Assessing the use of Process Mining techniques to monitor the work process of Commercial Drivers

Jennifer Cutinha
University of Twente
PO Box 217, 7500 AE Enschede
the Netherlands

jenniferalicecutinha@student.u
twente.nl

ABSTRACT

Process mining techniques have been applied to practical use cases within several domains. However, very few studies demonstrate the applicability of process mining within the logistics domain. Therefore, this research aims to apply process mining techniques to monitor the work process of drivers at a logistics and transportation company. The case study is carried through the steps mentioned in the PM^2 framework. The results obtained through process mining techniques are rendered inconclusive due to missing data and data quality issues. To solve this, we introduce a novel approach by positioning process mining within an industry data-sharing model, namely the OTM. The results from the case study are relevant to the logistics industry and the use of OTM for process mining can be used for further knowledge development.

Keywords

Process mining, Logistics, data-sharing, OTM

1. INTRODUCTION

1.1 Background

Process mining focuses on extracting knowledge from data generated and stored in the databases of information systems to build event logs [11]. An event log can be seen as a collection of cases and a case can have multiple traces or sequences of events. The event log should contain four minimal requirements namely, the case id which represents the process instance, the names of the events or activities in the process, the timestamps of the events, and the resource that conducted the activity. There are three main types of process mining: process discovery, conformance checking, and enhancement. Process discovery allows process models to be extracted from an event log; conformance checking monitors deviations by comparing a given model with the event log; and enhancement extends or improves an existing process model using information about the actual process recorded in the event log [11]. In this research, process mining will be studied in the logistics domain. Logistics is the process of planning and executing the efficient transportation and storage of goods from the point of origin to the point of consumption [12].

1.2 Previous Work and Research Contribution

The application of process mining techniques in different domains is quite prevalent already in existing literature. In a case study by Accorsi et al. [2], conformance checking techniques were deployed for security auditing in a bank scenario for a real-life loan application process. In the paper of Jans et al. [6], they conclude through a case study at a bank, that process mining of event logs is valuable to auditors when used as an analytical procedure. Wang et al. [15], proposed a compliance monitoring framework that was tested on customs supervision in the Netherlands. In another paper by Jans et al. [5], process mining was applied to the procurement process at a company for the mitigation of internal transaction fraud. However, few studies study the applicability of process mining in logistics. The application of process mining within the logistics domain has not been elaborated fully or have been part of larger studies. For example, in the case study by Wang et al. [16], process mining techniques were applied to the inward cargo handling process at a prominent Chinese bulk port. Additionally, Becker et al. [12] used real-life manufacturing data and applied Markov Chains as a sequence clustering technique to improve the quality of process models discovered. Increasing the number of in-depth case studies within the logistics domain is especially important since logistics processes are highly complex and dynamic, and data can come from heterogeneous data sources and be unstructured [4].

In this research, a case-study is conducted at a logistics and transportation company to discover and monitor the work process of commercial drivers. The company will be addressed as ABC though the course of the paper. The work process of drivers as a use case for process mining has not been investigated in existing literature, and therefore, this research deviates from existing studies in this regard. Additionally, through the insights gained from the case study, we introduce a novel approach to process mining within the context of a data-sharing model, namely the OpenTripModel (OTM) which is a standard being adopted by the Dutch logistics industry. Therefore, the contribution of this study is twofold - a case study, whose insights are relevant for the logistics industry; and the introduction of process mining within OTM which is valuable to a broader audience and can be used for further knowledge or research development.

1.3 Problem Context and Case Definition

Monitoring the activities of commercial drivers constitutes a major part of the process of fleet management. Many commercially available software offer real-time monitoring solutions for fleet managers to track the working and rest periods of drivers, and their whereabouts. However, only fleet managers are aware of the whereabouts of drivers. Furthermore, at ABC, the drivers and fleet managers are spread across 4 different countries – Netherlands, Germany, Poland, and the Czech Republic. Higher management at ABC would like to obtain an aggregated view of the working process of drivers and their performance.

2. RESEARCH QUESTIONS

The prime focus of this research is the case study, and accordingly, we have defined the following research questions:

PRQ: To what extent can process mining be used to help higher management at ABC monitor and improve the working process of drivers?

To answer the primary research question, the following sub-research questions are proposed:

RQ1: What process mining techniques and tools are suitable for achieving this purpose?

RQ2: What are the results obtained upon the application of process mining techniques?

RQ3: Do the results obtained from process mining techniques provide new knowledge to higher management? If so, what improvements can be made based on this new knowledge?

