (2017) ascribe the difficulty of analysing event logs of logistic systems to the heterogeneity of the data. They account for this heterogeneity by making the dataset “context-aware” by using Machine Learning in Process Mining for logistic processes. They demonstrate the benefit of including contextual data in the event log dataset in a logistic scenario.

Jokonowo, Sarno, Rochimah, and Priambodo (2019) show a clear use case of Discovery in the logistic domain. They describe that determining the time a container has stayed at a port (Dwell-Time) is often challenging because of the size and complexity of the dataset. By using Discovery, they were able to unravel the dataset and make precise measurements of the Dwell-Time of containers in different processes.

Conformance checking

Wang, Caron, Vanthienen, Huang, and Guo (2014) describe the logistic processes as a highly human-centric process, being both complex and flexible. As such, it is not uncommon for the process to deviate from the process model. The article describes the use of data from a sizable Chinese port, where the authors successfully show deviations from the process model by using Conformance checking.

Li and Deng (2009) emphasize the changing nature of the delivery process and the challenge in keeping the model of that process up to date. They present an algorithm to update the model of an e-commerce delivery process. As discussed in section 2.1, the final step in an e-commerce delivery process is the last mile in a B2C context. Although strictly different then the last mile in a B2B context, we can state that this article is a good example of event log analysis in the context of the last mile.

Enhancement

As mentioned, Enhancement is challenging but also brings the most high-value insights. Kecman and Goverde (2013) first apply Discovery and then Enhancement (*Prediction*) techniques to predict times of the next event. This by using an event log produced by a train network.

Vasyutynskyy, Gellrich, Kabitzsch, and Wustmann (2010) describe the analysis of internal logistic systems. By using Discovery and Enhancement, they analyse a baggage handling system. They indicate that simply simulating the system with Discovery methods does not fully counter the complexity and diversity of the process. They then continue to describe a generic approach to Enhancing internal logistic systems.

Dunkl, Rinderle-Ma, Grossmann, and Fröschl (2014) focus on Enhancement's *Decision mining*. They develop a system that adds additional data to the standard event log and use this combined dataset to find attributes that are decisive for the branching of process paths within a Discovered model. Finally, they apply this method to a container transportation scenario, where they show that their model can indicate where in the process a decision is being made and based on which attribute. Lau, Ho, Zhao, and Chung (2009) developed an algorithm that specifically targets large datasets of logistical data. Combining Process Mining and fuzzy-logic, they are able to determine which factors in the supply-chain of a case study contribute to higher customer satisfaction.

Table 1. List of Process Mining studies in a logistical context

Process Mining Category	**Sub-category**	**Reference**
Discovery	Basic	Becker and Intoyoad (2017)
		Janssenswillen, Depaire, and Verboven (2017)
		Gou et al. (2016)
		Song, Jacobsen, Ye, and Ma (2016)
		Pulshashi, Bae, Sutrisnowati, Yahya, and Park (2015)
		Sutrisnowati et al. (2015)
		Sutrisnowati, Bae, Park, and Ha (2013)
		Krathu, Pichler, Zapletal, and Werthner (2012)
		Rozsnyai, Lakshmanan, Muthusamy, Khalaf, and Duftler (2012)
		Haigh and Yaman (2011)
		Gerke et al. (2009)
		Gonzalez, Han, and Li (2006)
		Soffer et al. (2019)
		Repta and Stanescu (2017)
		Jokonowo, Sarno, Rochimah, and Priambodo (2019)
	Advanced	Besri and Boulmakoul (2017)
Conformance checking	General	Wang et al. (2014)
		Li and Deng (2009)
Enhancement	Bottleneck analysis	Vasyutynskyy et al. (2010)
		Lau, Ho, Zhao, and Chung (2009)
	Organisational perspective	Wang, Zhu, Wang, and Huang (2016)
	Resource allocation	Zhao, Zeng, Zheng, and Yang (2017)
	Drift	-
	Decision mining	Dunkl et al. (2014)
	Prediction	Kecman and Goverde (2013)
	Multi-dimensional	Sutrisnowati, Yahya, Bae, Pulshashi, and Adi (2017)
		Cai et al. (2018)
Supporting areas	Gathering and cleaning data	Kumar, Thomas, and Annappa (2017) Van Cruchten and Weigand (2018)

5.4 Answer to Research Question 2

The goal of this literature review is to answer Research Question 2:

Which insights can be gained by analyzing last mile delivery processes using Process Mining?

Section 5.2 showed that there are three main categories of Process Mining; Discovery, Conformance checking and Enhancement. Since literature describing the application of Process Mining to the last mile was lacking, the scope was broadened by reviewing the application of Process Mining to the overarching domain of logistics in section 5.3. This chapter takes the results from section 5.3 and tries to estimate what the application of Process Mining to the logistic domain can say about gaining insights from last mile event logs.

The distribution of Process Mining methods that have been applied to the logistic domain are shown in Table 1 and may give an indication of which methods can gain insight in the last mile. More than half of the articles apply Discovery to logistic processes. Often mentioned reasons for using Discovery on logistic event logs are the complexity, diversity and size of the logistical dataset. To understand the logistic process of a delivery process tracked with RFID technology, for example, requires applying Discovery combined with pre-processing of the data. Since the last mile is a sub-process of the delivery process, the two processes should be comparable. It is therefore expected that using Discovery for analysing last mile event logs should be possible and result in gaining a better understanding and overview of the last mile.

Conformance checking was also successfully applied to logistic event logs. The two articles that use Conformance checking (Wang et al., 2014; Li & Deng, 2009) state that logistic processes are constantly changing, quickly outdating the model of a process. To understand if the current model of a process is still accurate, both papers successfully apply Conformance checking. Since the case study described by Li & Deng (an e-commerce delivery case) is particularly close to a last mile delivery process, it is expected that Conformance checking can assess the correctness of a process model for the last mile.

The articles that describe the use of Enhancement methods to a logistical event log have in common that the analysis does not typically start with Enhancement. Rather, first a Discovery analysis and/or Conformance checking is applied. This simply to create a good understanding of the process before improving it through Enhancement. Five of the six sub-categories of Enhancement (*Organisational perspective*, *Bottleneck analysis*, *Resource allocation*, Decision mining and Prediction) were found to be applied to logistic processes in the literature. There seems to be no clear reason not to apply the remaining sub-category, *Drift,* to the logistic domain or the last mile. The potential insights these six sub-categories of Enhancement can bring forth is great, as they have the potential to improve the process as is shown by Vasyutynskyy et al. (2010) and Lau et al. (2009). Some logistical processes included in the Enhancement case studies are not directly comparable to the last mile, such as a train network (Kecman & Goverde, 2013) or an internal logistic system (Vasyutynskyy et al., 2010), which makes it difficult to assess if some methods can be applied to the last mile. However, Bottleneck analysis, Decision mining and Prediction are implemented in a process closely resembling the last mile, as shown by Lau et al. (2009) and Dunkl et al. (2014).

Finally, it is worth mentioning that recent applications of Process Mining to logistics have taken a *Multi-dimensional* approach (Kumar, Thomas, & Annappa, 2017; Van Cruchten & Weigand, 2018). This approach entails combining multiple Process Mining techniques to answer a question. The main reason for using this Multi-dimensional approach is the complexity of the question, analysed process and/or the dataset. Since the last mile process is said to be a complex and dynamic process as well, taking a Multi-dimensional approach to answering difficult process questions may prove fruitful.

Chapter 6. Experimentation with a transformed event log

Chapter 5 answers Research Question 2 by showing which insights can be gained by analyzing last mile delivery processes using Process Mining Discovery, Conformance checking and Enhancement. This chapter aims to see which of these methods can be executed with a transformed external dataset in a practical setting, by answering Research Question 3:

Which Process Mining methods can be executed with a transformed external last mile event log?

As far as the writer is aware, no attempt has been described in the literature where an event log of an external last mile process has been analysed. Moreover, the effects of the transformation from the original carrier events to the HS-code system, as described in Chapter 4, are largely unknown. It is therefore uncertain how suitable the available dataset is for analysis. As such, the Research Question will be answered through a practical experiment using a transformed dataset (i.e., containing HS-codes rather than original events) of one of Hilti's external carriers. Because of the experimental nature of the chapter and the time constrain of this thesis, the results should be seen as a first indication of how external carrier event log data can be analysed.

6.1 Approach

The literature review in Chapter 5 paints a clear picture of how standard event log data are commonly analysed, namely with Process Mining. As illustrated, this branch of data science takes an event log as input and returns a model that helps to understand, and potentially improve, a process. Chapter 5 also showed that the three main categories of Process Mining are often applied in sequence, meaning that first Discovery methods are applied, followed by Conformance checking and finally Enhancement. Such an order allows the researcher to understand the process before an attempt is made to improve it. Therefore, this experiment chapter follows the steps of Process Mining as described in the literature in Chapter 5, starting with Discovery, followed by Conformance checking and finally attempting Enhancement.

Before attempting the three Process Mining methods, a dataset is retrieved from the Data lake and prepared for the experiment. This step is vital as the dataset is the primary input for the analysis and heavily impacts the application of Process Mining methods.

The Process Mining tool used is the commercial tool PAFnow. The tool is an extension of the business analytics tool Power BI. This tool was made available for Hilti under a try-out license for a number of months to perform the experiment. The tools used for the exploration are PAFnow Premium Report (version 2020.3.1), the PAFnow Companion Setup (version 1.1.5) and Microsoft Power BI (desktop version 2.84.701.0).

6.2 Preparation of the dataset

This section introduces the structure of a dataset that is used for Process Mining analysis; the event log. The section shows that Hilti's transformed event log holds the right information to attempt Process Mining analysis. Finally, the properties of the dataset used for the experiment are determined.

6.2.1 Event log structure

Figure 9 describes the basic relational model of an event log. A standard event log contains multiple cases within one process, where each case has one or multiple activities occurring to them. The occurrence of an activity is recorded in the event log as an event. Each event is described by attributes, which detail information about the event, such as the activity.

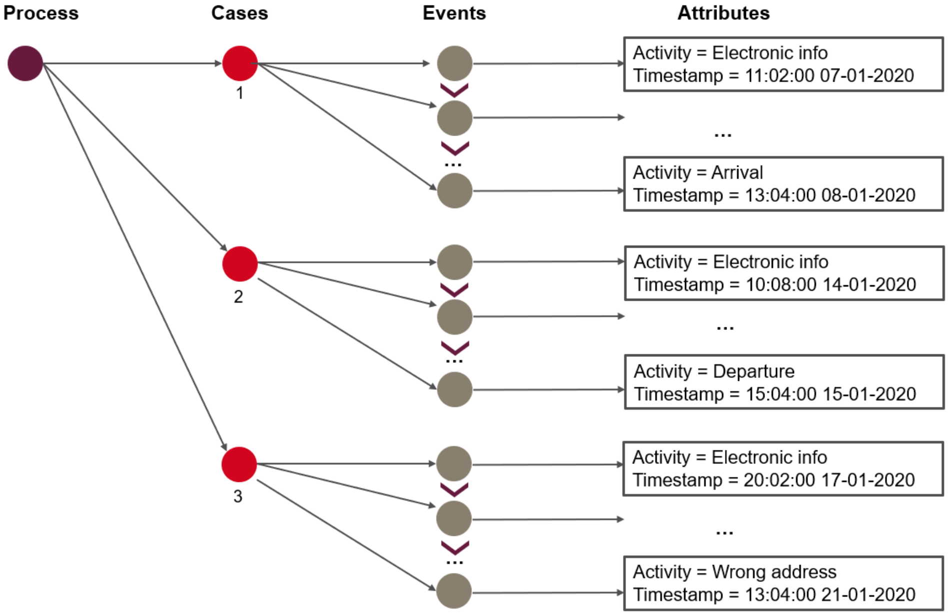

Figure 9. Event log structure Note*;* Adapted from "*Process Mining*", by Van der Aalst, W 2002, p. 130, Copyright Springer-Verlag Berlin Heidelberg 2011

An event log is a dataset containing a sequence of all the events that occurred in a process to the selected cases, as depicted in Figure 10. Information related to events and their respective cases is stored in additional columns in the event log. Case attributes are pieces of information that are true for the whole case (e.g., type of delivery) where event attributes are true for one event (e.g., time of an activity). Every row in an event log describes a new event and provides information related to that event and/or its case.

6.2.2 Meeting the Process Mining requirements with Hilti's event log

At the very least, an event log should specify three data points for each event:

1. a column with the Case ID of the unique object being tracked,
2. a column detailing the Activities occurring to that object, and
3. a column with the Timestamps of the events ("Data Requirements", n.d.).

As described in Chapter 4, Hilti´s event log contains these three data points. Consider a snippet of Hilti´s event log depicted in Figure 10. Each row of the table stands for an event. The parcel that is being shipped can be taken as the unique objects to be tracked. This unique object can be identified with its unique Handling Unit (HU) number. As such, the column *HU_Number* is taken as the case ID, satisfying the first requirement. As explained in Chapter 4, the HS-codes indicate the activities occurring in the last mile. Therefore, the HS-codes detailed in the column *Event_HS_Code* will be taken as the Activity column, satisfying the second requirement. Finally, every HS-code has its corresponding *Timestamp*, meeting the third basic requirement of an event Log suited for Process Mining.

HU_Number	Event_HS_Cod	Timestamp	Global_Service	Plant	Truckload_Time	Carrier_Depot_	ship-to_Party
198294767	HS102	10/25/2019 12:11	STD	8150	21:00.0	DEGLS-89	10242269
198294767	HS146	10/25/2019 16:43	STD	8150	21:00.0	DEGLS-89	10242269
198294767	HS148	10/28/2019 6:18	STD	8150	21:00.0	DEGLS-89	10242269
198294767	HS160	10/28/2019 6:22	STD	8150	21:00.0	DEGLS-89	10242269
198294767	HS201	10/28/2019 12:07	STD	8150	21:00.0	DEGLS-89	10242269
198294767	HS000	10/29/2019 11:38	STD	8150	21:00.0	DEGLS-89	10242269
198377811	HS000	10/28/2019 11:44	STD	8150	53:05.0	DEGLS-89	26557552
198377811	HS102	10/28/2019 14:04	STD	8150	53:05.0	DEGLS-89	26557552

Figure 10. Hilti´s Last mile event log snippet

On top of the minimum data requirements described above, Van der Aalst (2016) has listed secondary requirements which should be met by an event log suitable for Process Mining. If and how these secondary requirements are met by Hilti's transformed event log is described below.

1. Correlation

The event log should be groupable per case. In the case of the carrier GLS in Germany, Hilti has linked the HU number to every occurring event. As such, the events can easily be grouped on this HU number during data extraction, meeting the requirement.

2. Timestamps

Within a case, the events should be ordered by the sequence in which they occurred. Every event recorded in Hilti´s event log has a timestamp recorded to it. As such, it is straight forward to put the events per case in order of occurrence by ordering the events grouped per case on timestamp during the extraction of the data from the Data lake.

3. Snapshots

To manage the total volume of data, it is common to select a snippet of the available data consisting of events over a limited period. As a result, the problem may arise that some events fall outside of this 'snapshot', cutting off processes of one or more cases halfway. To avoid this, the time constraint is put on the starting date of the delivery, and not on the date of the events themselves. This way, the complete event log of cases starting in the selected period is selected, even if some events occurred after the end of the selected period.

4. Scoping

The issue of scoping describes the dilemma of how much and which data to include. The process of selecting a subset of the event log is described in section 6.2.3.

5. Granularity

Granularity describes the availability of data at different levels of detail. This can be challenging if one tries to merge different sets of data and combine them to one event log and attributes. Within the dataset of GLS in Germany, this is not the case, as all the data can naturally be linked to the HU number. Higher levels of granularity are recorded in event attributes and can be used in later stages to analyse the event log on higher levels.

In summary of this section, it can be concluded that at the core, the data available to Hilti contain the right information and meets the requirements for Process Mining analysis. In the next section, a dataset is selected for the experiment.

6.2.3 Scoping

The last mile delivery data available in Hilti's Data lake hold over 15 years of event log data and can be linked to many internal datasets, increasing the potentially relevant number of attributes. Such a dataset is too large and diverse for this experiment, which is why a demarcated dataset will be selected.

As can be seen in Figure 10, the total size of an event log is dependent on two factors:

1. which cases are tracked, affecting the number of rows, and
2. which information describing the events and cases are included (attributes), affecting the number of columns.

Case selection

To limit the number of cases and define clear boundaries of the dataset, the following constraints were applied. The;

1. delivery direction is Outbound,
2. delivery service is Standard Delivery,
3. deliveries are executed within Germany (no border crossing),
4. deliveries are executed by one carrier, namely GLS, and
5. start of the delivery lies between 05-01-2020 and 05-02-2020 (one month of data).

