Evaluating the Use of the Open Trip Model for Process Mining: An Informal Conceptual Mapping Study in Logistics

Jean Paul Sebastian Piest[a], Jennifer Alice Cutinha[b], Rob Henk Bemthuis[c]
and Faiza Allah Bukhsh[d]

University of Twente, Drienerlolaan 5, 7522 NB, Enschede, The Netherlands

Keywords: Open Trip Model, Process Mining, Logistics, Event Data.

Abstract: When aggregating logistic event data from different supply chain actors and information systems for process mining, interoperability, data loss, and data quality are common challenges. This position paper proposes and evaluates the use of the Open Trip Model (OTM) for process mining. Inspired by the current industrial use of the OTM for reporting and business intelligence, we believe that the data model of OTM can be utilized for unified storage, integration, interoperability, and querying of logistic event data. Therefore, the OTM data model is mapped to a generic event log structure to satisfy the minimum requirements for process mining. A demonstrative scenario is used to show how event data can be extracted from the OTM's default scenario dataset to create an event log as the starting point for process mining. Thus, this approach provides a foundation for future research about interoperability challenges and unifying process mining models based on industry standards, and a starting point for developing process mining applications in the logistics industry.

1 INTRODUCTION

The logistics sector can be referred to as a network where multiple organizations come together for the planning, organization, coordination, and execution of transportation of goods and logistics services. Typically, a logistic process involves multiple parties (e.g., shippers, logistics service providers, transport operators, or carriers), different entities within an organization, and is spread across different countries. As a result, logistic processes are exceedingly complex and dynamic, and data usually comes from heterogeneous data sources in various (un)structured formats (Intayoad and Becker, 2018).

Shipment data are commonly administered in multiple information systems (e.g., ERP, WMS, TMS, and FMS) (Evofenedex, TLN, and Beurtvaartadres, 2019). As a result of business transactions, the status of shipments and whereabouts of goods are tracked and traced during the physical handling processes. Data are exchanged in different formats (e.g., e-mail and EDI) and supported by industry standards and in-

teroperability models (Evofenedex, TLN, and Beurtvaartadres, 2019).

The Open Trip Model (OTM) is such a data exchange standard and adopted by the Dutch logistics industry as part of a federated data sharing infrastructure (Bastiaansen et al., 2020). Figure 1 depicts the default OTM scenario for data exchange and data sharing between involved stakeholders and their information systems.
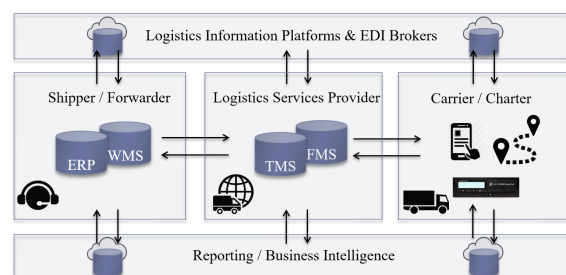

Figure 1: Interaction of multiple logistic parties (adapted from OTM presentation).

In the industry, the OTM data model is also used for reporting and business intelligence, as illustrated in Figure 1, however, to our best knowledge, not for process mining. Process mining focuses on extracting knowledge from data generated and stored in the

[a] https://orcid.org/0000-0002-0995-6813
[b] https://orcid.org/0000-0003-4706-2228
[c] https://orcid.org/0000-0003-2791-6070
[d] https://orcid.org/0000-0001-5978-2754

databases of information systems in the form of event logs (Van der Aalst et al., 2012). Logistic processes generate large amounts of event data. Event data are expected to be a rich source for behavior analysis as it comprises data concerning the dynamic behavior of people, objects, and systems at a detailed level. Process mining techniques can be helpful to produce insightful information based on logistic event data.

When event data are aggregated from different systems for process mining, interoperability, data loss, and data quality issues are common challenges. Existing approaches aim to increase the accuracy of process mining techniques despite noisy data. Consequently, various tools and algorithms have been built in process mining tools to eliminate the effect of noisy data and determine the actual control-flow of a process. However, as logistic event data are typically complex, dynamic, and heterogenous, it remains challenging to generalize their results (Intayoad and Becker, 2018). More precisely, much of the current literature pays little attention to unified standards of logistic process definitions. The OTM provides such a standard for sharing logistic data. We believe that the usage of the OTM, in addition to addressing (some) interoperability issues, can provide a promising foundation for a more unified implementation of process mining practices in the logistics domain and its heterogeneous environment.