3. METHODOLOGY

To guide the process mining project, the PM^2 framework by van Ech et al. [14] has been used. This methodology has been selected because it supports iterative analysis which is suitable for more complex projects where domain knowledge is required throughout the analysis. This is not supported by other process mining methodologies such as Process Diagnostics Method (PDM) [3] or the L* lifecycle model [1]. A pictorial representation of this framework has been included in the appendix for readability (appx. A, figure 1). The results of this case study will be presented following this framework.

4. RESULTS

4.1 PM^2 Methodology

4.1.1 Stage 1 & 2: Planning and Extraction

In the planning stage, the research questions addressed in section 2 were formulated, and accordingly, the data was extracted. In the extraction stage, the required data was extracted by a business intelligence expert at ABC. The resulting CSV file contained a total of 359,604 logistics events relating to shipments carried out for a period of four months, from July 2018 to October 2018. The data set consisted of a list of logistics activities or events with the timestamps that happened for a shipment. This included activities such as loading, unloading, driving, refueling, etc. In addition to these data fields, the driver who performed the logistic activity, and the username of the responsible fleet manager (CorrectionUser) was also included. An illustration of the source file is depicted in Table 1. In process mining, the ShipmentID, Activity, Start/End, DriverID represent the caseid, event, timestamp, and resource, respectively. The table does not contain true information to maintain anonymity.

4.1.2 Stage 3: Data processing

For several events in the original file, the ShipmentID was null and these were filtered out. This resulted in 153,878 events or 45% of the original dataset. Furthermore, there are 12 different

Table 1. Illustration of the source data set

ShipmentID	Activity	Start	End	DriverID	CorrectionUser
SHP1234	Arrival Loading	7/4/2018 8:51:00	7/4/2018 9:15:00	345	Switt
SHP4536	Refueling	7/15/2018 13:17:00	7/15/2018 13:31:00	543	Rdul
SHP1234	Loading	7/4/2018 9:15:00	7/4/2018 9:35:00	345	Switt
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.

activities that happen for a shipment. A shipment should contain at least half of these activities. Any shipment that contained less than 6 activities would mean missing information, and therefore these rows were filtered out using Python. This resulted in 43,544 events or 12% of the dataset. Another issue that was observed in the dataset was overlapping time intervals for an event. This is illustrated in Figure 1 along with the transformation made. This was achieved by creating a macro in excel.
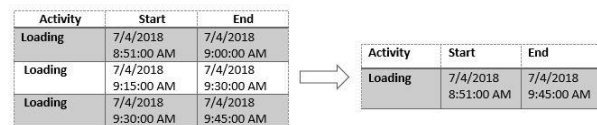

Figure 1. Transformation for overlapping time intervals

Lastly, to make country-wise comparisons, the data set was divided into four subsets representing drivers managed from Germany, the Netherlands. Poland and the Czech Republic based on the last column in Table 1.

4.1.3 Stage 4: Mining and Analysis

4.1.3.1 Intermediate findings

The first step in process mining is the process discovery phase. This step produces the models from the event data. Table 2 indicates the basic case statistics per country.

Table 2. Case Statistics per country

Statistics	NL	DE	CZ	PL
No of cases or shipments	67	258	512	927
Total No of events	1,006	6,358	10,388	21,398
Total No of Activities	12	12	12	12

Each country displays a total of 12 activities. These are connect, arrival loading, loading, waiting, traffic jam, refueling, driving, rest, arrival unloading, unloading, standstill, and disconnect. Arrival loading/ arrival unloading is the time spent at the loading/unloading location, but no loading/unloading has begun yet. Standstill indicates periods of vehicle inactivity. Rest refers to the rest periods according to the EU regulation that the driver is obligated to take in between continuous driving and at the end of work. Sometimes, the logistics activities are carried out by 2 or more drivers. For example, loading for a shipment could be carried out by one driver and the unloading for that shipment by another. This is represented by connect and disconnect. Figure 2 represents the process model for Poland generated by the process mining tool Apromore. Similar models were created for other countries (appx. B).