Constraint 1 and 2 ensured that the outbound delivery process was selected, which is the equivalent of the standard last mile delivery process within Hilti's data frame.

Constraint 3 was chosen to only have the process from one country. Analysing data from multiple countries was avoided as different countries have implemented the HS-code system slightly different, potentially distorting results. Here, Germany was chosen because they are known to have the biggest dataset due to many daily deliveries that have been recorded over a long period.

Constraint 4 was selected for a similar reason as constraint 3, needing only one carrier's data to ensure consistent use of the HS-code concept.

Constraint 5 was based on selecting recent data of a workable size. The carrier's version of the data, before transformation to the HS-code system, is available 3 months after the occurrence of the event. Therefore, recent data were selected, so that possible uncertainties in the interpretation of the data could be checked. As the analysis was executed in March 2020, data at the beginning of 2020 were selected. As the first few days of the new year tend to have deviating processes, the starting date was set at January the 5th, 2020.

Attribute selection

Besides the three "bare minimum" columns described in section 6.2.2, case and event attributes can be added to an event log detailing extra information of a case or event. Hilti's Data lake pools data from many different sources together, making the potential number attributes to include very large. Which event and case attributes to include was decided during discussions with Hilti's managers and by experimenting with the data. Leading here was 1) which attributes were known to be accurate, 2) which information was relevant for the last mile delivery process and 3) which attribute may be of interest to analyse in a later stage. The selected case attributes are listed in Table 2 and the selected event attributes in

Table 3 in Appendix D.

This selection of cases and attributes returned 654.993 rows of events with 18 columns of data. A dataset much too large for a program like excel, but small enough for the Process Mining tool PAFnow.

6.2.4 Data description

The selected event log contains all standard, outbound deliveries within Germany executed by carrier GLS during the period between 05-01-2020 and 05-02-2020. This returned a total of 102.034 cases (i.e., parcels) to which 654.993 events occurred, making the average number of events per case slightly more than 6.4. The average lead time (i.e., the average time a case took from start to finish,) is 36 hours and 26 minutes and 30 seconds. The distribution of cases and their respective lead times in days is displayed in Figure 11. Furthermore, 3509 different process variants were found in the event log. The 10 most common processes account for 83.345% of the cases, indicating many variants only occur a few times. Furthermore, 32.94% of cases were found to have a (self) loop i.e., the same activity occurring twice in one process. Finally, 22.37% of cases occurred over the weekend, meaning that either a Saturday, Sunday or both occurred during the delivery process. For the remainder of this thesis, the term 'event log' will be used interchangeably with the term 'dataset'.

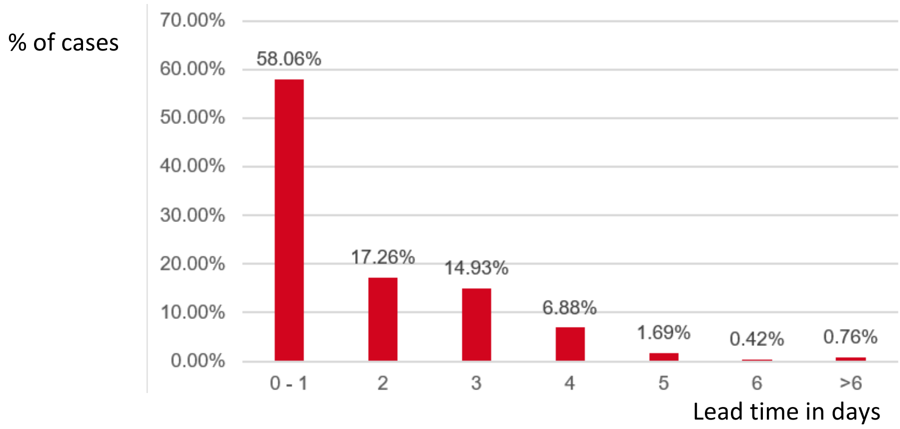

Figure 11. % of cases per total lead time in days

6.3 Discovery

Having selected a dataset, it can be seen for which Process Mining analysis the dataset is suitable, starting with Discovery. The goal of Discovery is to understand the process, as it occurs in real life, through the event log produced by the process. In this experiment, the goal of Discovery is slightly altered as a result of the transformed data that are used as input.

Consider Figure 12, detailing the steps involved in understanding the real-life process, using the (transformed) event log and process models. Normally, the event log is created during the process (Step 1), after which the process detailed in that event log is extracted using Discovery (Step 2 option a). In this thesis, however, this event log is first transformed (Step 2 option b) and then Discovery is applied (Step 3). Using Discovery to analyse the original event log returns a model of the process that is usually close to real-life. In this experiment, however, we are interested in insights gained by analysing a transformed event log. In our scenario, the transformation of the event log may have altered the data and therefore describes a different process. Discovery by itself is not able to value if the process model is true to the real-life process but will simply return whichever process is present in the data. Therefore, this section performs Discovery experiments with the dataset to create a model that helps to understand the process captured in the dataset, which may or may not be the real-life process. In section 6.4, Hilti´s understanding of the carrier´s process, which is based on Hilti's experience and collaboration with the carrier, is compared to the model discovered from the transformed event log (Step 4). Using this comparison, it is possible to understand how the model extracted from the transformed event log by Discovery in Step 3 represents the real-life process. Put differently, this thesis needs to use both Discovery and Conformance checking to deduce if and how the transformed event log represents the real-life process.

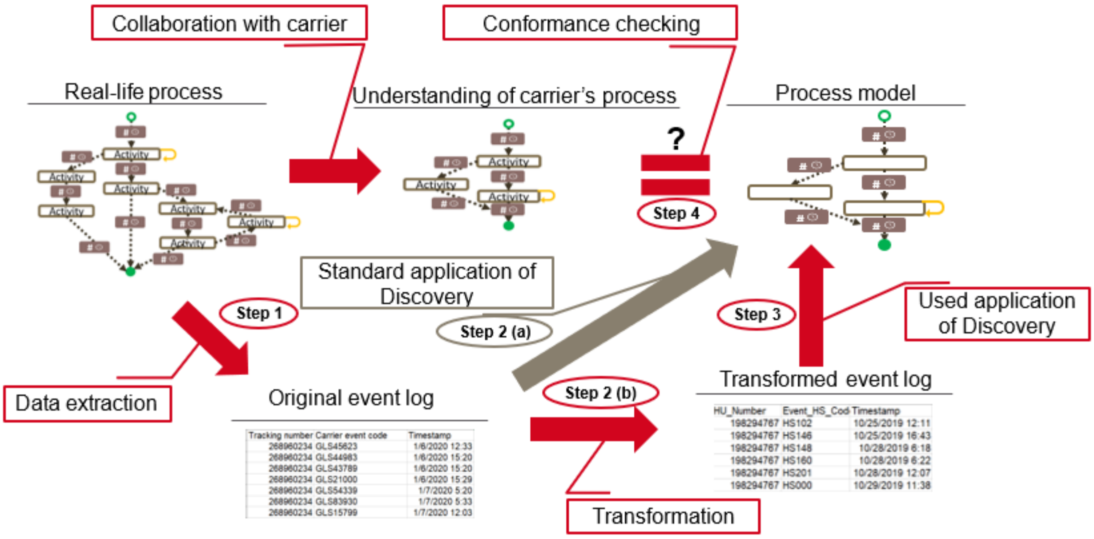

Figure 12. Understanding the carrier's process using Discovery and Conformance checking

6.3.1 Experiment 1: Creation of the first process model

<u>Model creation</u>

In this section, the main objective is to create a model that helps understand the process described in the transformed data. The first model was created from the dataset described in section 6.2. During the extraction of the dataset from Hilti's Data lake, the data were ordered in the structure of an event log as determined by Van der Aalst (2016). This entailed grouping the events per case and sorting them in order of occurrence. The ordered dataset was then imported into the PAFnow Companion, where parameters for the final model were set. This meant manually selecting which columns of the dataset indicated the Case ID, Activity Name, Timestamp, Event attributes and Case attributes, telling the program how to build the model. The PAFnow Premium Report was then used to display the discovered process. Figure 13 contains the complete process by showing the activities for all the cases in the dataset.

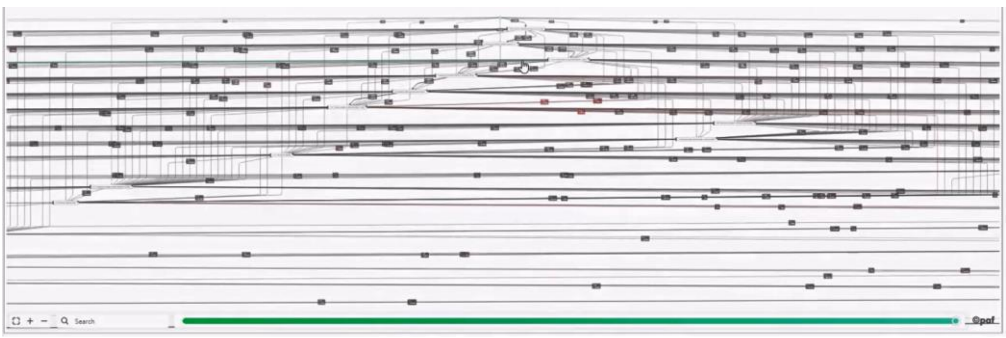

Figure 13. Complete process

The model in Figure 13 is very elaborate and hard to interpret due to a high number of process variants in the dataset. Such diversity of the process is not uncommon but does make interpreting the process challenging. A more insightful representation of the discovered process is “the most common process variant”, displayed in Figure 14. The most common process variant is the path of activities which is taken by most of the cases and is often interpreted as the standard process.

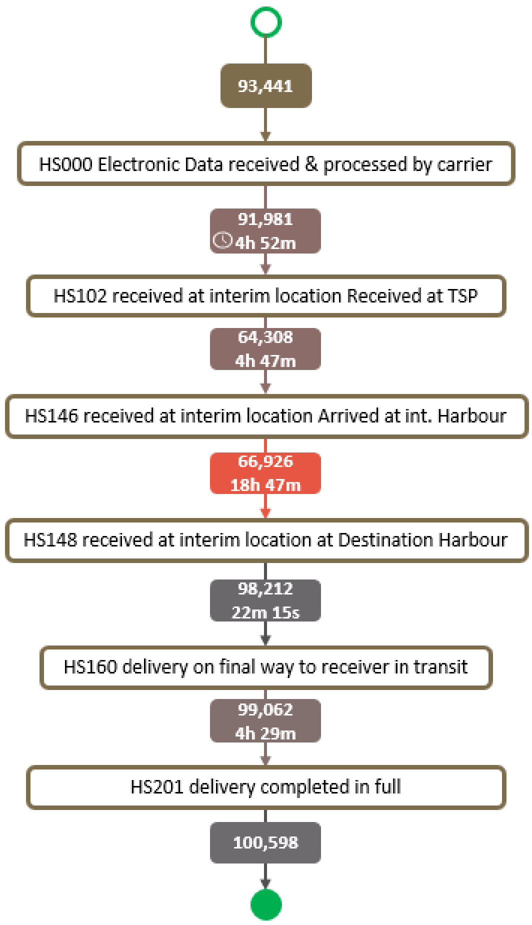

Figure 14. Most common process

Figure 14 can be understood as follows: the process starts at the open green circle at the top of the figure after which multiple activities occur in sequence until the end of the process, indicated with the closed green circle at the bottom. The large and open rectangles represent the activities within the process (i.e., the HS-codes and their descriptions). The arrows in between the activities indicate the order in which the activities occurred (i.e., the flow of activities). The small and solid rectangles positioned on the arrows between the activities indicate the number of cases that flow from one activity to the next and the average time between the two activities. The colour of the connection between the activities represents the relative time between the activities, making it easier to spot sections of the process that take a relatively long time. For example, 66.926 cases flow from activity “HS146 received at interim location Arrived at int. Harbour” to activity “HS148 received at interim location at Destination Harbour”, in 18 hours and 47 minutes on average. The red colour indicates that the average time between these two activities is relatively long when compared to the average time between other activities.

Figure 14 shows the most common process variant, only displaying the most frequent flow of activities. In contrast, Figure 13 shows all activities and flows in the dataset. Both figures display the same process, but simply show a different number of variants. The Process Mining tool used for the experiment allows exploring this spectrum of displayed variants, as illustrated in Figure 15. Exploration of this spectrum aids in understanding (less) common flows and in determining (self) loops, sections of the process that are repeated within the same case.

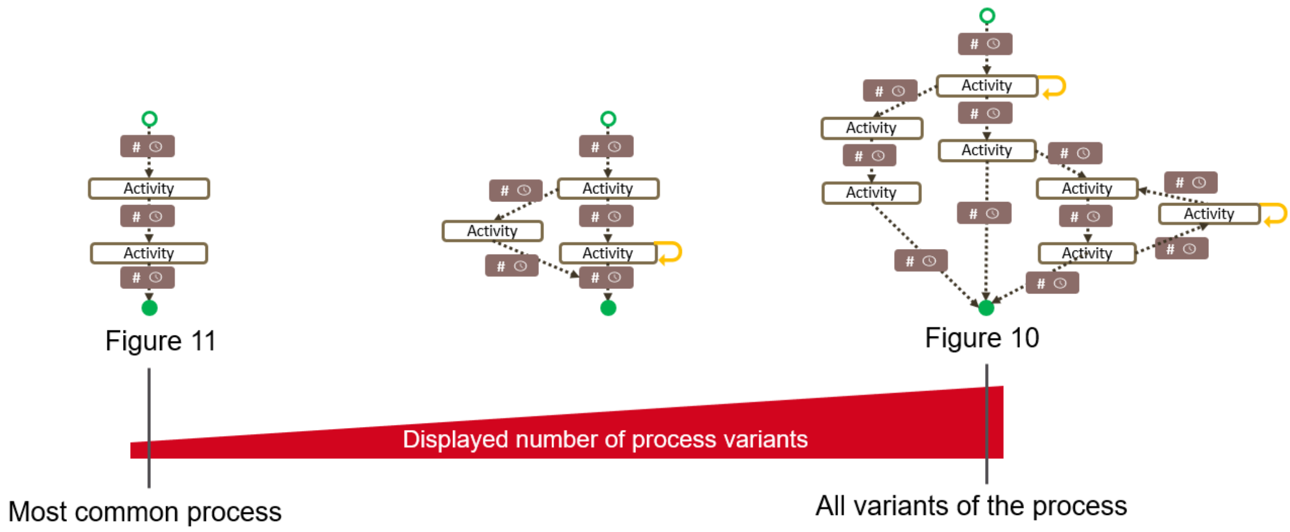

Figure 15. Spectrum of complexity

Insights

The following insights were gained by this first Process Mining Discovery experiment:

- The most common six activities and order of activities (depicted in Figure 14).
- The most time in the process is spent between arrival at an interim hub (HS146) and arrival at the final hub (HS148).
- Semi-common processes often include a repetition of arriving at the interim hub (HS146), arrival at the final hub (HS148) or both.
- Most cases start with the exchange of electronic data (HS000) and end with successful delivery at the customer (HS201), meaning that most variety of the process occurs between these two activities
- There is a large number of different process variants (3509) present in the analysed event log.

It should be noted that these specific uses of the HS-codes only apply to the transformed dataset of GLS Germany, and do not represent how the HS-code system is implemented at other carriers. However, the following general conclusions can be drawn from this first experiment:

- It is possible to use Process Mining Discover on a transformed event log dataset.
- In principle, the same types of insights that can be gained from a standard event log can be gained from a transformed event log. Insights such as start and end activities, common activities, common process variants and (self) loops.
- The level of detail present in the HS-codes results in having a lot of different process variants, which makes it challenging to understand the full process model as well as less common process variants.

6.3.2 Experiment 2: Simplifying the model

The process model created in Experiment 1 was difficult to interpret because displaying more process variants quickly resulted in an unreadable and complex process model. The most probable reason for this complexity is that many activities include a level of detail which is not relevant for each analysis. For example, the HS-codes “HS331 Customer not available - no message left” and “HS333 Customer not available - message left”, have a very similar meaning. To check if the order of activities is logical, as will be done in the Conformance checking section, one does not need to know whether a message was left or not. If similar activities such as HS331 and HS333 discussed above would be interpreted as being identical, the total number of different process variants could be drastically reduced, making the process model less complex and more straight forward to interpret.

To cluster similar activities but retain information by which Conformance checking can determine if the order of the activities is logical, event attributes were sought which could categorize similar activities. Two event attributes already present in the selected dataset were selected to achieve this. The first event attribute, called *HSCodeMoment*, describes the process stage the activity refers to. Possible stages of the process ordered from beginning to the end of the process are: “Pre Pickup”, “InTransit” and “Drop”. Every activity included in the HS-code system belongs to one of these stages. The second event attribute that was used, called *HSCodeType*, describes how the activity should be interpreted, i.e., if the activity signals Success, Error or if the activity is Informational. Similar to the HSCodeMoment event attribute, each activity is classified as one of the HSCodeTypes.