This paper aims to evaluate the use of the OTM for process mining. An informal conceptual mapping study is conducted to determine whether the basic requirements for process mining can be fulfilled. More specifically, the data model of the OTM is mapped to a generic event log structure. Based on a step-by-step walkthrough, we demonstrate how the OTM data model can be used to extract event data and create an event log. The event data are imported in the process mining tool Disco to generate a process model. This way, we provide researchers a common ground for solving interoperability issues related to process mining in heterogeneous environments using industry standards, including OTM, and practitioners a dedicated, potentially generalizable, process model and approach to develop process mining applications.

The remainder of this paper proceeds as follows. Section 2 discusses related work. Section 3 positions the use OTM for process mining. Section 4 discusses preliminary results. Section 5 concludes and provides an outlook for future research.

2 RELATED WORK

The practice of process mining has gained attention in many domains, such as healthcare (Mans et al., 2012; Erdogan and Tarhan, 2018), education (Bogarín et al., 2018), finance (De Weerdt et al., 2013), logistic, and supply chain processes.

2.1 Process Mining in Logistics

There are several studies available on process mining in the logistics domain.

A systematic mapping study (dos Santos Garcia et al., 2019) illustrates that less than 5% of their paper sample is about the logistics domain. This mapping study identified 27 studies with a focus on logistic processes, including transportation, storage of goods, and stock management. Most of the studies focus on process discovery in the logistical context. Specific studies study process mining in regard to network analysis, resource configuration, prediction of event times, and remodeling of business processes. These studies indicate a rich spectrum of use cases.

Additional studies examined logistic processes through process mining, mainly focusing on the internal logistics of case-specific scenarios (Knoll et al., 2019a; Knoll et al., 2019b). Others developed a process mining system for determining the root causes of quality problems in a supply chain (Lau et al., 2009). Based on daily captured logistic data, the authors fine-tuned configuration parameters to improve operational performance.

2.2 Interoperability Challenges

There are relevant studies in the logistics domain that address the interoperability challenges using process mining techniques and the need for standardization.

On a high abstraction level, the interoperability issues are addressed in the four levels of big data interoperability (Singh and van Sinderen, 2016). More specifically, the study of (Lont et al., 2018) shows how different systems and devices can be linked to the data model of OTM, eliminating certain interoperability issues.

Some studies emphasize the complexity of monitoring logistic processes (Cabanillas et al., 2013; Wang et al., 2014). The authors pinpoint the importance of new research, novel contributions on discretizing, aggregating, and correlating events in a way that the overall business process can be better traced. This work indicates that further research should be done on improving the quality of the event log data by including a reconciliation of the data.

3 INFORMAL CONCEPTUAL MAPPING OF OTM FOR PROCESS MINING

3.1 Aim and Data Model of OTM

The OTM is an open-source, flexible data sharing model that contributes to uniform and consistent exchange of information across various information systems. This model is managed by the Stichting Uniforme Transport Code (SUTC) and its goal is to help logistics companies in the Netherlands share real-time logistic data efficiently (Stichting Uniforme Transport Code (SUTC), 2019).

The constructed data model of the OTM, as shown in Figure 2, is centered around event data and considers eight entities and four lifecycles.
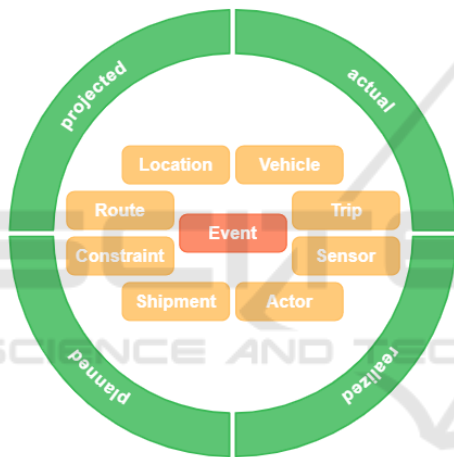

Figure 2: OTM data model (Open Trip Model, 2020).

Entities are used to represent various objects within a logistic process, e.g., vehicles. All dynamic behavior of actors is modeled as (a series of) event(s) and are related to shipments, trips, and routes. The order of these events indicates the workflow over time, and this is depicted by the lifecycle. This way, a trail of event data is created. The lifecycle expresses the different phases in the transport process and enables different views on the operation (e.g., look ahead at events that have been planned, what is taking place right now or look back at what has been realized). Event data, together with related entities and the lifecycles, provide the foundation to develop process mining applications, behavioral analysis, and performance management.