The second step of process mining is conformance checking. Here, a pre-defined process model usually in the form of a Petri Net or a BPMN model is compared to the event log to see if the modeled behavior reflects the behavior of the event log. A BPMN model was created on the flow of logistics activities in the Netherlands. To perform conformance checking, the ProM
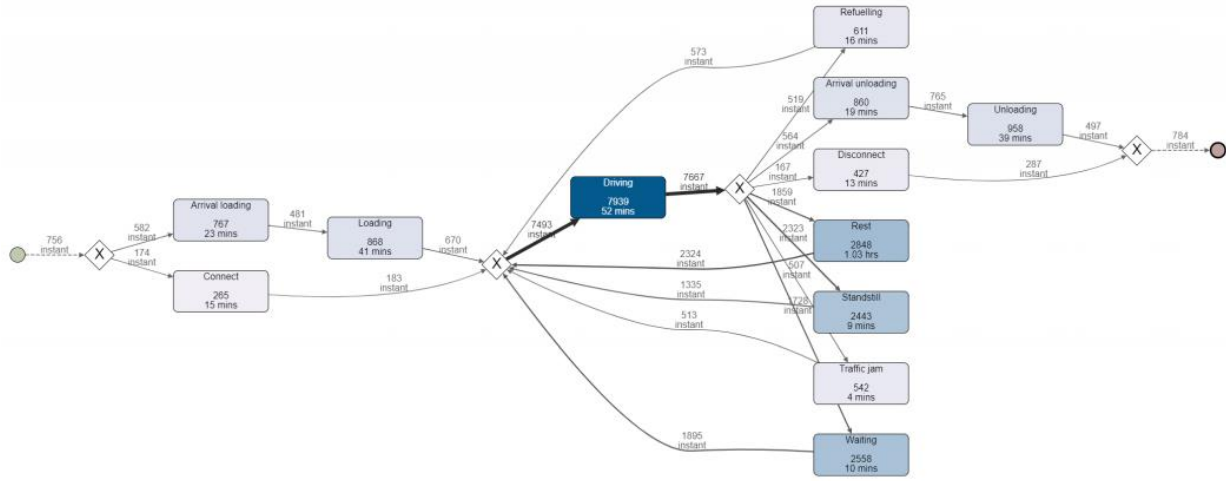
Figure 2. A process model for Poland generated through Apromore

plugin replay log on Petri net was used to determine the fitness. Fitness is a quality dimension that determines how well the event log can be replayed on the model. This revealed a very low fitness metric of 0.39. This is expected since, although parts of activities in the process models may appear to have some sequence upon discovery, they are not necessarily sequential in terms of the control-flow perspective. Therefore, this was not computed for other models.

The performance of drivers per country can be determined by the amount of time that drivers from each country take to perform logistics activities. ABC has determined a set of advised times for logistic activities. The median duration of the logistic activity in the process model gives an approximation of the actual time taken in reality by drivers per country to perform the activities. Table 3 indicates the advised times as well as the median duration of the main logistics activities performed by drivers per country. Another approach would be to use the compliance checking plugins in ProM 6, namely, the elicit compliance rule and configure compliance rule plugins. This would determine the exact number of cases where the drivers did not meet the requirements mentioned in Table 3. However, these plugins are still in the experimental phase and are, therefore, not used.

Table 3. Advised duration and actual durations of logistics activities per country

Activity	Advised Duration	NL	DE	PL	CZ
Arrival Loading	5 mins	45 mins	14 mins	23 mins	21 mins
Loading	45 mins	4 mins	39 mins	41 mins	43.5 mins
Arrival Unloading	5 mins	55 mins	19 mins	19 mins	21 mins
Unloading	45 mins	3 mins	39 mins	39 mins	36.5 mins
Refueling	10 mins	15 mins	15 mins	15 mins	45 mins
Waiting	5 mins	35 mins	7 mins	10 mins	18 mins

4.1.3.2 Main findings
Loading and unloading are supposed to follow arrival loading and arrival unloading, respectively. However, from Figure 2, it is evident that all arrival loading/arrival unloading instances are not followed by loading/ unloading. Also, the number of arrival loading/arrival unloading activities are less than that of loading/unloading, when this is expected to be equal. This is observed in all countries (appx. B). Although the actual duration of the activities carried out by the drivers exceeds the advised duration, they are comparable across countries. However, longer durations of arrival loading/ arrival unloading than the actual loading/ unloading time are observed in the Netherlands. These findings are evaluated in the next stage.

4.1.4 Stage 5: Evaluation
The generated process models were shown to various professionals at ABC. However, for determining validity, the main findings were discussed mainly with three people through video meetings ranging from 30-60 minutes. A table indicating the feedback can be found in the appendix (appx. C, table 1).

4.1.5 Reiteration of Stage 4
Having received feedback from various professionals, the following could be deduced:
- Presence of data loss due to parts of information being captured by different systems (events are captured in the board computer of the truck, while shipment data is captured in another system)
- Anomalous reporting behavior (an arrival loading activity could mean that a loading took place or only a loading activity was indicated by the driver in the board computer)

Moreover, the results were validated using various tools, namely Disco, Apromore, and ProM 6 (discussed in sect. 5).

4.1.6 Stage 6: Process Improvement and Support
Bi-weekly meetings were held at ABC for the discussion of progress and results. This research was able to provide ABC insight into how process mining techniques can be utilized for other business functions as well. However, the underlying issues presented in section 4.1.5 result in inaccurate conclusions. Drivers are trained to record events on the board computer so that fleet managers can monitor their whereabouts. But a lack of a centralized standard of what should and should not be recorded means that this is done differently in different countries. Sometimes, it may also not be necessary to record an event (appx. C, table 1, PL fleetmanager). However, from a data analysis perspective, this is the difference between a correct and incorrect interpretation of analytic results. Therefore, drivers and fleet managers need to be made aware of recording data.