By combining these two event attributes, a new attribute, *HSCodeMoment&Type* was created and each event was categorized on this attribute. As such, each activity is classified as belonging to a stage of the process and type of message. HSCodeMoment&Type clusters multiple activities together, reducing the total number of activities while simultaneously retaining the information by which Conformance checking can see if the order in which the activities are received. For example, if an activity classified as “InTransit & Success” is followed by an activity “Pre Pickup & failure”, we know that the order in which these activities are received is illogical, without having to know which activities were received. Similarly, the classification of HSCodeMoment&Type allows us to see if errors have been resolved. For example, if an event belonging to the class “In Transit & Error” is received, indicating an error during transportation, receiving an activity from the class “In Transit & Success”, indicates that error has been solved.

The scheme described above can be used by Discovery to make the process in the transformed event log easier to understand, as is the goal of Discovery. To see if such assessments could be made by using Process Mining, a new model was created with HSCodeMoment&Type set as activity name instead of the HS-codes (i.e., unclassified events). The combined event attribute was computed during the extraction of the data from the Data lake and set as the activity column when constructing the Process Mining Discovery model. The results are depicted in Figure 16.

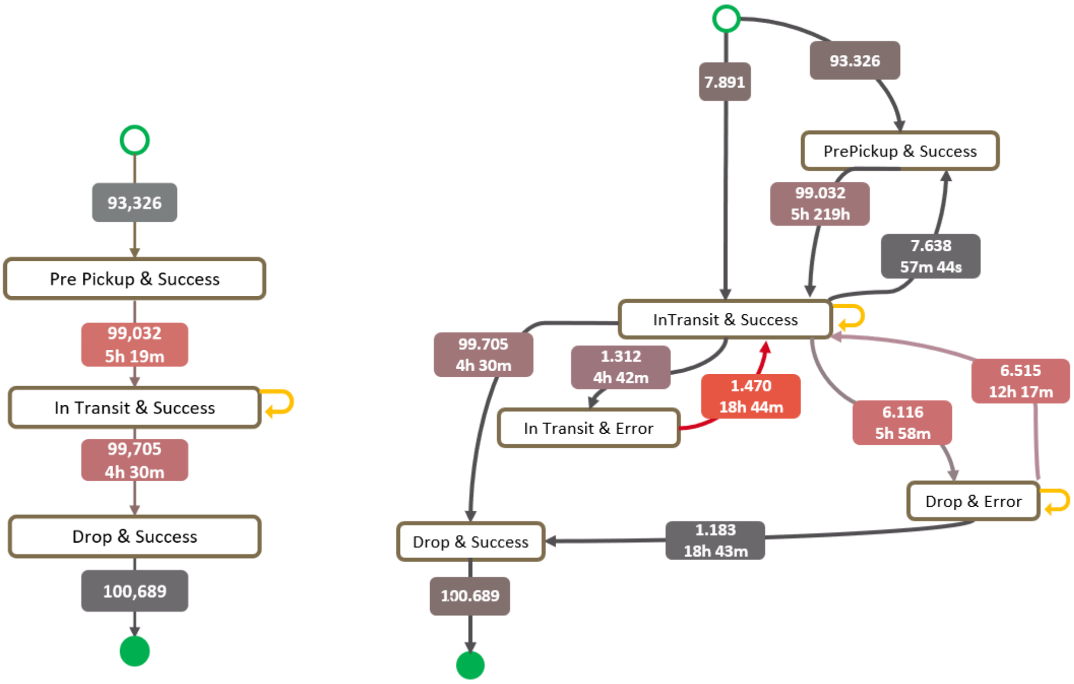

Figure 16. Discovered processes using classified events, most common variant (left) and extended process (right)

By the process of aggregating multiple events into categories and taking these new categories as activities as described above, the total number of activities process variants was indeed drastically reduced from 3509 to 1902 process variants. When one expends the simple model presented in Figure 16 by allowing for more process variants, the increase in complexity is far less rapid and distorting, and the extended process model is produced.

The goal of this Process Mining Discovery experiment is to understand the process that is described in the transformed event log. Experiment 1 achieved this by gaining specific insights into the process and Experiment 2 contributed to comprehending the process by reducing its complexity. Now that the event log data have been transformed into an interpretable process, we can attempt to estimate what our process model can tell about the carrier´s process using Conformance checking.

6.4 Conformance checking

The goal of Conformance checking is to understand similarities and differences between a discovered process and a model of that process. To see if Conformance checking can be used on the selected dataset, interest is in comparing the discovered process to Hilti´s understanding of the carriers process, as depicted in Figure 12. The first element of this comparison, the process discovered from the transformed event log, was produced in section 6.3.2 with Experiment 2. For the second element, a model of this process, Hilti's understanding of the carrier's process is used. This understanding has arisen by close collaboration with the carrier over the last years, aided by correspondence with the carrier during the experiment to verify assumptions about the last mile process. The result is a general understanding of the carriers last mile delivery process. A full model of the carriers process could not be obtained, but as this experiment aims to see if a comparison is possible, a general understanding of the model should be sufficient.

First, it is seen how the discovered process resembles Hilti's understanding of the carrier process. Secondly, a small case study is discussed to see if Conformance checking can already help (dis)confirm statements in the reports of the carrier with regards to on-time delivery.

6.4.1 Conformance between discovered process and understanding of carrier process

By using the models created in the Discovery section, we can see if the discovered process seems to correspond to the current understanding of the carrier's process. Firstly, this is done by checking if the detailed discovered process simply "makes sense" i.e., are the activities, the order in which they occur and their relative intervals similar to what one would expect of the last-mile delivery process. Reviewing the detailed most common process in Figure 14, we may note the following points:

- Hilti management confirmed that in principle, the order of activities as they are portrayed in Figure 14 is the logical order in which the activities should occur. The delivery process does indeed start with electronic transmission of delivery data and ends with delivery at the customer. Furthermore, the arrival of the parcel at multiple hubs in the middle of the process is also in line with the common perception of the carrier delivery process. Since Figure 14 is the most occurring variant of the process this indicates that, in principle, the transformed event log data display a logical order of events.
- The total average lead time being 36 hours and 26 minutes seems to be a realistic unit of time for an average delivery, albeit somewhat on the short side for standard deliveries.
- The part of the process between departure from the final hub (HS160) and arrival at the customer (HS201) indicates that the HS-code system interpreted the carrier codes properly and shows that the discovered model can capture real-life processes. The time of 4 hours and 29 minutes between activity departure and arrival (HS160 and HS201), indicating the average time between the final hub and the customer resonates with common lead times in this step according to management. This part of the discovered model is therefore an initial confirmation of the transformed event log accurately catching part of the delivery process.
- However, there are parts of the process that seem less accurate:
  - The event log shows that the parcel has arrived at harbours (HS146 and HS148), even though it is known that within Germany these goods are transported over land and should be indicated with other HS-codes.
  - When slightly expanding the model i.e, allowing for more connections between the activities, the process showed many "self-loops" for arrival at the interim hub (HS146) and arrival at the final hub (HS148). This is an indication that these codes were sent multiple times in a row. For arrival at the interim hub, this is acceptable, as there can be multiple interim hubs in the process. However, one would expect only one arrival at the final hub. This raises the suspicion that the many-to-one relationship between the carrier codes and the HS-codes indicating arrival at the final hub (HS148) may create in-accurate duplicate activities. To determine the true cause of these self-loops, the original carrier data should be available. This could clarify if these self-loops occurred in reality, if the original carrier data were inaccurate or if the self-loops were created by a many-to-one relation between the carrier events and the HS-code system. These original data are however unavailable in large quantities at present.

It should be noted that these specific uses of the HS-codes only apply to GLS Germany, and do not represent how the HS-code system is implemented at other carriers. However, the following general conclusions can be drawn from this first Conformance check:

1. The most common process variant of the discovered process generally corresponds with Hilti's understanding of the real-life process.
2. Some parts of the process were difficult to interpret, most probably due to the transformation from carrier event codes to the HS-code system.
3. Having a general knowledge of the carrier´s process helps to see if the discovered process is logical in a last mile context, but does not allow detailed analysis of where the process model and the process in reality differ.

As stated above, Conformance checking cannot fully verify the transformed dataset because the exact process of the carrier is unknown. To demonstrate its use, a small section of the last mile is selected which is better known to Hilti management and combined with the model from Experiment 2.

6.4.2 Experiment 3: A case study

This case study aims to see if Conformance checking can be applied to the transformed dataset. This is done by checking if a sub-process of the carrier's last-mile delivery process is behaving as expected by Hilti management i.e., comparing a (sub)process model to the discovered (sub)process.

For Hilti, it is vital to know if a parcel arrived on-time at the customer, to:

- know which customer received their parcels on-time,
- determine the performance of the carrier, and
- find potential improvements of the process.

For many years, the carrier was responsible to report which parcels arrived on-time and which ones arrived late at the customer in so-called "carrier reports". More recently, Hilti started to verify these reports by basing calculations on the carrier data. As discussed in section 1.3.1, the true on-time delivery rate is currently ambiguous due to a discrepancy of carrier reports and initial computations based on transformed event log data. Consider Figure 17, which displays a theoretical model of the last mile delivery process divided into process stages, as introduced in Experiment 2. At the last step of the process, when the parcel arrives at the customer, Hilti determines whether the parcel arrived within the agreed-upon time window. This assessment is based on comparing the timestamp of the last activity belonging to the category "Drop & Success" (when the parcel arrives at the customer) to the delivery deadline.

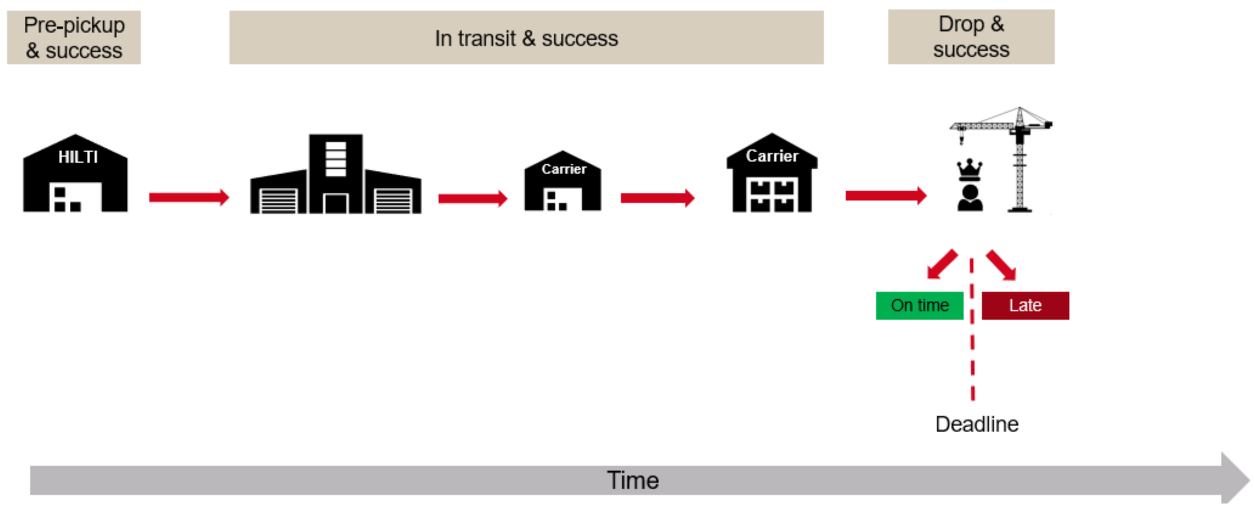

Figure 17. Hypothetical process model

However, when reviewing the process model created in Experiment 2, process variants were spotted where the activity "Drop & Success" was not the last activity received. This raised the question of what activities did occur after Drop & Success and what influence this has on the on-time status of the parcels in question (see Figure 18), which will be the topic of this case study.

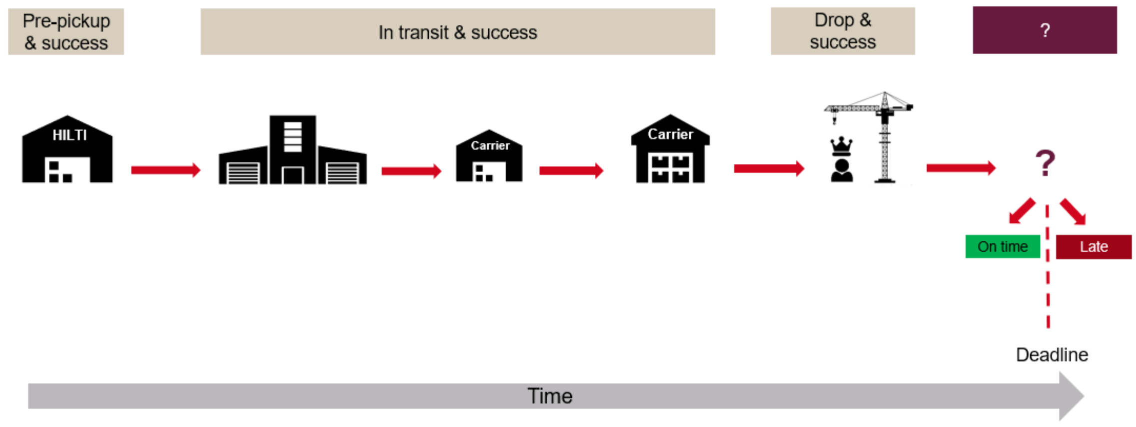

Figure 18. Case study of pattern found in the event log

Model preparation

A model was prepared to see which activities occurred after "Drop & Success". As we are interested in activities occurring after any activity belonging to the "Drop & Success" category, the process model created in Experiment 2 was needed because the activities belonging to the "Drop & Success" stage are grouped within this model. If we were to use the process discovered in Experiment 1, it would be more challenging to select the right sequence of activities.

When selecting the data, the following cases were of interest:

1. cases where a "Drop & Success" has occurred at one point in the process, and
2. cases where another activity has occurred after "Drop & Success", which was not followed by a second "Drop & Success".

These cases were selected by using the Process Mining tool´s filter function, which returned a total of 243 cases. This is less than 1% of the total dataset but is still a significant number of parcels within one month of data. Moreover, 90% of these 243 cases were regarded as having reached the customer on-time in the Data lake, based on the timestamp of the "Drop & Success" activity. Since the data seem to suggest that other activities occurred to these cases after "Drop & Success", the question arises if these deliveries were correctly marked as reaching the customer in time. A new model was created based on the event log of the 243 cases, to see which activities occurred after Drop & Success. The result is presented in Figure 19.

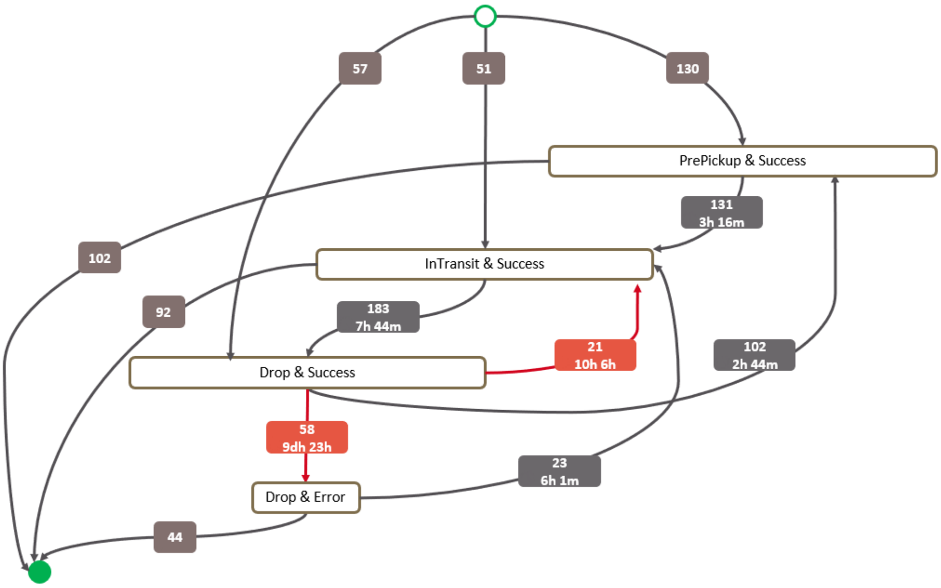

Figure 19. High-level model of Activities after drop and success

<u>Model analysis</u>

Looking at Figure 19, interest is in the lines leading away from the activity “Drop & Success” as the model indicates that these activities occurred after the activity “Drop & Success”. Judging from the graph, these are the activities “InTransit & Success”, “Drop & Error” and “PrePickup & Success”.