3.2 Using OTM for Process Mining

In addition to data exchange, the data model of the OTM can be used for storage, integration, and querying of logistic event data originating from multiple information systems as the foundation for process mining use cases relevant to various stakeholders in the logistics industry. The event log should contain four data elements to fulfill the minimal requirements for process mining, 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 (Van der Aalst et al., 2012).

3.3 Evaluating and Demonstrating the Use of OTM for Process Mining

The OTM walkthrough scenario and example data described on the website (Open Trip Model, 2020) are used to evaluate the use of the OTM for process mining. Figure 3 shows how version 4.2 of the data model of the OTM can be mapped to satisfy the minimum process mining requirements.
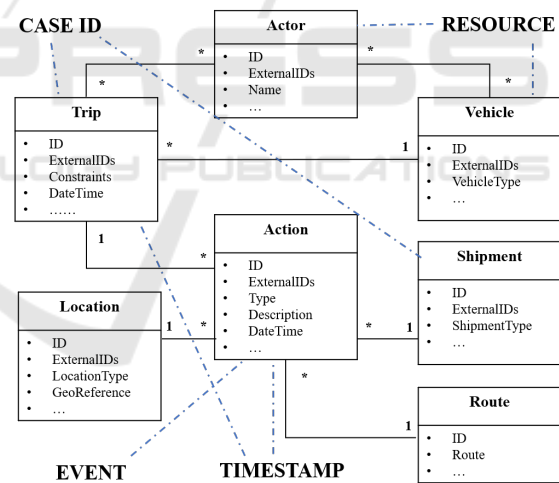

Figure 3: Adapted OTM data model version 4.2 linked to the minimum process mining requirements.

In the following part, we will walk through an operational logistics scenario and discuss how this scenario can be expressed in OTM entities and events, and eventually used for process mining.

3.3.1 Describe the Scenario

The example scenario is related to the logistics operation of a supermarket chain that contracts transport companies to transport goods from their warehouse

to the stores. Only a limited part of the operations is discussed here, as that will be enough to highlight the most important concepts of OTM.

The OTM scenario (Open Trip Model, 2020) starts with the planning department of the supermarket chain, sending the following instructions to a contracted transport company:

"I have a transportation request, requiring a diesel-powered boxtruck to transport refrigerated goods from "warehouse A", dock 14 (start loading at 6:15 AM, 30 minutes loading time) to two stores: "store B", where there is 30 minutes unloading time and "store C". From store C some goods must be returned to "terminal D" at the warehouse location. At store C there's also 30 minutes for loading and unloading. According to our route planner, you should be able to go from A to B, B to C and C to D in a given time and by driving no more than the given distance. The set access routes to the stores are taken into account."

3.3.2 Identify and Map Entities

We can identify the following OTM entities in this text: vehicle, shipment, location, route, and trip. The scenario contains a refrigerated boxtruck, which can be mapped to the vehicle entity as follows:

```
Vehicle:
    id: 1
    type: refrigerated boxtruck
    fueltype: diesel
```

The scenario informally describes three types of goods that need to be transported and can be mapped to the shipment entity as follows:

```
Shipment:
    id: 1
    contents: refrigerated goods
    from: location A
    to: location B
Shipment:
    id: 2
    contents: refrigerated goods
    from: location A
    to: location C
Shipment:
    id: 3
    contents: returned goods
    from: location C
    to: location D
```

The shipments are transported to four locations and can be mapped to the location entity as follows:

```
Location:
    id: A
    name: warehouse A, dock 14
Location:
    id: B
```

```
    name: store B, loading bay
Location:
    id: C
    name: store C, loading bay
Location:
    id: D
    name: returned goods, terminal D
```

The route and trip are informally formulated and considered identical in this scenario. Event data typically originates from multiple actors and systems.