4.2 Need for a logistics data sharing standard and OpenTripModel
The two issues addressed in section 4.1.5 are addressed in the four levels of big data interoperability in the paper of Singh et al. [13]. The logistics sector can be referred to as a network where multiple organizations come together. These can be external organizations, for example, shippers, authorities, transporters, or carriers (Figure 3). But it could also mean multiple parties within an organization spread across different countries, business units, etc. as was seen in the case of ABC. All these parties work together with parts of information being distributed across various systems. Therefore, when data is aggregated from different systems for analysis, data loss and
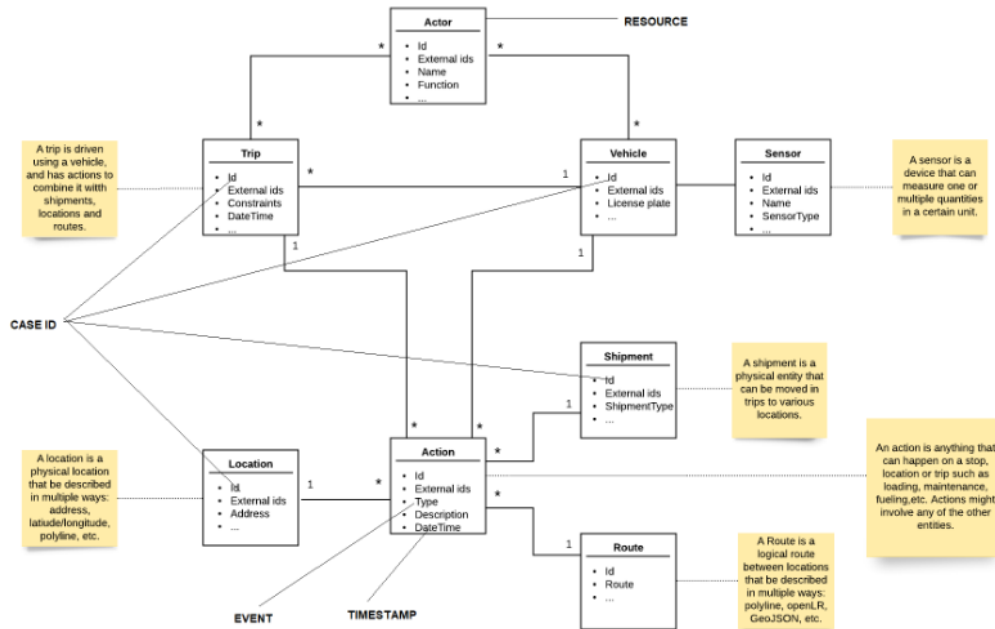
Figure 5. Adapted OTM model [10] linked to the minimum process mining requirements

quality issues are obtained. This was observed in the case-study where the dataset contained information from distributed systems, namely the transport management system (TMS) and the fleet management system (FMS).
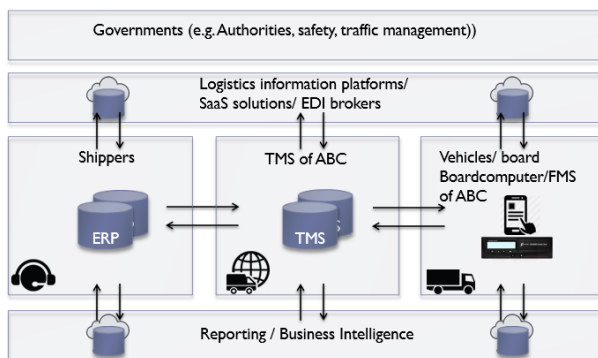

Figure 3. Interaction of multiple logistics parties (adapted from SUTC presentation)

OpenTripModel is a license-free flexible data-sharing model that allows for the uniform and consistent exchange of information across various information systems. This model is managed by Stichting Uniforme Transport Code (SUTC) and its goal is to help logistics companies in the Netherlands to share data efficiently.[1] In the paper of Lont et al. [7], it was shown through a case study that different devices can be linked to this data model, eliminating issues of interoperability between systems. In the industry, this model has been used for many Business Intelligence (BI) use cases, however not for process mining. In this paper, we want to show how this model can be suitable for process mining.