Now that we have established the general outlines of what occurs after drop and success, we are interested in which specific activities (i.e., HS-codes) occurred, and their impact on on-time delivery.

Therefore, we add one level of specificity by constructing a new model from the same 243 cases, but now with the HS-codes themselves set as activities, resulting in Figure 20.

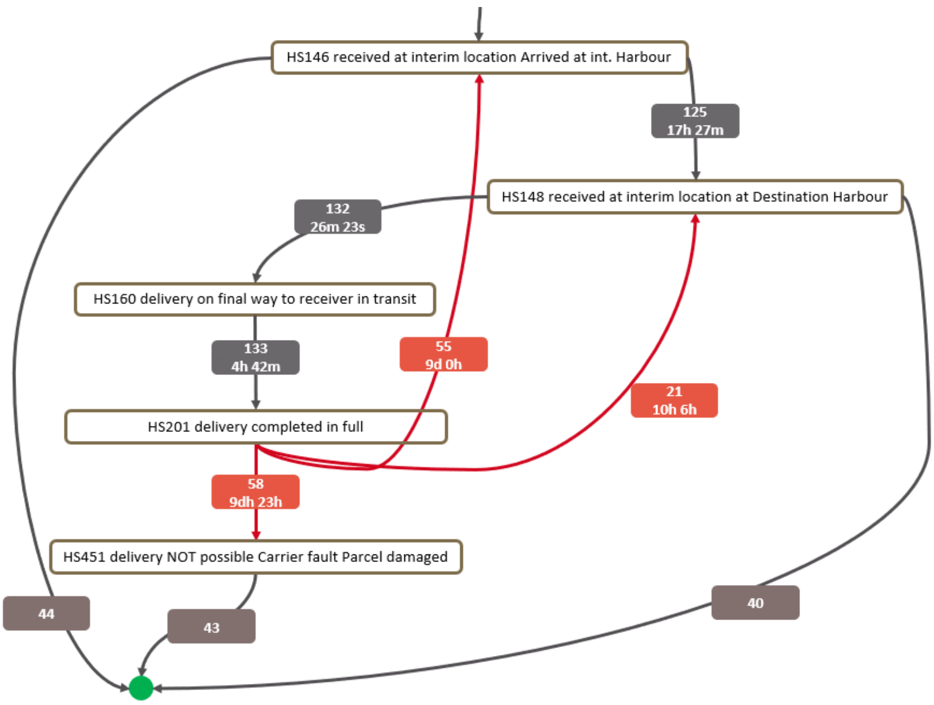

Figure 20. Detailed-level activities after drop and success

Insights

The model in Figure 20 shows the specific activities expressed in HS-codes codes. The HS-code "HS201 delivery completed in full" is the standard code to indicate successful delivery, which was registered as "Drop & Success" in the high-level model in Figure 19. Looking at the red lines leading away from this activity, we see that it is succeeded by the activities "HS146 received at interim location Arrived at int. Harbour", "HS148 received at interim location at Destination Harbour" and "HS451 delivery NOT possible Carrier fault Parcel damaged".

Besides knowing which codes are received, it is worth noting the time difference between supposedly on-time delivery and the other codes positioned on the red lines leading away from the successful delivery activity. Where the time difference between regular steps of the process ranges between mere minutes or maximum 1 or 2 days, the time difference on these red lines is between 9 and 10 days. This is a very long time, especially considering that Hilti's last mile deliveries processes usually only take between one and four days. It is, therefore, safe to assume that these codes are received more than a week after the delivery deadline, marking these deliveries as being late.

After having determined which codes followed the "successful delivery" and assessing their impact on on-time delivery calibration (the codes were received more than a week after the deadline), the following insights were gained for the case study at hand:

1. Analysis of the event log using Discovery and Conformance checking revealed that extra activities are communicated when Hilti is under the impression that the delivery is completed.
2. By using different levels of specificity of the activities in Experiment 1 and 2, the specific activities that occurred after delivery were identified. Two activities refer to the parcel arriving at a hub in the carrier's network (HS146 and HS148) where the third activity indicates an address error (HS451).
3. The true cause of why these codes were received cannot be identified form the event log alone, and the next step is to ask the carrier where these events come from.
4. If one would assume that the transformed data are accurate and therefore reflect what occurred in real-life, it could be concluded that 90% of the 243 reviewed cases, which were originally marked as being delivered in time, did in reality not reach the customer in time. At this moment, however, too little is known about the effects of the external generation and transformation of the data to make this assumption.

In a quick alignment with the carrier presenting the issue described above, the carrier communicated that these three activities should not occur after confirmation of successful delivery. They suspect that these codes are generated due to re-usage of the packaging material of Hilti's parcels, and decided to investigate the matter further, most probably changing the way it uses old packaging material.

In conclusion of Experiment 3 and weighing the insights described above, it is most probable that the parcels selected for this case study did indeed arrive on-time in real-life and that the extra codes were generated by an error at the carrier. Regardless of this outcome, the case study demonstrates that combining Process Mining Discovery and Conformance checking has the potential to gain insights into the correctness of on-time delivery calculation. If an anomaly is found, this reveals a true discrepancy between the carrier report and real-life delivery or exposes an error in the dataset. Either way, applying Discovery and Conformance checking to the transformed event log provides insight into the carrier's process and/or the transformed dataset.

6.5 Reflecting on Experiments 1, 2 and 3

Before addressing Process Mining Enhancement in the next section, we reflect on the three experiments performed in section 6.3 and 6.4 which have shown that:

1. Process Mining Discovery can be applied to the transformed event log and Discover the model that is described in this transformed dataset. This helps to understand various elements of the process including time-consuming parts of the process, common activities and common process variants.
2. Two types of models can be created using Discovery. A process model with classified events used to check (in)correct sequences of activities and secondly, an unclassified process model showing (in)correct usage of HS-codes and/or activities.
3. The data available for Process Mining describe an external process and are transformed from the original event log into the HS-code system. These two factors limit the ability of Conformance checking to make a comparison between the real-life process and the discovered process model.
4. However, well-understood sub-processes can be further analysed for correctness by applying Conformance checking to the discovered models. If a discrepancy is found, it may result in strengthening data quality or directly contribute to challenging the carrier by (dis)confirming the carrier´s report, or even let the carrier change its process.

Taking these points into consideration, we may conclude that both Process Mining Discovery and Conformance checking can be applied to the transformed event log, albeit with (practical) limitations. Their main function can be summarised as aiding to understand the carrier's process, improving data quality, checking the carrier on well-known parts of the process and potentially change the real-life process.

The applicability of Discovery and Conformance checking to the transformed dataset should also be linked to the six Opportunities presented in Chapter 3. After all, these Opportunities were identified by Hilti to positively impact the external last mile delivery process. Do Discovery and Conformance checking gain insights that may help to use these Opportunities?

Three out of the six Opportunities identified seem to be related to Discovery and Conformance checking:

- Opportunity 1, concerning Data quality,
- Opportunity 2, creating a live overview, and
- Opportunity 5, motivating carriers by creating transparency.

Opportunity 1 states that improving the data quality allows Hilti to expand the range and certainty of analysis methods that can be applied to the transformed event log. As shown in the experiments, combining Discovery and Conformance checking shows which process is captured by the transformed event log and if this Discovered process deviates from Hilti's understanding of the carrier´s process. For those parts of the carrier process that are well understood, discrepancies between the transformed event log and the real-life can directly lead to improving data quality or real-life process. We can therefore conclude that Opportunity 1 can be achieved using Discovery and Conformance checking, although it does depend on the shipper´s understanding of the carrier´s process.

Opportunity 2 entails the creation of a live overview of active parcels in the carrier network. Such a "bird-eye perspective" on active deliveries may support managers in keeping track of active parcels and can also be valuable when Hilti management needs to assist the carrier in deliveries that experience some issue. From the experiments performed so far, Discovery seems the perfect tool to create such an overview. Although the technical details of creating an overview that is truly “live” have not been reviewed, Experiment 1 showed that in principle Process Mining Discovery can be used to display the process described by the transformed event log.

Opportunity 5 is based on Hilti’s experience that their efforts in creating transparency in the carrier’s process, by exploiting the carrier’s event log data, have resulted in a greater commitment of the carrier to improve delivery service. As such, there is a strong belief that creating transparency is one of the main ways in which the last mile can be improved in the long term. Assuming that Hilti management is correct, Process Mining Discovery and Conformance checking seem to be the right tools to achieve this transparency. Experiment 3 is a good example of this, by showing that important measures such as on-time arrival can be verified by comparing a discovered process and a model of that process. This way, the transformed event log becomes a better representation of the carrier process and/or the true on-time delivery rate is discovered, both increasing transparency.

Even though the Opportunities described above are to be further developed and are affected by the quality of the data and understanding of the carrier´s process, the first experiments of applying Process Mining to the transformed event log show potential to help Hilti to improve the last mile delivery process.

6.6 Enhancement

Section 6.5 shows the potential of applying Discovery and Conformance checking to the transformed dataset and its potential to exploit three of the Opportunities identified in Chapter 3. Having experimented with applying Discovery and Conformance checking to the transformed dataset, Process Mining Enhancement can be addressed. Where Discovery and Conformance checking generally aid in understanding and verifying how the process occurs in real-life, Process Mining Enhancement typically tries to use the event log to improve the process. As discussed in Chapter 5, Enhancement is seen as having the highest potential impact on the process but also to be the most challenging.

Given the:

- time remaining after the Discovery and Conformance checking experiments,
- general difficulty of Process Mining Enhancement, and
- challenges and novelty of working with an external process and dataset

a full Enhancement experiment lies outside the scope of this thesis. To still review if Enhancement applies to the transformed dataset, we may do the following; by reviewing the results of Chapter 3 and Chapter 5 it can be determined which Enhancement method(s) fit to the Opportunities identified by Hilti management. Furthermore, the transformed dataset can be reviewed to see if it can be used for the selected Enhancement method(s). These two elements are worked out in the remainder of section 6.6.

6.6.1 Matching Opportunities to Enhancement methods

Out of the six Opportunities presented in Chapter 3, it was shown that Discovery and Conformance checking can assist in exploiting Opportunities 1, 2 and 5. In this section, it will be estimated if Enhancement methods can assist in exploiting the remaining Opportunities 3, 4 and 6. These estimations are based on the current understanding of the Opportunities, Enhancement methods available and transformed event log data.

Opportunity 3 proposes a notification system. Such a system would use live events to estimate the risk of parcels arriving late. In case (a group of) parcels are predicted to arrive late, a notification would be sent to those Hilti managers that may be able to solve current issues (e.g., provide correct address information) or warn the customer that a parcel will arrive late. This Opportunity can be linked to Enhancement´s Prediction method, as described by Van der Aalst, Schonenberg and Song (2011). This method would use historic process variants, event attributes and case attributes to predict which active cases may be experiencing issues or may arrive late.

Opportunity 4 suggest to first identify "high-risk" areas, which have a high ratio of late deliveries. This knowledge may then be used to change how the carriers' trucks are loaded, aiming to optimise the way high-risk parcels are handled in the carrier network. Determining the high-risk areas can be based on the on-time delivery calculation already present in the transformed dataset. Successively, a combination of Process Mining Discovery, Enhancement´s Prediction and Bottleneck analysis could be used to see if there are critical time-windows in the carrier network that may be met by changing the order in which the carriers' trucks are loaded.

Finally, Opportunity 6 aims to find bottlenecks within the carrier network explaining why deliveries arrive late (e.g., due to malfunctioning hubs or late pickups). Logically, Enhancement Bottleneck analysis method, as described by Van der Aalst (2012) seems to be the right tool to find these places in the carrier network that often lead to late deliveries.

Summarizing this section, it is estimated that (a combination of) Enhancement´s Prediction and Bottleneck analysis may help to exploit Opportunities 3, 4 and 6. As stated before, exploiting these Opportunities with the Enhancement methods selected lies outside of the scope of this thesis. What will be included, however, is to see if the transformed dataset is suitable for application of these methods. Based on the matching of Opportunities and Enhancement methods described above, the transformed dataset should be reviewed on two factors:

Factor 1: Which data points in the transformed dataset can be used for Prediction and Bottleneck analysis? Both Bottleneck analysis and Prediction base their insights on events in the real-life process. Knowing which real-life events are recorded in the event log and how accurate these points are is crucial for later applications of the methods to the dataset.

Factor 2: Can the on-time delivery rate be determined per area? Opportunities 4 and 6 are based on the assumption that areas with a low on-time-delivery rate can be identified from the transformed dataset. Especially Opportunity 4 relies on this assumption since it aims to alter the order in which parcels are loaded into the carrier´s truck based on their destination. Therefore, it can be seen if the on-time delivery rate can be determined per region and gain insight into which areas are problematic.

Understanding these two elements for the transformed data is no guarantee that the selected methods will be able to exploit the Opportunities, as other challenges may arise during the execution of the projects. Furthermore, it should be noted that the exploration of this dataset simply displays the usability of GLS´s transformed dataset, which may be different from using another carrier´s dataset. However, Factor 1 and 2 are minimum requirements which presence in the dataset can currently be verified. Furthermore, reviewing this dataset will contribute by showing what data are important to collect/verify for future experiments. This way, the correct carrier and dataset can be selected before applying Enhancement methods to the transformed dataset.

6.6.2 Reliability of data points in the transformed dataset

Based on the understanding of the carrier´s process, multiple departure and arrival events occur in the last mile by which the last mile can be tracked. Reviewing the transformed event log and Discovered process in Experiment 1, six out of eight of these events are represented in the dataset. Figure 21 shows these data points, their descriptions and corresponding HS-codes, if applicable. The accuracy of these data points varies; in Figure 21, missing data points are indicated with red, unclear data points with orange, low accuracy data points with yellow and high accuracy data points are marked green.

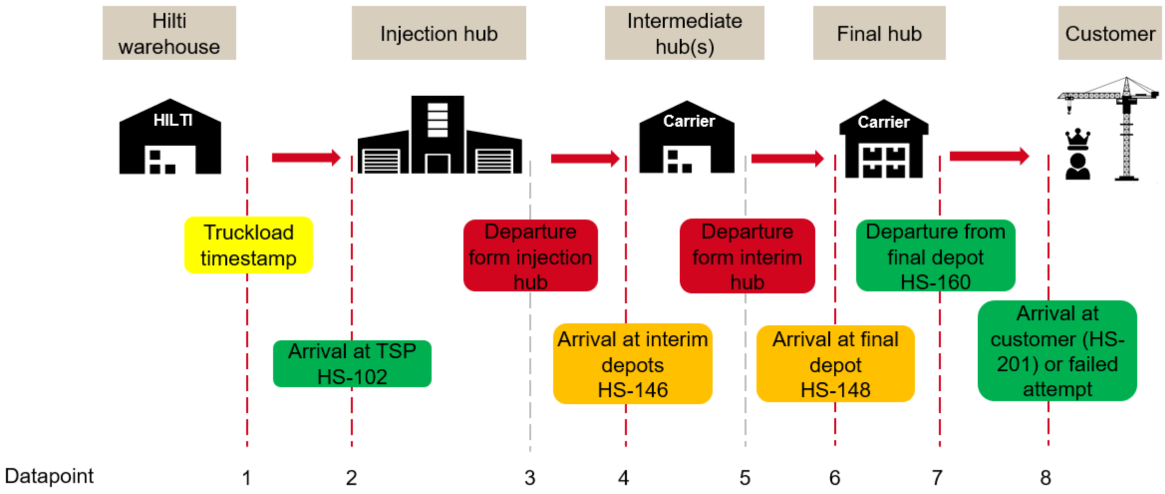

Figure 21. Last mile stages and data points

Missing data points

Data points 3 and 5, departures from the injection hub and departure from intermediate hubs, are not captured by the event log. The most probable impact of these missing data points is lower accuracy of Prediction and Bottleneck analysis. For example, if the bottleneck of a process lies between arrival at the interim hub and arrival at the final hub, it will be difficult to determine whether the problem lies inside or outside the interim hub. However, since the arrival at both the injection hub and the interim hub is measured, there is at least one point of measurement in every physical location in the last mile. The exact effect of the missing data points should be determined with an experiment.