3.3.3 Identification of Events

Based on the identification and mapping of entities from the text, the following events can be identified:

```
Planned event on Trip 1:
    start loading  6:15 AM
Planned event on Vehicle 1:
    load Shipment 2
Planned event on Vehicle 1:
    load Shipment 1
Planned event on Trip 1:
    stop loading  6:45 AM
Planned event on Trip 1:
    start driving from A 6:45 AM
Planned event on Trip 1:
    stop driving at B 7:45 AM
Planned event on Trip 2:
    start loading/unloading 7:45 AM
Planned event on Vehicle 1:
    unload Shipment 1
Planned event on Trip 2:
    stop loading/unloading 8:15 AM
Planned event on Trip 2:
    start driving from B 8:15 AM
Planned event on Trip 2:
    stop driving at C 9:15 AM
Planned event on Trip 3:
    start loading/unloading 9:15 AM
Planned event on Vehicle 1:
    unload Shipment 2
Planned event on Vehicle 1:
    load Shipment 3
Planned event on Trip 3:
    stop loading/unloading 9:45 AM
Planned event on Trip 3:
    start driving from C 9:45 AM
Planned event on Trip 3:
    stop driving at D 10:45 AM
Planned event on Trip 3:
    start loading/unloading 10:45 AM
Planned event on Vehicle 1:
    unload Shipment 3
Planned event on Trip 3:
    stop loading/unloading 11:05 AM
```

3.3.4 Extract the Event Data

Based on the identified entities and events, event data are created based on the scenario. Table 1 presents

Table 1: Event Log Created based on the OTM Walkthrough and Example Data.

Shipment	Trip	Activity	Start time	End time	From	To	Vehicle
1	1	Loading	6-1-2021 06:15	6-1-2021 06:45	A	A	1
2	1	Loading	6-1-2021 06:15	6-1-2021 06:45	A	A	1
1	1	Driving	6-1-2021 06:45	6-1-2021 07:45	A	B	1
2	1	Driving	6-1-2021 06:45	6-1-2021 07:45	A	B	1
1	1	Unloading	6-1-2021 07:45	6-1-2021 08:15	B	B	1
2	2	Driving	6-1-2021 08:15	6-1-2021 09:15	B	C	1
2	2	Unloading	6-1-2021 09:15	6-1-2021 09:45	C	C	1
3	3	Loading	6-1-2021 09:15	6-1-2021 09:45	C	C	1
3	3	Driving	6-1-2021 09:45	6-1-2021 10:45	C	D	1
3	3	Unloading	6-1-2021 10:45	6-1-2021 11:05	D	C	1

the test dataset. The dataset contains three shipments, three trips to visit three locations, three types of activities, planned start and end times, three locations, and an assigned vehicle.

3.3.5 Create the Event Log

We used the process mining tool Disco (developed by Fluxicon) to import the event data using the following script:

```
Import column mapping:
    'Shipment' → Case ID
    'Activity' → Activity
    'Start time' → Timestamp
    (Pattern: 'yyyy/MM/dd HH:mm:ss')
    'End time' → Timestamp
    (Pattern: 'yyyy/MM/dd HH:mm:ss')
    'Vehicle' → Resource
```

Based on the import script, the event log can be created in Disco. Figure 4 shows how the minimal requirements for process mining are fulfilled.
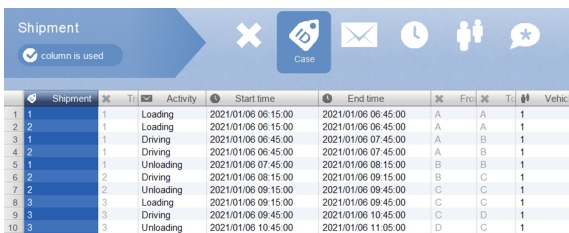

Figure 4: Screenshot of the event log created in Disco.

3.3.6 Generate the Process Model

Based on the event log, the process model can be generated. Figure 5 depicts the generated process model, including some basic model statistics (e.g., frequencies, repetitions) and performance indicators (e.g., duration).

Using process mining techniques, the event data are split up into three cases and two variants, as shown in Figure 6, which can be analyzed for patterns. When
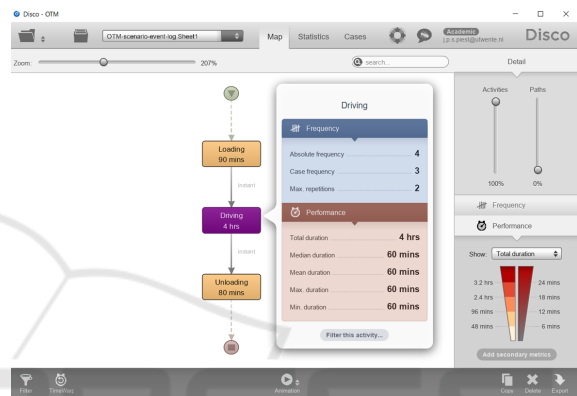