Figure 4 shows the construction of the data model. Entities (represented in yellow) within this model are used to handle various parts within a logistics process. Events indicate the relationships between entities. The order of these events indicates the workflow over time, and this is depicted by the lifecycle (represented in green). Figure 5 shows how this model can be mapped to the minimum process mining requirements. A

small experiment (appx D) has been conducted using a small OTM scenario [8]. The lifecycles in Figure 5 can be used for various process mining phases. For instance, planned and realized events can be used for discovery and conformance checking, the actual phase can be used for real-time process mining and the projected phase can be simulated to produce model enhancements. This model can be used to represent many process mining use cases relevant to various stakeholders in the logistics industry (Table 4). It is to be noted that these use cases are not exhaustive and can be expanded.
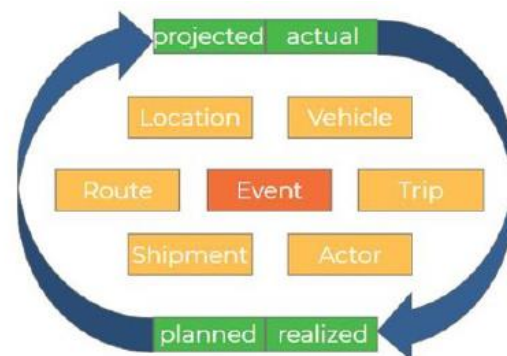

Figure 4. The OpenTripModel [9]

5. DISCUSSION

The issues of missing data and poor data quality from the case study resonate with existing process mining literature. As a consequence, various tools have been built in ProM to eliminate the effect of noisy data and determine the actual control-flow of a process. However, the case study presented in this paper investigates a process whose activities are not sequential, and therefore, tries to focus on the performance perspective rather than the control-flow perspective. Missing data has a significant effect on the performance since the results of the data reflect the performance of only a small percentage of reality. This is exemplified when data quality issues arise. Existing studies try to increase the accuracy of process mining techniques despite noisy data. However, we think that effort should also be made to address the root cause. Therefore, this research proceeds differently by attempting to position process mining within an industry data-sharing standard.

---

[1] https://dutchmobilityinnovations.com/spaces/1168/connected-transport-corridors/files/27520/algemene-presentatie-sutc-pptx?version=1

Lifecycle Phase in the OTM	Shipper	Transporter/Carrier	Authorities
Realized/Planned	Trip, Shipment, Event, Location, Actor, Vehicle		Vehicle, Actor, Event
	Analyzing the ordering process, track& trace event	On-time analysis of Shipments, loading & unloading norms, customer service level monitoring Comparative process mining to determine fast/slow trips on a route Comparing route to actual position	Compliance with EU driving & rest time regulations
	Actor, Shipment, Event	Vehicle, Event, Actor	
	Billing and invoicing process	Performance comparisons of fleet, drivers, employees per business unit, country etc.	
		Shipment, Actor, Event	
		Customer damages, fines, and claim handling	
		Vehicle, Actor, Event	
		Idle waiting times of vehicles for determining CO_2 emissions.	
Actual		Shipment, Actor, Event, Location, Vehicle	
		Real-time monitoring of shipments	
Projected			Location, Event
			Mapping of logistic traffic data

Table 4. List of process mining use cases per stakeholder linked to OTM entities in the different lifecycle phases

Also, in this research, different process mining tools have been used, namely Disco, Apromore, and ProM 6. The prime reason for this was to determine the validity of the results. For process discovery, Disco's interface seemed intuitive and easy to understand, especially by business professionals at ABC. Apromore is similar to Disco, however, it has an extra feature to view models in BPMN. An effort was also made to produce models in ProM 6 using the alpha miner and the inductive miner. However, a 'flow' could not be identified. To determine the performance, only ProM 6 had the required plugins. However, they were in the experimental phase (sect.4.1.3.1).

Finally, The PM^2 methodology proved to be very useful in the stepwise execution of the case study.

6. CONCLUSION

In this paper, we conducted a case study where we aimed to show higher management of ABC the overall performance of drivers per country. However, the two main issues of data loss and anomalous reporting behavior rendered the process mining results from a performance perspective inconclusive. Higher management at ABC was, however, able to visualize the work process of drivers in terms of the control-flow perspective. To achieve this, we used various tools to determine the validity of the results. Logistics data provides a rich source to conduct process mining. However, logistics processes involve multiple parties with data being spread over multiple information systems resulting in interoperability issues. Data sharing through the OTM standard can tackle these interoperability issues and therefore add value to results obtained through process mining techniques. Therefore, we introduced OTM in the context of process mining. We mapped the minimum process mining requirements to the OTM model, conducted a small experiment based on an event log in the OTM format, and explored various generalized process mining use cases linked to the OTM that could be relevant for different parties in the logistics sector. Further directions for research could include applying process mining techniques to these explored uses cases in companies or organizations that implement the OTM model. A comparison-based study, in this regard involving organizations that implement and do not implement the OTM, would be an interesting follow-up.