Unclear data points

Data points 4 and 6 are unclear in the sense that they are difficult to interpret. This because often, multiple of these activities are received when only one of these codes was expected. As such, it is unclear which activity is the actual time of arrival. It is expected that the transformation of the carrier´s event log to the HS-codes creates this surplus of data points 4 and 6. When asking the carrier, they indicate that the first time data point 6 is sent, is when the parcel arrived at the final hub in real-life. Any other instance than the first time data point 6 is recorded should simply be ignored in future analysis. For data point 4, it is more difficult to interpret which code is correct, because the parcel may go through multiple intermediate hubs in real-life. The only way to determine if the code signifies a real-life activity from the dataset is by looking at the interval in which the multiple data points are received. It is therefore challenging to determine which of the received code(s) should be interpreted as data point 4 (i.e., arrival at the interim hubs). Data points 4 and 6 are examples of where the transformation of carrier codes to Hilti codes can cause loss of information. If the GLS dataset were to be used for Enhancement Prediction and Bottleneck analysis, data point 4 and 6 could form a serious barrier to getting insights since they make it difficult to know which activity occurred in real-life. However, there are ways to deal with this data issue in the future. A data-pre-processing step could be applied to the transformed event log, determining which of the received codes signify real-life events based on interval for example. Another way to accomplish this is to ask the carrier (i.e., GLS) for the original event log and use this information to filter out irrelevant events. Thirdly, the transformation rules between the carrier event log and HS-codes could be revised, together with the carrier, building transformation rules that only map relevant data. Finally, it is most likely that this issue has arisen from the transformation of the GLS event log to the HS-code system. Consequently, it is possible that if one were to select another carrier, these specific issues would not arise. It is recommended for Hilti to select a dataset for which the HS-codes can be clearly traced to one of the eight real-life events in the carrier process that are presented in Figure 21.

Low accuracy data points

Data point 1 (Truckload timestamp) is not generated by the carrier but by Hilti itself. Hilti indicates that data point 1 is not to be considered the exact moment when the truck departs from the Hilti warehouse. Instead, it is the time when the bridge where the truck was loaded at the Hilti warehouse, has closed. As a result, the truck could have left half an hour earlier or later according to Hilti, diminishing the accuracy of data point 1. This may impact the Bottleneck analysis and Prediction capabilities of Enhancement. Especially in case the injection hub is calculated to be a bottleneck, the exact departure time at the Hilti warehouse may become increasingly important. If the injection hub turns out to be the bottleneck, Opportunities 3, 4 and 6 may be affected by the inaccuracy of data point 1. It may also turn out that later stages of the process are more problematic, in case the impact of low accuracy data points may be mitigated.

High accuracy data points

Data points 2, 7 and 8 are considered accurate. It shows from the data that no unwanted duplicates exist for these data points and that their timestamps are identical to real-life events. Furthermore, almost all data point 7´s are followed by data point 8, signalling a successful or a failed delivery. The data point 7´s that are not followed with one of these (dis)confirmation codes all have a new data point 7 the following day. This signals that the delivery has not reached its target and a new delivery attempt is made the next day. Most of the delivery attempts are successfully confirmed by a data point 8 however, signalling that the order has been completed in full at the customer. This perhaps detailed analysis of the order in which the data points are received does signal that the last section of the last mile is indeed logical and close to the real-life process. These accurate data points are important for the overall usability of the data. Considering that both data point 7 and 8 are considered reliable, it would be possible to perform future experiments on this part of the process.

Impact on Process Mining Enhancement

The exact impact of the data point quality on Enhancement methods is difficult to determine without a practical Enhancement experiment. However, based on the assessment of the data point quality described above, we deduct the following:

1. Six out of the eight potentially measurable data points on which Enhancement Prediction and Bottleneck analysis could be based, are represented in the dataset. It is expected that selecting a dataset which has 7 or 8 data points will increase the precision of Enhancement's Prediction and Bottleneck analysis.

2. Data point 1 can be considered to have low accuracy. It is speculated that this may make Enhancement methods less precise but should not block these methods from being applied.
3. Any dataset to be used for future experiments should be checked for duplicate activities that do not reflect real-life events, such as the duplicates of data points 4 and 6 in this dataset.
4. Of the currently reviewed dataset, the part between data point 7 and 8 was found to be accurate. This part of the process can be used for initial experiments with Enhancement.
5. As other issues in the transformed dataset may arise, it is recommended to:
    - select a carrier that is eager to cooperate to make the transformation to the HS-code system representable for the original event log, and
    - collect the original event log of the transformed event log being analysed, so that it can be known how the translated dataset was created

The exploration of the dataset for Enhancement purposes may give an indication of which part of the transformed dataset can be used for Bottleneck analysis and Prediction experiments. Furthermore, some practical recommendations for the selection and verification of the transformed dataset for future applications of Enhancement were deducted. In the next section, the part of the process between datapoint 7 and 8 and found to be most accurate, is used to determine if high-risk areas can be identified from the transformed dataset. An insight that is needed for Bottleneck analysis.

6.6.3 Experiment 4: Determining areas with a low on-time delivery rate

There are two case attributes available in the dataset that, if combined, should allow us to determine high-risk areas:

1. “on-time delivery”, a Boolean stating whether the case arrived at the customer on-time, and
2. “postal-code”, detailing the postal code of the customer.

These two case attributes were added to the transformed event log described in section 6.2. With Python package pgeocode (version 0.1.0), the postal codes of the delivery address were used to gather additional geographic information (i.e., coordinates, community code, region, etcetera). Combining this information with the “on-time delivery” case attribute, high-risk areas were calculated on multiple levels of specificity, as shown in Figure 22.

Postal_Code	mean	count
79379	0.800000	10
79539	0.937198	207
79106	1.000000	3
79618	1.000000	18
79599	1.000000	1
79427	1.000000	1
79410	1.000000	2
79400	1.000000	11
79350	1.000000	28
79312	1.000000	18
79294	1.000000	10
79268	1.000000	1
79639	1.000000	1

community_code	mean	count
9565.0	0.777778	9
9563.0	0.777778	45
9361.0	0.780488	41
9564.0	0.888268	179
9271.0	0.892857	56
9176.0	0.916667	12
9574.0	0.976190	42
9373.0	0.983051	59
9575.0	1.000000	2
9573.0	1.000000	9
9571.0	1.000000	4
9474.0	1.000000	6
9376.0	1.000000	2

Figure 22. On-time delivery ratios calculated per postal code (left) and community code (right)

Figure 22 displays on-time ratios per postal code and community code. The number in the "mean" column indicates the ratio of on-time deliveries for that region. The figure shows that the calculation of on-time delivery rates are possible per region, where the region can be defined on different levels of specificity. Furthermore, using the coordinates of the postal codes combined with the "on-time delivery" Boolean, high-risk areas can be visualized on a map as depicted in Figure 23.

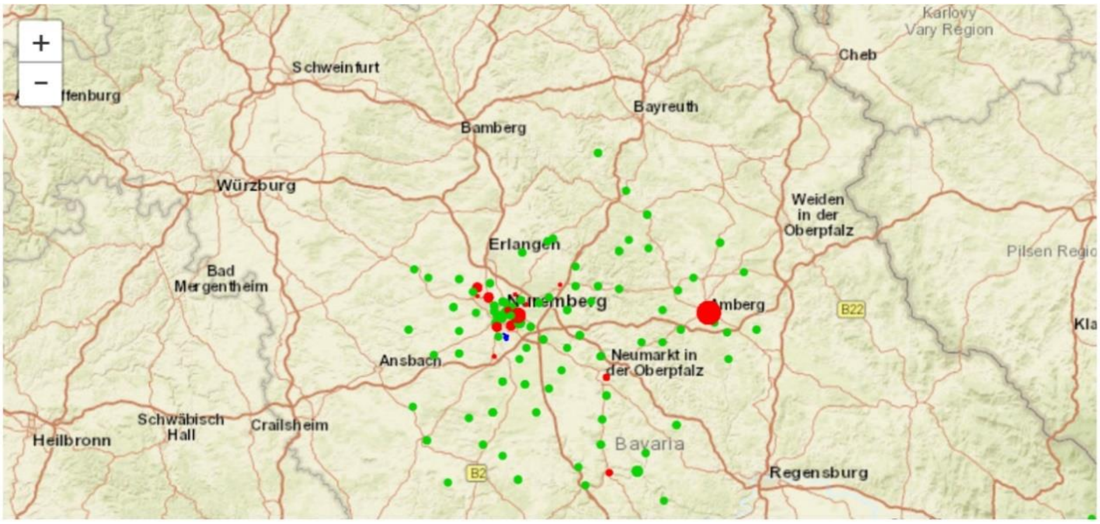

Figure 23. Areas performing high and low in on-time delivery displayed on a geographical map

Based on the examples provided above, it can be concluded that it is possible to identify which areas are performing low on on-time delivery. By turning the on-time delivery rates into case-attributes, the calculated on-time delivery rate can be added back into the event log to be used during Enhancement. This can make it easy to select those cases that belong to a certain low-performing area for Bottleneck analysis, possibly enabling Opportunities 4 and 6.

6.7 Answer to Research Question 3

Based on the findings in Chapter 6, Research Question 3 will be addressed:

Which Process Mining methods can be executed with a transformed external last mile event log?

In response to this Research Question, it was seen if Process Mining Discovery, Conformance checking and Enhancement could be executed with the transformed dataset. The results of this assessment are discussed below.

<u>Discovery</u>

The goal of Process Mining Discovery is to extract the process that is registered in the transformed event log. Experiment 1 and 2 achieved the following:

- extracting a process overview from the transformed event log
- discovery of the most common process, activities and time-consuming process steps
- a version of the process with grouped activities, simplifying the process model and facilitating Conformance checking analysis

Conformance checking

The goal of Process Mining Conformance checking is to explore the (dis)similarity between the model of a process and the discovered process. In Experiment 3, Conformance checking compared the model discovered from the transformed event log to Hilti's understanding of the carrier process, which resulted in;

- an assessment of the logic of the discovered process,
- identification of the points on which the discovered process was (dis)conform with Hilti's understanding of the carrier process, and
- a case study in which it was shown that a small percentage of the parcels were attributed activities after the delivery deadline, even though most of these parcels were marked as reaching the customer on-time; the case showed that Conformance checking can verify carrier reports and change the real-life process.

Enhancement

As stated before, a complete Enhancement experiment similar to the Discovery and Conformance checking experiments was deemed too extensive to be included in this thesis. As an alternative, it was reviewed which Enhancement methods fit Hilti's Opportunities and how a transformed dataset could potentially be used to apply these methods. As a result:

- Enhancement's Prediction and Bottleneck analysis were identified as most relevant to exploit Hilti's Opportunities
- It is better understood which data points in the reviewed dataset are usable for Enhancement´s Prediction and Bottleneck analysis
- It is known which carrier and accompanied dataset/data points should be selected for Enhancement´s Prediction and Bottleneck analysis
- It is shown that the event log contains the right information to determine areas with low performance in on-time delivery, which is needed for Bottleneck analysis.

Effect of the transformation and externality of the data on the application of analysis methods

For a complete answer to Research Question 3, the impact of using transformed data of an external process should be discussed. During the application of Discovery, Conformance checking and Enhancement, the effect of using transformed data became apparent in the following ways:

1. Discovery is only able to capture the process described in the transformed dataset, which may or may not be representative for the event log of the real-life process.
2. Similarly, Conformance checking can only compare the discovered model to an understanding of the carrier process, which may differ from the real-life process.

The above points show the altered goals of the Discovery and Conformance checking in the context of a transformed dataset. It could be stated that in case a transformed event log is analyzed, an important function of the current Discovery and Conformance checking experiments is a better understanding of the (dissimilarities) between the transformed dataset and Hilti's understanding of the carrier's process. However, to use the full potential of Process Mining Discovery, Conformance checking and Enhancement, one out of the following two gaps should be bridged at a minimum:

- the gap between the real-life process and Hilti's understanding of this process, or
- the real-life arrival and departure events and their representation in the transformed event log.

Finally, we may turn to answer Research Question 3 in stating which event log analysis methods can be applied to a transformed event log.

Process Mining Discovery and Conformance checking have gained insight into the transformed dataset. This insight differs from ‘classic’ Discovery and Conformance checking insight, which is gained in the real-life process. Now, the real-life process could only be approached. The main function of Discovery was to extrapolate the process contained in the transformed event log and make this process interpretable. Applying Conformance checking gained insight into the logic of the discovered process and how it related to Hilti’s understanding of the carrier’s process. All things considered, it showed that Process Mining Discovery and Conformance checking have successfully gained insight into the transformed dataset.

With regards to Enhancement, more research should be performed to see if Enhancement can improve the process by using a transformed event log in practice. However, it was shown that the transformed dataset holds the potential for application of Enhancement methods, and which steps have to be taken during collection of the transformed dataset to facilitate Enhancement´s Prediction and Bottleneck analysis.

Chapter 7. Discussion

In this chapter, findings and their implications are discussed, as well as the limitations and recommendations of this thesis.

7.1 Summary of findings

The three Research Questions and their respective findings are summarized in Figure 24 and are elaborated upon below.

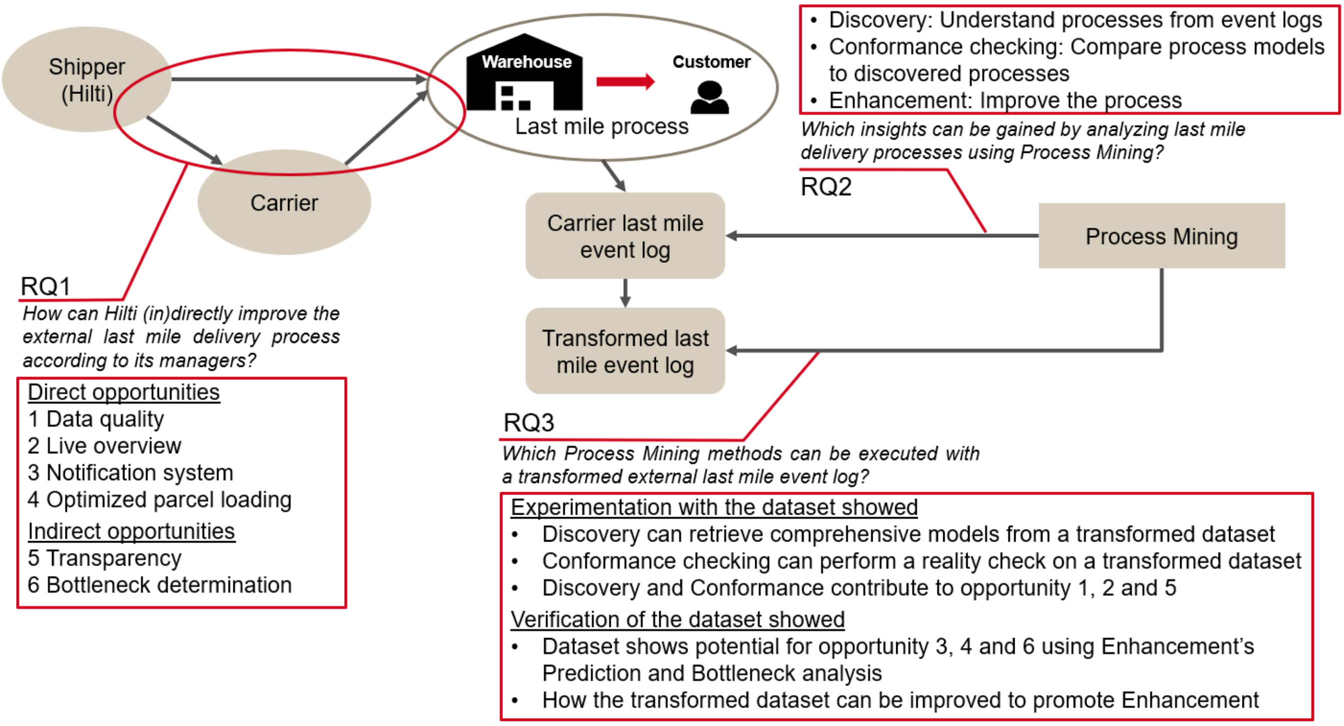

Figure 24. Findings to Research Questions 1, 2 and 3

Through a qualitative interview with Hilti management, Chapter 3 explored Opportunities for Hilti to improve the external last mile process in answer to Research Question 1. Six Opportunities to (in)directly improve Hilti´s external last mile were identified in collaboration with Hilti management.

Direct Opportunities
Opportunity 1: Improve the data quality to structure the shipper´s last mile management

Opportunity 2: Use data to construct a (live) overview of parcels in the external process

Opportunity 3: Notify customers of a (potentially) late delivery based on live carrier data

Opportunity 4: Change the order of parcel in the carrier's truck to reach high-risk areas

Indirect Opportunities (depending largely on collaboration with the carrier)
Opportunity 5: By creating transparency within the carrier network, the carrier improves the process

Opportunity 6: Find bottlenecks within the carrier network to avoid or solve them with the carrier

In answer to Research Question 2, a literature review was performed which presented insights gained using three categories of Process Mining that are often applied in sequence of increasing difficulty, starting with Discovery, followed by Conformance checking and ending with Enhancement. Discovery creates understanding in the process recorded by the event log. Conformance allows for comparison between a discovered process and a process model. Enhancement aims to improve the process. No literature was found which described the analysis of a transformed event log detailing an external last mile delivery process.