Figure 5: Screenshot of the process model, statistics, and performance in Disco.

the actual process is executed, the planned lead times could be used as a norm for performance monitoring based on the OTM data model. The lifecycles might be used to detect deviations and compliance checking (e.g., using rules and regulations regarding driving and rest times). In addition, the lifecycles could potentially be used for simulations, optimizations, predictions, and model enhancements.
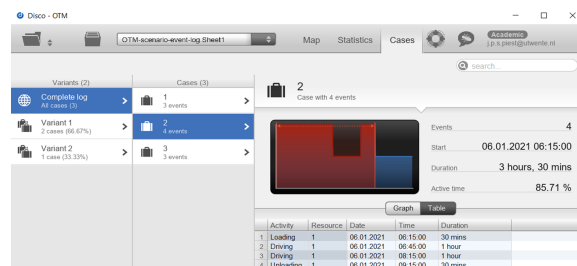

Figure 6: Screenshot of the detailed view with cases and variants in Disco.

4 DISCUSSION

Current literature indicates a rich spectrum of process mining use cases in logistics. The absence of unified process definitions or standardized process models makes it hard to generalize their results. Furthermore, the aggregation of event data in logistics has proven difficult due to the complex and heterogeneous nature of logistics. Interoperability issues are addressed in several studies. However, established approaches and tools focus on working with noisy data.

The ideas put forward in this position paper are based on a different approach, and propose the use of OTM for process mining, addressing the interoperability challenges and develop applications based on a dedicated but generalizable process model. Although this position paper contains promising preliminary results and demonstrates how the basic requirements for process mining can be fulfilled, the support is limited. The conducted informal conceptual mapping study in logistics requires further experimental research and comparative studies.

This position paper describes and traces the use of OTM for process mining through an example, mapping OTM to the requirements for process mining and demonstrating its use. Nevertheless, it is likely that even if OTM is widely adopted that handling data integration and extraction challenges for other systems would still need to be addressed.

Besides that, the work is still in an initial stage and the approach should be tested in industry to determine how it helps decision-makers in the process of conducting process mining analysis. Additional examples and more complex use of the proposed approach should be explored, considering the relevant concerns and issues of the decision-makers in the logistic processes. More precisely, to verify whether OTM has the potential to be considered.

5 CONCLUSIONS

Inspired by the current industrial use of the OTM for reporting and business intelligence, its use for process mining is evaluated and demonstrated in this paper.

5.1 Preliminary Results and Findings

Based on an informal conceptual mapping study in logistics, it is shown how the minimum process mining requirements can be satisfied based on the OTM data model. Based on the default OTM scenario and example data, the step-by-step walkthrough discusses and demonstrates how event data can be extracted to create an event log. The process mining tool Disco is employed to generate a process model.

5.2 Implications and Limitations

This demonstration provides initial support that the OTM can fulfill the minimum requirements for process mining. However, this paper's demonstration only covers one implementation of the OTM and is based on synthetic data. Therefore, further experimental research and development is required, involving decision-makers from the logistics industry.

The approach proposed in this paper requires the adoption of the OTM by all involved actors and their supporting information systems. The OTM is currently mainly used in the Netherlands. However, given its prominent position within the federated data sharing infrastructure for the Dutch logistics industry, it is considered to be a promising direction for future research and development. To increase the applicability and potential generalization of the proposed approach, similar global industry standards should be evaluated and compared.

5.3 Future Research Directions

Future research could focus on systematically mapping the process mining spectrum to the OTM based on formal methods and techniques.

To extend the preliminary research, a full implementation of OTM is required and this approach should be tested for robustness with real-life datasets in multiple use cases. The identified use cases provide a starting point to conduct case study research. Future work should aim to determine if implementations of OTM and real-world data are also as straightforward to map for process mining.

Further research directions could also include applying process mining techniques to these explored use cases in organizations that implement the OTM. A comparison-based study, in this regard involving organizations that implement and do not implement the OTM, would be an interesting next step.

Furthermore, a comparison study on solution alternatives (e.g., the GS1 EPICS) and alternative approaches (e.g., data mining, machine learning) should be conducted to evaluate the use of industry standards in a broader sense.