7. REFERENCES

[1] Aalst, Wil et al. (2011). Process Mining Manifesto. Lecture Notes in Business Information Processing. 99. 169-194. DOI = https://doi.org/10.1007/978-3-642-28108-2_19.

[2] Accorsi, R., and T. Stocker (2012). On the exploitation of process mining for security audits: the conformance checking case. Proceedings of the 27th Annual ACM Symposium on Applied Computing. Trento, Italy, Association for Computing Machinery: 1709–1716. DOI=https://doi.org/10.1145/2245276.2232051

[3] Bozkaya, M., Gabriels, J., Werf, J.: Process Diagnostics: A Method Based on Process Mining. In: Information, Process, and Knowledge Management, 2009. eKNOW'09. International Conference on. pp. 22–27. IEEE (2009). DOI= https://doi.org/10.1109/eKNOW.2009.29

[4] Intayoad, W., and T. Becker (2018). Applying Process Mining in Manufacturing and Logistics for Large Transaction Data, Cham, Springer International Publishing. DOI = https://doi.org/10.1007/978-3-319-74225-0_51

[5] Jans, M., et al. (2011). "A business process mining application for internal transaction fraud mitigation." Expert Systems with Applications 38(10): 13351-13359. DOI= https://doi.org/10.1016/j.eswa.2011.04.159

[6] Jans, M., et al. (2014). "A Field Study on the Use of Process Mining of Event Logs as an Analytical Procedure in Auditing." The Accounting Review 89: 1751-1773. DOI= https://doi.org/10.2308/accr-50807

[7] Lont, Y., van Duin, R., Jens, D-P., & van Lier, B. (2018). Demasking the black hole of transportation? A blocking road naar de ontwikkeling van een CO2 Blockchain-georienteerde (product) app. In Bijdragen vervoerslogistieke werkdagen 2018 (pp. 1-12). University Press Zelzate.

[8] OpenTripModel walkthrough. (2018). https://www.opentripmodel.org/docs/walkthrough

[9] OpenTripModel. (2018). https://developer.opentripmodel.org/#

[10] The OTM5 model. (2019). OpenTripModel. https://www.opentripmodel.org/v5.0/docs/the-otm5-model

[11] Rojas, E., Munoz-Gama, J., Sepúlveda, M., & Capurro, D. (2016). Process mining in healthcare: A literature review. *Journal of Biomedical Informatics, 61*, 224-236. DOI= https://doi.org/10.1016/j.jbi.2016.04.007

[12] Rouse, M. (2019). What is Logistics? Retrieved from https://searcherp.techtarget.com/definition/logistics

[13] Singh, P. M., & van Sinderen, M. (2016). Big data interoperability challenges for logistics. In M. Zelm, G. Doumeingts, & J. P. Mendonca (Eds.), *Enterprise interoperability in the digitized and networked factory of the future* (pp. 325-335). ISTE Press.

[14] van Eck, M. L., et al. (2015). PM2: A Process Mining Project Methodology, Cham, Springer International Publishing. DOI = https://doi.org/10.1007/978-3-319-19069-3_19

[15] Wang, Y. X., et al. (2018). Regulatory Supervision with Computational Audit in International Supply Chains. New York, Assoc Computing Machinery. DOI = https://doi.org/10.1145/3209281.3209319

[16] Wang, Y., et al. (2014). "Acquiring logistics process intelligence: Methodology and an application for a Chinese bulk port." Expert Systems with Applications 41(1): 195-209. DOI= https://doi.org/10.1016/j.eswa.2013.07.021