Research Question 3 aimed to see which Process Mining methods can be applied to a transformed event log. Experimenting with a subset of Hilti's data showed that Discovery can extract the process that is captured in a transformed dataset. Furthermore, it showed that Conformance checking can:

- validate if the discovered process is logical,
- compare the discovered process to the understanding of the carrier's process, and
- challenge the carrier on real-life processes.

Combining Discovery and Conformance checking were shown to enable Opportunities 1 and 5. Furthermore, Discovery laid the foundation for achieving Opportunity 2.

Since no practical Enhancement experiment was performed, it cannot be said if Enhancement methods can be applied to the transformed dataset. However, it was determined that:

- to attempt Opportunities 3, 4 and 6, Enhancement's Prediction and Bottleneck analysis are best suitable
- which data points can and should be used as input for Enhancement, and
- how the improve the (collection of) the dataset to promote Enhancement.

Finally, it was shown that based on the information available, analysing the transformed dataset and comparing the discovered model to Hilti's understanding of the carrier's process is a comparison relatively far removed from the real-life process. This since neither the transformed event log nor Hilti's understanding of the carrier's process is currently verified by the carrier.

7.2 Implications

The implications affect two domains. On the one hand, it needs to be discussed how the findings may contribute to improving Hilti´s last mile process. On the other, the question of what these findings entail for shippers that want to improve their external last mile, in general, is addressed.

7.2.1 Implications for Hilti AG

This thesis presented six Opportunities for Hilti to improve the external last mile in Chapter 3. In Chapter 5 and Chapter 6, it was reviewed how these Opportunities can be exploited using the transformed dataset. Although each Opportunity aims to improve the last mile, it is worth noting the different (potential) implications they may have.

Opportunities 1, 2, 3 and 5 can be said to primarily assist Hilti management in their current activities. As was shown in Chapter 2, Hilti is responsible for supportive tasks in the last mile process, such as communication with the customer and assisting the carrier in solving pressing problems. Chapter 1 and Chapter 3 described the responsive nature of how Hilti currently executes these tasks and the potential that lies in a more pro-active and structured way of supporting the last mile. Opportunity 1, 2 and 3 support this shift to pro-active and structured last mile management.

Opportunity 1 does this by making the transformed dataset a better representation of what occurs in real-life. This is crucial in enabling Hilti to base their actions on the dataset rather than mostly on sporadic emails or calls from the carrier as is the case now. As was shown in Chapter 6, the data quality can already be improved, meaning that Opportunity 1 can be used directly.

Opportunity 2 has two (potential) effects. Firstly, an overview of historic data allows Hilti to understand the process that is described in the transformed event log. This is fundamentally different from the analysis performed with the data thus far, where KPI's were calculated on single data points, no process could be deducted from the data and no information was available concerning (common) process variants. As such, application of Discovery and Conformance checking gives Hilti the ability to understand the various processes, enabling them to understand cause and effect in the delivery process. Secondly, a live overview could help Hilti move beyond understanding historical cases and assist in managing active parcels. This does not only increase Hilti´s overview of daily operations, it is expected to structure managing parcels for which an error occurred. This since the process for that delivery thus far is known as well as common effects of the error are available and understood, supporting decision making. The first effect of Opportunity 2 is already obtained by the overviews created. However, to obtain the effect of a live overview, technical details of collecting and using live data should be worked out further.

Opportunity 3 aims to go one step further in aiding Hilti management in managing troublesome deliveries. When based on single events, Process Mining tools can be used to automatically send a message to Hilti managers or a customer with information about a possibly delayed parcel. However, if such a notification system is combined with Enhancement's Prediction, a message could be generated if a sequence of events is recognized that historically has led to delayed deliveries. Based on the preliminary outcomes of this Thesis, Hilti has decided to prepare such a notification system using a commercial Process Mining tool. This effort has started with setting up a system using single events but aims to include Prediction to recognize possible delays from a particular sequence of events. Future experiments will show if such a notification system can produce relevant messages to Hilti managers and/or customers. If such a system works, however, it is a strong enabler to pro-actively inform customers about delayed parcels, which in turn may raise the perceived service quality. If and how such notifications can raise perceived service quality could be further researched to strengthen Opportunity 3.

Opportunity 5 is based on Hilti's experience that creating transparency within the carrier network has resulted in motivation on the part of the carrier to improve the last mile together with Hilti. If we assume that further increasing transparency has a similar effect, Opportunity 5 implies that using Process Mining can lead to a more fruitful collaboration with the carrier. However, it is imaginable that some carriers may respond differently and find the increased transparency a reason to return to a more transactional collaboration. Therefore, more research on the effect of increased transparency is required to understand how to exploit Opportunity 5.

In contrast to how Opportunities 1, 2, 3 and 5 aid Hilti in its current, mostly supportive responsibilities, Opportunities 4 and 6 describe entirely new ways for Hilti to improve the external last mile. Opportunity 4 exploits one of the few points where Hilti can physically influence the last mile. It remains to be seen if changing the order in which the parcels are loaded in the carrier's truck does indeed make a difference when aiming to reach high-risk areas on-time. It is expected that if a bottleneck is discovered at the injection hub, using this method has the highest probability of being effective. If the Opportunity is exploited successfully, it enables Hilti to directly raise the on-time delivery rate of critical areas, something that has been difficult in the past. Opportunity 6 also provides Hilti with the means to improve on-time delivery, presenting two options to deal with an identified bottleneck: Hilti may choose to avoid the bottlenecks (e.g., choose a different carrier for problematic routes) or collaborate with the carrier to solve these bottlenecks. Moreover, knowing that a route or hub is problematic may be of increased utility when managing an important delivery, e.g., Hilti's "VIP" customers.

An opportunity not mentioned by Hilti management but which presented itself during exploration of applying Process Mining to transformed even log data is the ability to structurally check if the process model of the carrier process corresponds to the discovered model Using Conformance checking. This may prove to be useful in the long term when the carrier´s process is relatively well known and stable, and discrepancies between the discovered process and process model are less frequent and more probable to reflect real changes in the carrier´s process, as described by Li and Deng (2009).

Finally, we may conclude that applying Process Mining to the transformed event log data has the potential to aid Hilti in meeting their current challenges as described in section 1.3. As was shown during Experiment 3 in section 6.4.2, Process Mining can easily select specific sequences of events, enabling Hilti to understand if a parcel indeed arrived on time. This should support them in calibrating, understanding and verifying their true on-time delivery rate. Furthermore, Opportunities 2 and 3 can increase overview and foresight, which may enable a more pro-active management style, possibly increasing perceived service levels. Finally, Opportunities 4 and 6, if executed successfully, may provide Hilti with the ability to improve the on-time delivery rate of their external last mile.

7.2.2 Implications for Shippers in general

This section discusses the implications of the findings for shippers in general.

The findings of this thesis suggest that Process mining Discovery, Conformance checking and Enhancement can enable the shipper to improve the external last mile delivery process.

The first step

The results imply that the first step in improving the external last mile is understanding how the transformed event log data represent the carrier´s last mile delivery process. During the experiments, it became clear that there can be a significant difference between the carrier's real-life process and the process detailed in the transformed event log. This means that analysing the transformed event log without verification of the data may give a distorted view of the carrier´s process and how it may be improved. Such differences may have many reasons. In this thesis, they are suspected to be the result of the transformation from the original event log to the shipper's interpretation of these data and due to faulty event creation by the carrier. Because the other analysis are based on the (transformed) event log, understanding if and how the event log truly represents the carrier´s process is suggested as the first step for any shipper. As indicated with Opportunity 1, Discovery and Conformance checking can be used to improve the data quality and understand differences between the transformed event log and the understanding of the carrier's process. Doing such analysis first allows the shipper to rely on the dataset. However, to understand the real-life process, a good understanding of the real-life process and/or the original event log is required.

Other Opportunities

If this first step is taken with Opportunity 1 and it is known which parts of the transformed event log represent the carrier's process, the data can be used by the shipper to improve the external last mile process. Both the literature review in Chapter 5 and the experiments in Chapter 6 show that one can best start with Discovery, followed by Conformance checking and finally attempt Enhancement. The overview and transparency created with Opportunities 1, 2 and 5 can be seen as the foundation for Enhancement with Opportunities 3, 4 and 6. As such, it is implied that shippers succeed Opportunity 1 by Opportunities 2 and 5, before attempting 3, 4 and 6.

Usability of Opportunities for different shippers

In principle, any shipper with a dataset that represents the carrier's process should be able to attempt Opportunities 1, 5 and 6, as they depend on historic event log data. Opportunities 2 and 3 depend on the availability of event data soon after the events occurred in real life, for the live overview or predictions to be relevant. As such, a shipper using carriers that make their data available for extraction as soon as it occurred in real-life, can attempt Opportunity 2 and 3. Opportunity 4 assumes that the shipper has an influence on the order in which the carrier's truck is loaded in the warehouse. As such, Opportunity 4 can only be attempted by the shipper if the shipper can influence the order of loading the parcels.

Although it is not a condition, there is another factor that is expected to influence the potential of some Opportunities for different shippers: the number of measurable data points in the dataset. For Opportunities 3, 4 and 6, it is expected that having more (accurate and relevant) data points in the last mile event log allows Enhancement´s Prediction and Bottleneck analysis to be more precise, increasing the chance that these methods are successful. This suggests that shippers using carriers that have an accurate, detailed and interpretable last mile event log have an advantage in exploiting these Opportunities. A final factor that may influence the applicability of the Opportunities is the availability of on-time delivery data combined with location data. In principle, on-time delivery can be calculated for each parcel from delivery timestamps and delivery deadlines, but these data may be (temporarily) unavailable for some carriers. Before these data are collected, it will not be possible to determine high-risk areas and therefore attempt Opportunity 4.

Readiness of Opportunities

The results also show varying levels of certainty between the Opportunities with regards to if and how they will currently work in practice. The experiments showed that the insights of Opportunity 1 (improved data quality), part of Opportunity 2 (historic overview) and Opportunity 5 (transparency of the carrier's process) can be obtained with a transformed event log. It should be noted however that the potential effect of Opportunity 5, increased transparency, might vary between carriers and should be further researched. No practical experiments have been performed for Opportunities 3, 4 and 6 and the live overview from Opportunity 2. As such, their usability in practice is therefore currently largely unknown.

7.3 limitations and Recommendations

In Process Mining, an event log is used to understand and improve a real-life process. Since Process Mining insights are based on the event log used as input, the relevance of these insights strongly depend on how well the event log represents the actual process. In this thesis, it is reviewed how a transformed event log, created by an external carrier, can be used to improve an external last mile delivery process. As mentioned before, the combination of using a transformed event log while analysing an external process increases the potential discrepancy between the event log used and the real-life process. If the used event log is indeed a misrepresentation of the real-life process, this could make the conclusions drawn by Process Mining less-relevant. In this thesis, this potential discrepancy was mitigated by directly consulting the carrier on vital assumption on his process and its potential effects on the transformed event log. To structurally reduce the risk of discrepancy between the used event log and the real-life process, shippers are recommended to do the following:

1. Improve the understanding of the carrier's process: by having a better understanding of the carrier process, a shipper will be able to know if their version of the event log corresponds to reality or not. A better understanding can be achieved by selecting carriers that see value in

jointly improving the last mile, or by working closely together with current carriers and verify assumptions about their process.

2. Understand the effect of transforming the original event log to the transformed event log: as detailed in Chapter 4, the transformation of the original events log to the HS-code system is executed by an external party, and only the transformed event log was available to the shipper. If the original event log is available, it is much easier to understand how the transformed log was created exactly and what the effects of the transformation are.

Another limitation of this thesis is the usage of only one dataset, consisting of parcels delivered by one carrier (GLS) in one MO (Germany). Limiting the dataset to one carrier in one MO was needed to ensure correct interpretation of the event log. However, it is possible that experimenting with data from different carriers and/or MO's could have gained extra insights, particularly with regards to data quality and applicability of Opportunities. While comparing carriers and MOs to select a dataset, no major differences in event logs were spotted that could fundamentally alter the results of this thesis. Similar events, in terms of HS-codes and frequency, were recorded in all datasets. Still, it would be valuable to experiment with other datasets, including event logs originating from different carriers or MO´s, to get a better understanding of how the Opportunities can be exploited by other shipper-carrier combinations.

Furthermore, the HS-code transformation system was now largely regarded as a given. It is expected however that analysing, and if an alternative setup would enable a transformation with less loss of valuable information is possible.

The opportunities aim to improve the external last mile. For the Opportunities using Enhancement (Opportunities 3, 4 and 6) it was already mentioned that more experiments need to be conducted to verify if these opportunities can be achieved. Moreover, even though it was confirmed that the effect of Opportunities 1, 2 and 5, can be achieved (i.e., improved data quality, a live overview of parcels in carrier's network and more transparency on the carrier process), it has yet to be proven that these Opportunities do indeed have a positive effect on the external last mile delivery process. It is therefore advised to implement these opportunities and evaluate if, for the customer, the last mile delivery service indeed improved.

Furthermore, it became clear towards the end of the experiments that many Process Mining tools have modules dedicated to cleaning the event log. Often, this involves filtering out duplicate or partial process instances, of which it is determined that they do not represent real-life parcels. This structural cleaning of the data has not been included in this thesis. In hindsight, using such methods could also have aided in getting the transformed event log to be a better representation of the carrier's process. However, such cleaning of the event log may discard uncommon but important process-deviances, such as the one described in Experiment 3, so caution is advised. Regardless, it is expected that such cleaning of the dataset will be most important when trying to use Enhancement methods such as Prediction and Bottleneck analysis. The "most common" processes that were compared to the understanding of the carrier's process with conformance checking are unlikely to be affected since faulty process instances are most commonly the exception. Still, it is recommended to use the event log cleaning modules in future experiments to improve the discovered process model.

Another limitation worth noting is the absence of academic literature that investigates the relationship between the shipper and the carrier. A good example where such an understanding of this relationship could have contributed is in estimating the effects of increased transparency (Opportunity 5) as discussed in section 7.2. Furthermore, understanding this relationship between the shipper and the carrier could have resulted in other opportunities to improve the external last mile for the shipper and the carrier. This thesis took the point of view of one shipper, from which six main Opportunities came forth. However, it is strongly suspected that more (in)direct opportunities exist to use the carrier data, among which using other Enhancement methods or opportunities dependent on collaboration with the carrier. It is therefore advised to review the academic literature on collaboration, look at opportunities from the carrier's perspective and conduct further experiments to strengthen the opportunities identified with this thesis and identify novel ones.

Finally, shippers are recommended to continue experimentation, with the Process Mining tools and Opportunities discussed. The case study described in Experiment 3 is just one of the ways in which the "logic" of the carrier's process can be verified, and it is expected that performing similar case studies will gradually decrease the discrepancy between the process described by the event log and the real-life process. Furthermore, shippers are encouraged to further experiment with the Opportunities linked to Enhancement (Opportunities 3, 4 and 6). This given their current high state of uncertainty on how they will work in practice, and their speculated potential to improve the service quality and/or on-time delivery rate. It is recommended that for these experiments, a carrier should be selected whose dataset holds many data points, i.e., points of arrival and departure in the last mile delivery processes that are recorded accurately in the event log. By selecting these datasets, Bottleneck analysis and Prediction can be more precise in their analysis, promoting Enhancement in improving the last mile.

Chapter 8. Conclusion

In answer to the main Research Question;

> *"How can carrier event log data be used to improve an external last mile delivery process?",*

six Opportunities were identified that can help improve the external process using the transformed event log data from a shippers perspective. Using Process Mining Discovery and Conformance checking, it was shown that the data quality can be increased (Opportunity 1), the carrier's process can be made more transparent (Opportunity 5) and that a (live) overview (Opportunity 2) should, in principle, be obtainable. These three Opportunities and Opportunity 3 allow the shipper to improve the external last mile by empowering the shipper in its current, mainly supportive responsibilities in the last mile.

It is speculated that Process Mining Enhancement's Bottleneck analysis and Prediction can notify the shipper or the customer of (potentially) late deliveries (Opportunity 3), increase on-time delivery rates by changing the order in which parcels are loaded (Opportunity 4) and find bottlenecks in the carrier network (Opportunity 6). Opportunity 3 has the potential to improve the customer's perception of service quality; Even if a parcel arrives late, the notification system avoids having them wait in vain. Moreover, Opportunities 4 and 6 may improve the on-time delivery rate itself by reaching high-risk areas on-time using altered truck loading (Opportunity 4) or by identifying and avoiding bottlenecks (Opportunity 6). Such concrete opportunities to improve the service quality and/or the on-time delivery rate using event log data and Process Mining Enhancement are, to the researcher's knowledge, new and potentially significant ways for shippers to improve their externally executed last mile delivery process.