ACKNOWLEDGEMENTS

COS Project. The authors thank the involved consortium partners for their support and contribution to this research. The authors also thank the anonymous reviewers for their constructive feedback.

REFERENCES

Bastiaansen, H., Nieuwenhuis, C., Zomer, G., Piest, J., van Sinderen, M., Dalmolen, S., and Hofman, W. (2020). The logistics data sharing infrastructure: whitepaper august 2020. Technical report. Available online: https://www.dinalog.nl/wp-content/uploads/2020/08/ Dinalog_Whitepaper-Data-Infrastructure_DEF.pdf (accessed on 8 March 2021).

Bogarín, A., Cerezo, R., and Romero, C. (2018). A survey on educational process mining. *WIREs Data Mining and Knowledge Discovery*, 8(1):e1230.

Cabanillas, C., Baumgrass, A., Mendling, J., Rogetzer, P., and Bellovoda, B. (2013). Towards the enhancement of business process monitoring for complex logistics chains. In *International Conference on Business Process Management*, pages 305–317. Springer.

De Weerdt, J., Schupp, A., Vanderloock, A., and Baesens, B. (2013). Process mining for the multi-faceted analysis of business processes—a case study in a financial services organization. *Computers in Industry*, 64(1):57–67.

dos Santos Garcia, C., Meincheim, A., Junior, E. R. F., Dallagassa, M. R., Sato, D. M. V., Carvalho, D. R., Santos, E. A. P., and Scalabrin, E. E. (2019). Process mining techniques and applications–a systematic mapping study. *Expert Systems with Applications*, 133:260–295.

Erdogan, T. G. and Tarhan, A. (2018). Systematic mapping of process mining studies in healthcare. *IEEE Access*, 6:24543–24567.

Evofenedex, TLN, and Beurtvaartadres (2019). Nationaal onderzoek data en digitalisering in de logistiek. Available online: https://www.tln.nl/app/uploads/2019/11/ Rapport-data-en-digitalisering.pdf (accessed on 8 March 2021).

Intayoad, W. and Becker, T. (2018). Applying process mining in manufacturing and logistic for large transaction data. In *International Conference on Dynamics in Logistics*, pages 378–388. Springer International Publishing.

Knoll, D., Reinhart, G., and Prüglmeier, M. (2019a). Enabling value stream mapping for internal logistics using multidimensional process mining. *Expert Systems with Applications*, 124:130–142.

Knoll, D., Waldmann, J., and Reinhart, G. (2019b). Developing an internal logistics ontology for process mining. *Procedia CIRP*, 79:427–432.

Lau, H., Ho, G., Zhao, Y., and Chung, N. (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. Transport Logistics and Physical Distribution Interlocking of Information Systems for International Supply and Demand Chains Management ICPR19.

Lont, Y., van Duin, R., Jens, D.-P., and van Lier, B. (2018). Demasking the black hole of transportation? A blocking road naar de ontwikkeling van een CO2 Blockchain-georienteerde (product) app. *Bijdragen vervoerslogistieke werkdagen 2018*, pages 1–12.

Mans, R. S., van der Aalst, W. M., Vanwersch, R. J., and Moleman, A. J. (2012). Process mining in healthcare: Data challenges when answering frequently posed questions. In *Process Support and Knowledge Representation in Health Care*, Lecture Notes in Computer Science, pages 140–153. Springer, Berlin, Heidelberg.

Open Trip Model (2020). The OTM5 model. Available online: https://www.opentripmodel.org/docs (accessed on 8 March 2021).

Singh, P. and van Sinderen, M. (2016). Big data interoperability challenges for logistics. In *Enterprise interoperability in the digitized and networked factory of the future*, pages 325–335. ISTE Press.

Stichting Uniforme Transport Code (SUTC) (2019). Algemene Presentatie SUTC. Available online: https://dutchmobilityinnovations.com/spaces/ 1168/connected-transport-corridors/files/27520/ algemene-presentatie-sutc-pptx?version=1 (accessed on 1 September 2020).

Van der Aalst, W. et al. (2012). Process mining manifesto. In *Business Process Management Workshops*, Lecture Notes in Business Information Processing, pages 169–194. Springer, Berlin, Heidelberg.

Wang, Y., Caron, F., Vanthienen, J., Huang, L., and Guo, Y. (2014). Acquiring logistics process intelligence: Methodology and an application for a chinese bulk port. *Expert Systems with Applications*, 41(1):195–209.