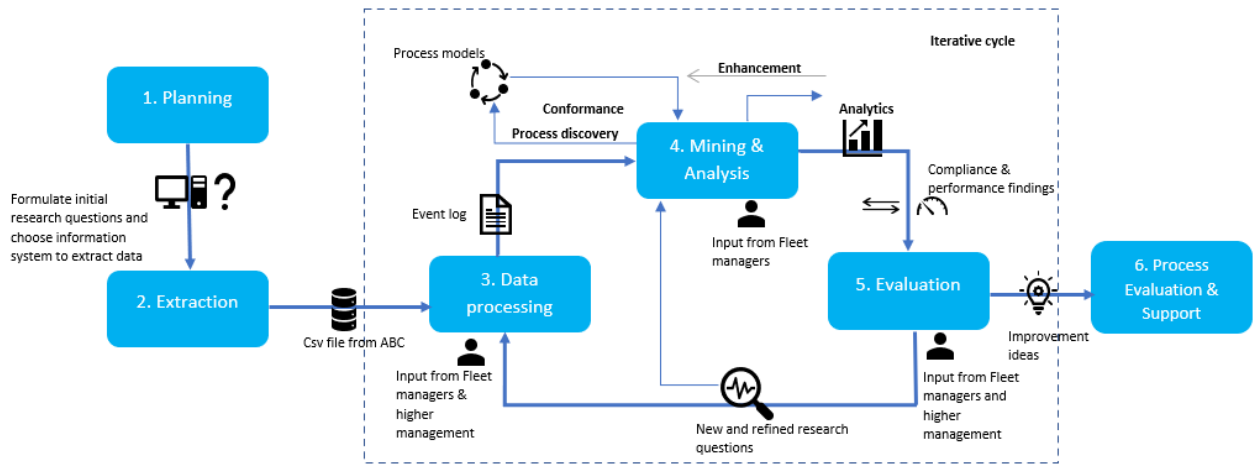# APPENDIX

A. PM² FRAMEWORK

Figure 1. An adapted version of the PM² framework by van Ech et al.
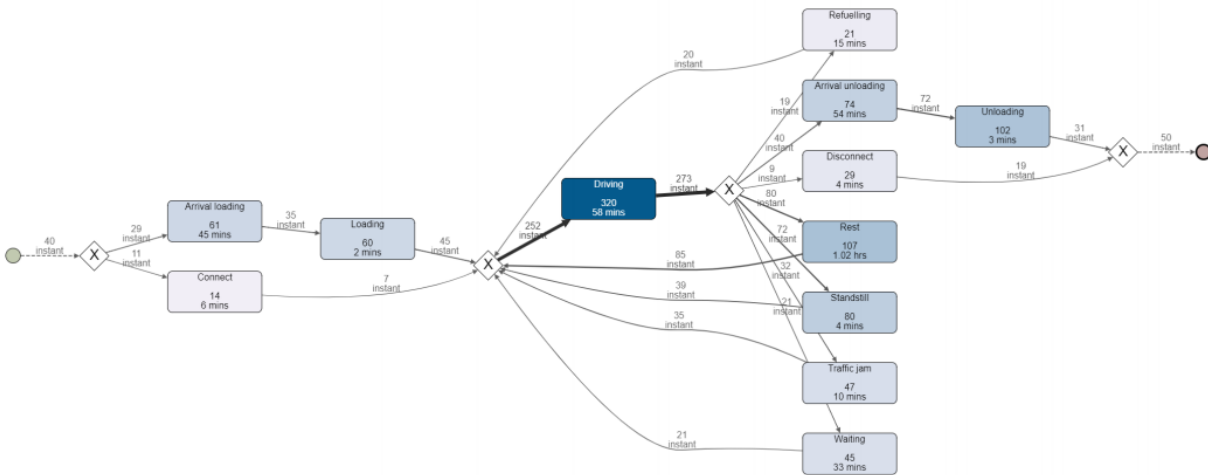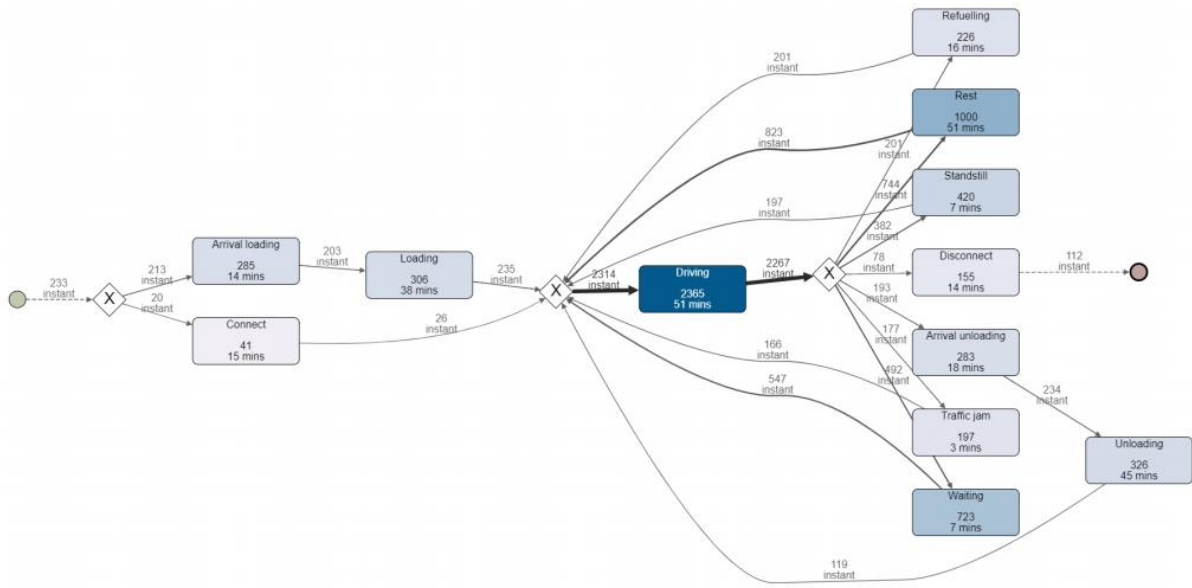
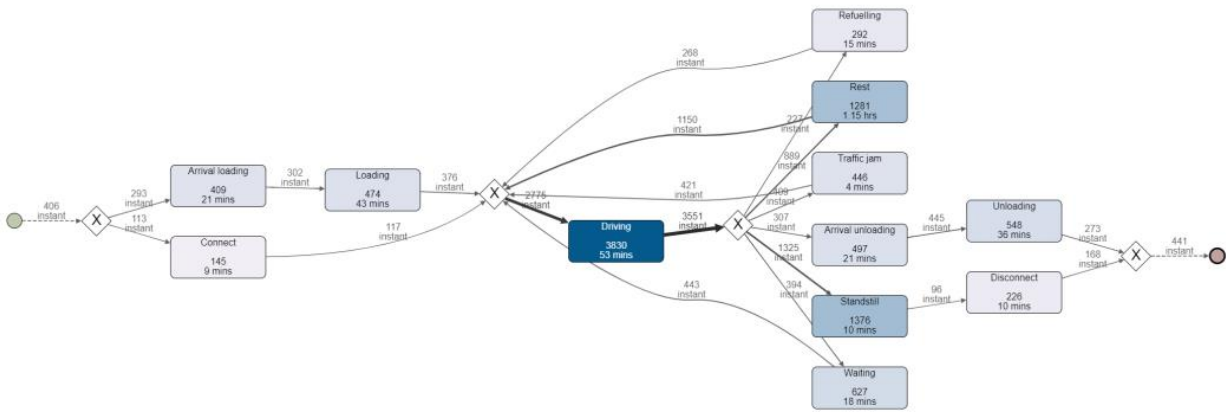## B. PROCESS MODELS