An important first step in using the carrier's event log to improve the external process is understanding how to interpret the event log and knowing how it represents the carrier's process. To accomplish this, it is suggested to collaborate with carriers to understand their process and to obtain the original event log. The latter suggestion is important when the transformation of the event log may have led to a distortion in the dataset. After it is understood how the event log represents the carrier's process, it is advised to start with the supportive Opportunities 1, 2, and 5, which are largely based on the more accessible Discovery and Conformance checking methods, before moving to the high impact Opportunities 3, 4 and 6 based on Enhancement.

References

Akeb, H., Moncef, B., & Durand, B. (2018). Building a collaborative solution in dense urban city settings to enhance parcel delivery: An effective crowd model in Paris. *Transportation Research Part E: Logistics and Transportation Review*, *119*, 223-233. https://doi.org/10.1016/j.tre.2018.04.007

Akman, B., & Demirörs, O. (2009). Applicability of processed discovery algorithms for software organizations. In *Proceedings of the 35th EUROMICRO Conference on Software Engineering and Advanced Applications,* 195-202. IEEE. https://doi.org/10.1109/seaa.2009.87

Becker, T., & Intoyoad, W. (2017). Context aware process mining in logistics. *Procedia Cirp, 63*, 557-562. Elsevier. https://doi.org/10.1016/j.procir.2017.03.149

Lune, H., & Berg, B. L. (2017). *Qualitative research methods for the social sciences* (9th ed.). Boston, MA: Pearson. Retrieved from jaipuria.ac.in

Berti, A., van Zelst, S. J., & van der Aalst, W. (2019). Process mining for Python (PM4Py): Bridging the gap between process-and data science. Retrieved from arXiv.org

Besri, Z., & Boulmakoul, A. (2017). Framework for organizational structure re-design by assessing logistics' business processes in harbor container terminals. *Transportation Research Procedia, 22*, 164-173. Elsevier. https://doi.org/10.1016/j.trpro.2017.03.023

Boyaci, T., & Ray, S. (2006). The impact of capacity costs on product differentiation in delivery time, delivery reliability, and price. *Production and Operations Management, 15*(2), 179-197. https://doi.org/10.1111/j.1937-5956.2006.tb00239.x

Burns, L. D., Hall, R. W., Blumenfeld, D. E., & Daganzo, C. F. (1985). Distribution strategies that minimize transportation and inventory costs. *Operations research, 33*(3), 469-490. https://doi.org/10.1287/opre.33.3.469

Cai, H., Xu, L., Xu, B., Zhang, P., Guo, J., & Zhang, Y. (2018). A service governance mechanism based on process mining for cloud-based applications. *Enterprise Information Systems, 12*(10), 1239-1256. Retrieved from www.sciencedirect.com

Castellanos, M., De Medeiros, A. A., Mendling, J., Weber, B., & Weijters, A. J. M. M. (2009). Business process intelligence. In Handbook of research on business process modeling (pp. 456-480). IGI Global. https://doi.org/10.4018/978-1-60566-288-6.ch021

Data Requirements. (n.d.) Retrieved from https://fluxicon.com/book/read/dataext/

De Medeiros, A. A., & van der Aalst, W. M. (2008). *Process mining towards semantics* [lecture notes]. Advances in web semantics I. doi: https://doi.org/10.1007/978-3-540-89784-2_3

De Smedt, J., van den Broucke, S. K. L. M., Obregon, J., Kim, A., Jung, J. Y., & Vanthienen, J. (2017). Decision mining in a broader context: An overview of the current landscape and future directions. In M. Dumas, & M. Fantinato (Eds.), *Business Process Management Workshops* (pp. 197–207). Cham, Switzerland: Springer International Publishing AG. https://doi.org/10.1007/978-3-319-58457-7_15

dos Santos Garcia, C., Meincheim, A., Junior, E. R. F., Dallagassa, M. R., Sato, D. M. V., Carvalho, D. R., Santos, E. A. P. & Scalabrin, E. E. (2019). Process Mining Techniques and Applications-A Systematic Mapping Study. *Expert Systems with Applications, 133,* 260-295. https://doi.org/10.1016/j.eswa.2019.05.003

Dunkl, R., Rinderle-Ma, S., Grossmann, W., & Anton Fröschl, K. (2015). A method for analyzing time series data in process mining: Application and extension of decision point analysis. In S. Nurcan, & E. Pimenidis (Eds.), *Information Systems Engineering in Complex Environments* (pp. 68–84). Cham, Switzerland: Springer International Publishing AG. https://doi.org/10.1007/978-3-319-19270-3_5

Evermann, J., Rehse, J.-R., & Fettke, P. (2017). Predicting process behaviour using deep learning. *Decision Support Systems, 100*, 129–140. https://doi.org/10.1016/j.dss.2017.04.003

Fancello, G., Paddeu, D., & Fadda, P. (2017). Investigating last food mile deliveries: A case study approach to identify needs of food delivery demand. *Research in Transportation Economics, 65*, 56-66. https://doi.org/10.1016/j.retrec.2017.09.004

Fernie, J., & Sparks, L. (Eds.). (2018). *Logistics and retail management: emerging issues and new challenges in the retail supply chain*. London, England: Kogan page publishers. Retrieved from www.iimsedu.com

Fluxicon (n.d.) The three minimum requirements for process mining: A Case ID, an Activity name and at least one Timestamp column. [Online Image] Retrieved from https://fluxicon.com/book/read/dataext/

Forslund, H., & Jonsson, P. (2010). Integrating the performance management process of on-time delivery with suppliers. *International Journal of Logistics: Research and Applications*, *13*(3), 225-241. https://doi.org/10.1080/13675561003712799

Gerke, K., Claus, A., & Mendling, J. (2009). Process mining of RFID-based supply chains. *2009 IEEE Conference on Commerce and Enterprise Computing*, 285-292. IEEE. https://doi.org/10.1109/cec.2009.72

Gevaers, R., Van de Voorde, E., & Vanelslander, T. (2009). Characteristics of innovations in last-mile logistics-using best practices, case studies and making the link with green and sustainable logistics*. Association for European Transport and contributors,* 1-21. Retrieved from www.psu.edu

Gonzalez, H., Jiawei Han, Xiaolei Li, & Klabjan, D. (2006). Warehousing and analyzing massive RFID data sets. *22nd International Conference on Data Engineering (ICDE'06)*, 83-83. IEEE. https://doi.org/10.1109/icde.2006.171

Gou, Z., Tsugawa, S., Korkaew, A., Yamaguchi, S., Khamket, T., Manaskasemsak, B., & Rungsawang, A. (2016). An interest-based tour planning tool by process mining from Twitter. *2016 IEEE 5th Global Conference on Consumer Electronics*, 1-4. IEEE. https://doi.org/10.1109/gcce.2016.7800542

Haigh, K. Z., & Yaman, F. (2011). RECYCLE: Learning looping workflows from annotated traces. *ACM Transactions on Intelligent Systems and Technology, 2*(4), 1-32. http://doi.acm.org/10.1145/1989734.1989746

Huang, Z., Lu, X., & Duan, H. (2011). Mining association rules to support resource allocation in business process management. *Expert Systems with Applications, 38*(8), 9483-9490. https://doi.org/10.1016/j.eswa.2011.01.146

Huang, Z., van der Aalst, W. M., Lu, X., & Duan, H. (2011). Reinforcement learning based resource allocation in business process management. *Data & Knowledge Engineering, 70*(1), 127-145. https://doi.org/10.1016/j.datak.2010.09.002

Hübner, A., Kuhn, H., Wollenburg, J., & Trautrims, A. (2018). From bricks-and-mortar to bricks-and-clicks–logistics networks in omni-channel grocery retailing. *Empirical Studies in Multi-Channel and Omni-Channel Retail Operations and Logistics, 102.* Retrieved from www.sementicscholar.org

Janssenswillen, G., Depaire, B., & Verboven, S. (2017). Detecting train reroutings with process mining. *EURO Journal on Transportation and Logistics, 7*(1), 1-24. https://doi.org/10.1007/s13676-017-0105-8

Jokonowo, B., Sarno, R., Rochimah, S., & Priambodo, B. (2019). Process mining: Measuring key performance indicator container dwell time. *Indonesian Journal of Electrical Engineering and Computer Science, 16*(1), 401-411. https://doi.org/10.11591/ijeecs.v16.i1.pp401-411

Karim, M. A., Samaranayake, P., Smith, A. J. R., & Halgamuge, S. K. (2010). An on-time delivery improvement model for manufacturing organisations. *International Journal of Production Research*, *48*(8), 2373-2394. https://doi.org/10.1080/00207540802642245

Kecman, P., & Goverde, R. M. (2013). Adaptive, data-driven, online prediction of train event times. *16th International IEEE Conference on Intelligent Transportation Systems (ITSC 2013)*, 803-808. IEEE. https://doi.org/10.1109/itsc.2013.6728330

Kin, B., Ambra, T., Verlinde, S., & Macharis, C. (2018). Tackling fragmented last mile deliveries to nanostores by utilizing spare transportation capacity—A simulation study. *Sustainability, 10*(3), 653. https://doi.org/10.3390/su10030653

Krathu, W., Pichler, C., Zapletal, M., & Werthner, H. (2012). Semantic inter-organizational performance analysis using the balanced scorecard methodology. In *2012 proceedings of the 35th international convention mipro* (pp. 1589–1594).

Kumar, M. V. M., Thomas, L., & Annappa, B. (2017). Distilling lasagna from spaghetti processes. In *Proceedings of the 2017 International Conference on Intelligent Systems, Metaheuristics & Swarm Intelligence*, 157-161. ACM Press. https://doi.org/10.1145/3059336.3059362

Langley, C. & Capgemini (2012). 2013 Third-Party Logistics Study: The State of Logistics Outsourcing Results and Findings of the 17th Annual Study. Retrieved from https://www.capgemini.com/wp-content/uploads/2017/07/2013_Third-Party_Logistics_Study.pdf

Langley, C. & Infosys (2019). 2019 Third-party Logistics Study: The State of Logistics Outsourcing: Results and Findings of the 23rd Annual Study. Retrieved from https://www.kornferry.com/content/dam/kornferry/docs/article-migration/2019-3PL-Study.pdf

Lau, H. C., Ho, G. T., Zhao, Y., & Chung, N. S. H. (2009). Development of a process mining system for supporting knowledge discovery in a supply chain network*. International Journal of Production Economics, 122*(1), 176-187. https://doi.org/10.1016/j.ijpe.2009.05.014

Lim, S. F. W., Jin, X., & Srai, J. S. (2018). Consumer-driven e-commerce. *International Journal of Physical Distribution & Logistics Management*. https://doi.org/10.1108/IJPDLM-02-2017-0081

Li, Y., & Deng, S. (2009). Automatic rediscovery and remodeling of logistic process in e-business logistic management. *2009 International Conference on Intelligent Human-Machine Systems and Cybernetics*, 107-111. IEEE. https://doi.org/10.1109/ihmsc.2009.35

Logistics Management (2013) In *Council of Supply Chain Management Professionals Definitions and Glossary of Terms.* Retrieved from https://cscmp.org/CSCMP/Educate/SCM_Definitions_and_Glossary_of_Terms.aspx

Márquez-Chamorro, A. E., Resinas, M., Ruiz-Cortés, A., & Toro, M. (2017). Run-time prediction of business process indicators using evolutionary decision rules. *Expert Systems with Applications, 87*, 1-14. https://doi.org/10.1016/j.eswa.2017.05.069

Mikalef, P., Boura, M., Lekakos, G., & Krogstie, J. (2019). Big data analytics and firm performance: Findings from a mixed-method approach. *Journal of Business Research, 98*, 261-276. https://doi.org/10.1016/j.jbusres.2019.01.044

Milgate, M. (2001). Supply chain complexity and delivery performance: an international exploratory study. *Supply Chain Management, 6*(3), 106-118. https://doi.org/10.1108/13598540110399110

Nakandala, D., Samaranayake, P., & Lau, H. C. (2013). A fuzzy-based decision support model for monitoring on-time delivery performance: A textile industry case study. *European journal of operational research, 225*(3), 507-517. https://doi.org/10.1016/j.ejor.2012.10.010

Olsson, J., Hellström, D., & Pålsson, H. (2019). Framework of last mile logistics research: A systematic review of the literature. *Sustainability, 11*(24), 1-25. https://doi.org/10.3390/su11247131

Process Analytics Factory (2019) PAFnow Companion Setup (version 1.1.5) [Computer software]. Retrieved from https://pafnow.com/

Process Analytics Factory (2019) PAFnow report (version 2020.3.1) [Computer software]. Retrieved from https://pafnow.com/

Pettersen, J. T., & Dahlstedt, F. (2019). *Trend spotting of needs within B2C & B2B goods deliveries in Sweden* (Master's thesis, Chalmers University of Technology, Gothenburg, Sweden). Retrieved from https://odr.chalmers.se/bitstream/20.500.12380/256836/1/256836.pdf

Microsoft (2019) Power BI (desktop version 2.84.701.0) [Computer software]. Retrieved from powerbi.microsoft.com

Pulshashi, I. R., Bae, H., Sutrisnowati, R. A., Yahya, B. N., & Park, S. (2015). Slice and Connect: Tri-Dimensional Process Discovery with Case Study of Port Logistics Process. In *Procedia Computer Science, 72*, 461-468. Elsevier. https://doi.org/10.1016/j.procs.2015.12.127

Repta, D., & Stanescu, A. M. (2017). Holistic Approach for process model discovery. *2017 21st International Conference on Control Systems and Computer Science (CSCS)*, 471-476. IEEE. https://doi.org/10.1109/cscs.2017.73.

Reyes, D., Savelsbergh, M., & Toriello, A. (2017). Vehicle routing with roaming delivery locations. *Transportation Research Part C: Emerging Technologies, 80*, 71-91. https://doi.org/10.1016/j.trc.2017.04.003

Rozinat, A., Wynn, M., van der Aalst, W., ter Hofstede, A., & Fidge, C. (2009). Work-flow simulation for operational decision support. *Data & Knowledge Engineering, 68* (9), 834–850. https://doi.org/10.1016/j.datak.2009.02.014

Rozsnyai, S., Lakshmanan, G. T., Muthusamy, V., Khalaf, R., & Duftler, M. J. (2012). Business process insight: An approach and platform for the discovery and analysis of end-to-end business processes. *2012 annual SRII global conference*. IEEE. https://doi.org/10.1109/srii.2012.20

Saenz, J., Figliozzi, M., & Faulin, J. (2016). Assessment of the carbon footprint reductions of tricycle logistics services. *Transportation Research Record, 2570*, 48-56. https://doi.org/10.3141/2570-06

Soffer, P., Hinze, A., Koschmider, A., Ziekow, H., Di Ciccio, C., Koldehofe, B., ... & Song, W. (2019). From event streams to process models and back: Challenges and opportunities. *Information Systems, 81*, 181-200. https://doi.org/10.1016/j.is.2017.11.002

Sohn, J. I., Woo, S. H., & Kim, T. W. (2017). Assessment of logistics service quality using the Kano model in a logistics-triadic relationship. *The International Journal of Logistics Management, 28*(2), 680-698. https://doi.org/10.1108/IJLM-09-2015-0172

Song, M., & Van der Aalst, W. M. (2008). Towards comprehensive support for organizational mining. *Decision Support Systems, 46*(1), 300-317. https://doi.org/10.1016/j.dss.2008.07.002

Song, W., Xia, X., Jacobsen, H. A., Zhang, P., & Hu, H. (2016). Efficient alignment between event logs and process models. IEEE Transactions on Services Computing, 10(1), 136-149. Retrieved from www.ieee.org

Stuckey, H. L. (2015). The second step in data analysis: Coding qualitative research data. *Journal of Social Health and Diabetes, 3*(1), 007-010. https://doi.org/10.4103/2321-0656.140875

Sutrisnowati, R. A., Bae, H., Park, J., & Ha, B.-H. (2013). Learning bayesian network from event logs using mutual information test. *2013 IEEE 6th international conference on service-oriented computing and applications*. IEEE. https://doi.org/10.1109/soca. 2013.38

Sutrisnowati, R. A., Bae, H., & Song, M. (2015). Bayesian network construction from event log for lateness analysis in port logistics. *Computers & Industrial Engineering, 89*, 53–66. https://doi.org/10.1016/j.cie.2014.11.003