Figure 1. Netherlands

Figure 2. Germany

Figure 4. Czech Republic

C. FEEDBACK FROM INTERVIEWEES

Table 1. Feedback received from interviewees

Role	Feedback
Quality Manager	• A loading activity should follow an arrival loading activity. The differences in the number of arrival loading and loading activities could indicate data loss or lack of board computer reporting/ recording. The same can be said about arrival unloading and unloading • Longer durations for arrival loading/arrival unloading could indicate that in reality, a loading/unloading event must have taken place. • There is no centralized standard of what should and should not be reported on the board computer by the driver. So, countries perform this differently
Fleet manager (Poland)	• Normally, drivers indicate their activities on the board computer so that fleet managers are aware of their activities & whereabouts at all times. When a driver indicates arrival loading and the next activity is driving and the required papers are signed, then it automatically means that loading must have taken place. So, it is then not necessary for a driver to indicate loading.
Thesis supervisor/ Part of the Research Group at ABC	• Shipment data is flushed out and not archived. Some of the activities indicated by drivers could just not be linked to a shipment. This is also indicated in the data processing stage when only 12% of the original dataset is left for analysis

D. OTM EXPERIMENT

Table 1. Test data created from OTM shipment scenario

| Shipment No	Trip ID	Activity	Start Time	End Time	From	To	Vehicle ID
1	1	Loading	6/3/2020 6:15	6/3/2020 6:45	A	A	1
2	1	Loading	6/3/2020 6:15	6/3/2020 6:45	A	A	1
2	1	Driving	6/3/2020 6:45	6/3/2020 7:45	A	B	1
1	1	Driving	6/3/2020 6:45	6/3/2020 7:45	A	B	1
1	2	Unloading	6/3/2020 7:45	6/3/2020 8:15	B	B	1
2	2	Driving	6/3/2020 8:15	6/3/2020 9:15	B	C	1
2	3	Unloading	6/3/2020 9:15	6/3/2020 9:45	C	C	1
3	3	Loading	6/3/2020 9:15	6/3/2020 9:45	C	C	1
3	3	Driving	6/3/2020 9:45	6/3/2020 10:45	C	D	1
3	3	Unloading	6/3/2020 10:45	6/3/2020 11:05	D	C	1
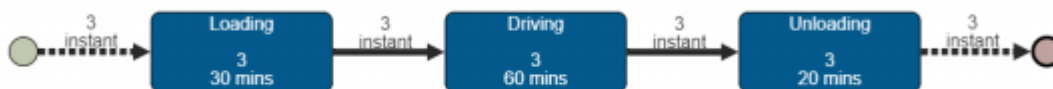
Figure 1. Process model generated using OTM test data (above)