Sutrisnowati, R. A., Yahya, B. N., Bae, H., Pulshashi, I. R., & Adi, T. N. (2017). Scalable indexing algorithm for multi-dimensional time-gap analysis with distributed computing. *Procedia Computer Science, 124*, 224-231. https://doi.org/10.1016/j.procs.2017.12.150

Symerio Revision (2018) pgeocode (version 0.3.0) [Computer software] Retrieved from https://pgeocode.readthedocs.io/en/latest/

Vakulenko, Y., Shams, P., Hellström, D., & Hjort, K. (2019). Service innovation in e-commerce last mile delivery: Mapping the e-customer journey. *Journal of Business Research, 101*, 461-468. https://doi.org/10.1016/j.jbusres.2019.01.016

van Beest, N. R., Dumas, M., García-Bañuelos, L., & La Rosa, M. (2016). Log delta analysis: Interpretable differencing of business process event logs. In H. R. Motahardi-Nezhad, J. Recker, & M. Weidlich (Eds.), *Business Process Management* (pp. 386-405). Cham, Switzerland: Springer International Publishing AG. https://doi.org/10.1007/ 978-3-319-23063-4_26

van Cruchten, R. R., & Weigand, H. H. (2018). Process mining in logistics: The need for rule-based data abstraction. *2018 12th International Conference on Research Challenges in Information Science (RCIS)*, 1-9. IEEE. https://doi.org/10.1109/RCIS.2018.8406653

Van der Aalst, W. M., Schonenberg, M. H., & Song, M. (2011). Time prediction based on process mining. *Information systems, 36*(2), 450-475. https://doi.org/10.1016/j.is.2010.09.001

Van der Aalst, W. M. (2011). *Process mining: Discovery, conformance and enhancement of business processes* (Vol. 2). Heidelberg, Berlin: Springer. https://doi.org/10.1007/978-3-642-19345-3

Van der Aalst, W. M. (2012) *Event log structure.* [Image] https://doi.org/10.1145/2229156.2229157

Van der Aalst, W. M. (2012). Process mining: Overview and opportunities. *ACM Transactions on Management Information Systems (TMIS), 3*(2), 1-17. https://doi.org/10.1145/2229156.2229157

Van der Aalst, W. M. (2016). *Process mining: Data science in action* (2nd ed.). Heidelberg, Berlin: Springer. https://doi.org/10.1007/978-3-662-49851-4

Vasyutynskyy, V., Gellrich, A., Kabitzsch, K., & Wustmann, D. (2010). Analysis of internal logistic systems based on event logs. *2010 IEEE 15th Conference on Emerging Technologies & Factory Automation (ETFA 2010)*, 1-8. IEEE. https://doi.org/10.1109/ETFA.2010.5641202

Vasilyev, E., Ferreira, D. R., & Iijima, J. (2013). Using inductive reasoning to find the cause of process delays. *2013 IEEE 15th conference on business informatics*, 242-249. IEEE. https://doi.org/10.1109/CBI.2013.41

Wang, J., Zhu, B., Wang, Y., & Huang, L. (2016). Mining organizational behaviors in collaborative logistics chain: An empirical study in a port*. 2016 International Conference on Logistics, Informatics and Service Sciences (LISS)*, 1-5. IEEE. https://doi.org/10.1109/ liss.2016.7854453.

Wang, Y., Caron, F., Vanthienen, J., Huang, L., & Guo, Y. (2014). Acquiring logistics process intelligence: Methodology and an application for a Chinese bulk port. *Expert Systems with Applications, 41*(1), 195-209. https://doi.org/10.1016/j.eswa.2013.07.021

Zhao, W., Zeng, Q., Zheng, G., & Yang, L. (2017). The resource allocation model for multi-process instances based on particle swarm optimization. *Information Systems Frontiers, 19*(5), 1057-1066. https://doi.org/10.1007/s10796-017-9743-5

Appendix A

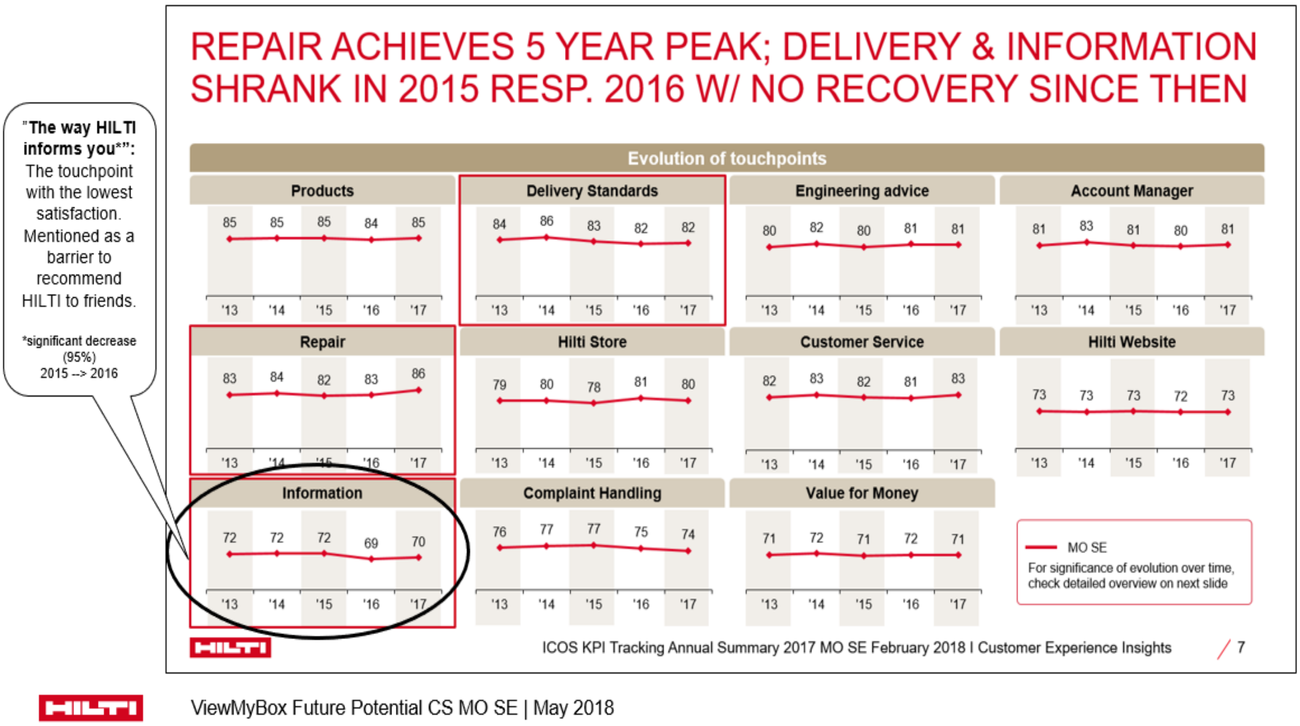

Figure 25. Overview of feedback of customers per topic

Appendix B

Global process managers

Oliver Weich; Distribution and Carrier integration

- Improvements: make the process more flexible, empower the customer to detail the order. Right now, not even Hilti can change anything once the order has left the warehouse.

- HS codes are not a priority for the carriers, and we pay good money for the service. It is better to do the mapping ourselves, this could gain Hilti control over the data and reduction of costs. It is easy to implement because of stable relations with carriers. IT sees more roadblocks.

- Use machine learning to identify errors early on and notify the transport manager; example, different depots [hubs] have different performance levels,

- The carrier could be interested in forecasting data and customer data, but not very likely.

Bogdan Raykov; Project business

- Could be interesting to see which project businesses are delivered on-time.

Johan De-Smedt; Thesis supervisor

- No alert system at the moment so dependent on the availability and willingness of Hilti employees.
- The order by which the parcels enter the carrier’s truck is determined by Hilti. If it can be shown that certain zip-codes are tricky to reach, they could be put in the truck last, so they are sorted first in the depot [hub] and have a higher chance of reaching
- Collaborate with carriers, the mutual dependency stays. Create transparency, bonus/malus won’t get you there.

- Collaborate with carriers, the mutual dependency stays. Create transparency, bonus/malus won’t get you there.
- The order by which the parcels enter the carrier’s truck is determined by Hilti. If it can be shown that certain zip-codes are tricky to reach, they could be put in the truck last, so they are sorted first in the depot [hub] and have a higher chance of reaching

Business Analyst

Ralf Eggenberger

- Another option is to look at the location, where do large problems occur. Perhaps you can see which hubs are often resulting in problems à perform clustering. See if you can make a link to the CCN.

- Possible directions: try to create a relationship between the variable time and other variables. See if you can discover patterns.

Head of transport regions / Transport managers

Thomas Krohn

- Opportunity: Simply to know where the goods are. Did the truck leave on-time, does the physical load match the paperwork of its content?
- Checks if everything leaves on-time from the depots [hubs].
- Where are the bottlenecks in the carrier network? How good is the network of the carrier?
- Did the goods arrive at the final depot [hub] before the last-mile of that day was run? If not, you know they are going to be too late.
- An example where TRR helps; Carrier Emons delivered half of the goods but checked the whole delivery as done. TRR gives insight in proving this and helps to show the carrier what can be improved
- Another example; showing that the carrier was later than the cut of point of the pre-time delivery. Now no need to pay the extra fee.

Marco Battista

- It would be useful to have the CCN number or code there to know what is going on and if the exception has been handled currently yet
- When an HS code is generated, there could be an automated action following it or help the region manager to take proactive measures. Sometimes emails going to customers, sometimes an alert an important customer will get his tools late, etc.
- There is no notification when the carrier missed his last-mile window. Even though it is certain that the last mile will then be covered another day, meaning the parcel will be late. When this is certain, Hilti should receive an update.

Olivier Haas

- Would like to have a way of checking if the data of the carrier are correct, especially in the tool service pickups where they are experiencing a lot of problems in the customer saying one thing and the carrier saying another.
- No updates or warning when deliveries are going to be late
- Tracking of important deliveries, communication with the sales team on alerting customers and communicating with carriers about late delivery is all done manually
- Info from customers on availability could also be used by transport to make delivery attempts more efficient

Transport specialists

Alejandro Guerra

- There are lots of unexploited opportunities with power BI; apps, AI, Microsoft flow
- Prediction may also help to plan the logistic process better

Luca Mandelli

- Italy uses 1 carrier (BRT) for 98% of deliveries. 6m channels are given to SpeedItalia who covers 1-2% of total deliveries. SpeedItalia is not integrated so no HS codes. BRT (98%) is integrated but very high % of delays not specified à No root cause analysis possible

Alexis Yon

- NTLS is highly dependent on Hilti as a customer and has a very high reliability (99%). Since Hilti is 70% of their income, they have to perform well.

Gerrit Budde

- Currently no capacity to implement TRR/analysis possibility and all carriers.
- Interested in where there are lags between events and timestamps of events to validate KPI's. Could check it out by following one parcel ourselves or borrowing a scanner from one of the carriers.
- Once in every one or two months, the integrated carrier and Hilti discuss the delivery results. After the TRR, this includes Hilti generated KPI's based on carrier data. Here, Hilti challenges the carrier on their data and their delivery performance
- The more transparent we get the carrier, the more of a priority Hilti becomes for that carrier
- Hilti produced reports have to be easy to understand in order to be used in challenging the carrier

Appendix C

Human resources

Opportunity: Hire dedicated Hilti personnel to manage a carrier from a TRR/data perspective and improve implementation if the TRR concept (Gerrit Budde)

Customer interaction

Opportunity: It would be useful to have the Customer complaint number matched with an error code, to know what is going on and if the exception has currently been handled. (Marco Battista)

Opportunity: Make the process more flexible, empower the customer to detail the order. (Oliver Weich)

Data quality

Opportunity: Knowing what data are reliable helps us use it more (Oliver Haas)

Opportunity: Detailing errors gives more insight into root-problems (Luca Mandelli)

Opportunity: Doing the mapping from carrier codes to the HS-system could gain Hilti Control over the data and would reduce costs (Oliver Weich)

Live Overview

Opportunity: Having a live overview of parcels in the carrier network would make managing more effective. (Thomas Krohn)

Opportunity: Knowing the timeliness of the data might help in creating a semi-real time data overview (Gerrit Budde)

Notifications

Opportunity: When an HS code is generated, there could be an automated action following it helping the region manager to take proactive measures (Marco Battista, Johan De-Smedt, Olivier Haas)

Opportunity: Use the data to check if everything leaves on-time from the depots[hubs], and receive a notification otherwise. (Thomas Krohn)

Opportunity: Logistical data could be used to see if a parcel is able to make it to the customer and Hilti could receive an update if the parcel will be too late.
(Marco Battista, Thomas Krohn, Oliver Weich, Oliver Haas)

Data analyses

Opportunity: look at the location, where do large problems occur. Perhaps you can see which hubs are often resulting in problems (Ralf Eggenberger, Johan De-Smedt)

Opportunity: Try to create a relationship between the variable time and other variables. See if you can discover patterns. (Ralf Eggenberger, Johan De Smedt)

Opportunity: Use the data to check where the bottlenecks in the carrier network are present and to evaluate the quality of the carrier network. (Thomas Krohn)

Opportunity: Could be interesting to see which project businesses are delivered on-time (Bogdan Raykov)

Opportunity: There are lots of unexploited opportunities with power BI; apps, AI, Microsoft flow (Alejandro Guerra)

Motivating the Carrier to perform better

Opportunity: Collaborate with carriers. Hilti and the carrier are dependent on each other, so collaboration works best. Create transparency, bonus/malus won't get you there. (Johan De Smedt)

Opportunity: A carrier that strongly depends on Hilti is much more likely to have high performance (Alexis Yon)

Opportunity: Cases of discrepancies between the data and the carrier report are discussed with the carrier every 1-2 months. This keeps carriers sharp. (Thomas Krohn, Gerrit Budde)

Opportunity: The more transparent we get the carrier by using the data, the more of a priority Hilti becomes for that carrier. (Gerrit Budde, Johan de Smedt)

Opportunity: If it shows from the data that the carrier was too late at the customer, there is no need for Hilti to pay the extra fee. (Thomas Krohn)

Collaboration in the last mile

Opportunity: The order by which the parcels enter the carrier's truck is determined by Hilti. If it can be shown that certain zip-codes are tricky to reach, they could be put in the truck in such a way so they are sorted first in the depot [hub] and have a higher chance of reaching the customer. (Johan de Smedt)

Opportunity: Making Hilti reports easy to understand and mapped to carrier codes allows for easy comparison to the carrier reports(Gerrit Budde)

Opportunity: Info from customers on availability or sales prediction could also be used by transport to make delivery attempts more efficient (Oliver Haas, Alejandro Guerra).
Side-note: This could be interesting for the long term, but currently this is not on carriers' horizon. (Oliver Weich)

Appendix D

Table 2. Selected Case attributes

Event log data	HU number	A unique number assigned to each parcel by Hilti
Delivery data	c_Carrier	A number detailing which carrier executed the delivery. There is only one carrier involved per HU, but multiple carriers may execute one delivery
	c_Plant	The Hilti Warehouse (also called Plant) from which the parcel departed
	c_Delivery_Number	A unique number attached to every single delivery. A delivery can contain multiple parcels and therefore multiple HU numbers
Computed data	c_on-time-bool	A Boolean indicating if the delivery was delivered on-time, from a carrier perspective
	c_Carrier_Depot_ Identifier	The ID number of the final hub of the carrier. This is determined based on a mapping table where postal codes are linked to carrier hubs. The maps need to be manually maintained. Therefore, changes in the carrier's network may result in the wrong hub.
	c_StartDate	The starting date of the delivery
	c_EndDate	The ending date of the delivery
	c_Datediff	The total amount of days between the start and ending date
	c_Workdaysdiff	The number of working days between the start and ending date
	c_Error_case	A Boolean stating if an error code was received for this delivery at some point
	c_Overtheweekend Bool	A Boolean which is positive when the delivery has crossed over the weekend (Saturday and Sunday)
	c_latest_confirmation_ to_deadline	The time difference between the last confirmation of delivery and the deadline. If the number returned is positive, there was time to spare and the delivery was on-time. If the number is negative, the delivery arrived after the deadline
	c_first_confirmation_ to_deadline	The time difference between the first confirmation of delivery and the deadline. If the number returned is positive, there was time to spare and the delivery was on-time. If the number is negative, the delivery arrived after the deadline

Table 3 Selected Event attributes

Event log data	HS-code	The HS-code, which is the event code transformed to Hilti's HS-code system
	Timestamp	The timestamp of the activity
Interpretation of event	HS-code descripion	The description of the HS-code
	HSCodeMomentDesc	The classification of which sub-process of the last mile the HS-code belongs to
	HSCodeTypeDesc	The type of message of the HS-code, which can be 'Success', 'Error' or 'Informative'