Available online at www.sciencedirect.com

ScienceDirect

Procedia CIRP 79 (2019) 427–432

www.elsevier.com/locate/procedia

12th CIRP Conference on Intelligent Computation in Manufacturing Engineering, 18-20 July 2018, Gulf of Naples, Italy

Developing an internal logistics ontology for process mining

Dino Knoll[a,]*, Julian Waldmann[a], Gunther Reinhart[a]

[a]*Institute for Machine Tools and Industrial Management, Technical University of Munich, Munich, Germany*

* Corresponding author. Tel.: +49-89-289-15550; fax: +49-89-289-15500. *E-mail address:* dino.knoll@iwb.mw.tum.de

Abstract

Process mining offers the potential for internal logistics process improvement using data. While there exists a massive amount of data, identifying and understanding relevant data across different information systems and complex data models is difficult. To address this issue, ontologies can be used to formalize a shared understanding. This paper aims to provide an extended internal logistics ontology focusing the process perspective. Existing internal ontologies are reviewed, compared and merged. Additionally, related resources such as products and packaging is integrated. In conclusion, this paper proposes a domain ontology to support process mining within internal logistics.

Peer-review under responsibility of the scientific committee of the 12th CIRP Conference on Intelligent Computation in Manufacturing Engineering.

Keywords: Process mining; Preprocessing; Internal logistics; Ontology

1. Introduction

Within last decades, manufacturing industry has been driven by globalization and mass customization. On the one hand, this results in a high model-mix production with a large variety of parts. On the other hand, companies are faced by massive pricing pressure which requires a continuous improvement of processes. To enable a high model-mix assembly line production, internal logistics has to provide the right quantities of goods most efficiently at the right place in the right order within the right time [1]. Internal logistics refers to the receipt of parts, warehousing (e.g. storing, sequencing), line feeding up to line side presentation [2, 3, 4]. Consequently, the complexity within internal logistics has increased dramatically and today's logistics operations require various information systems (e.g. ERP, WMS and individual software).

Recent developments in information technology aim to re-use existing event data to improve processes using process mining. 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 (information) systems [5]. Process mining has already been successfully applied to manufacturing [6, 7, 8] and logistics [9, 10, 11].

However, in practice the data is often not correlated as events and significant effort is required to create event logs [5]. First, there exists a large variety of complex, connected information systems within production and logistics which hold relevant data. The data extraction cannot be driven by experts of the domain who do not have any technical knowledge about the underlying information systems and concrete storage mechanisms [12]. Second, to correlate the data, numerous data tables have to be merged, and the data attributes that are interesting for process mining have to be explicitly located in each data table [13]. Calvanese et al. (2016) reported this as the "process-orientation" problem, that the underlying data must be understood from conceptual lenses before process mining can be applied [12]. Third, the extraction process spans across several levels of abstraction and there is no such notion for a single event log [12]. In particular, for applying process mining within SAP a main challenge is that there is no defined consistency for how all documents, change events and resource dependencies are stored [13]. Fourth, vocabulary used in the IT implementation differs radically from domain knowledge [12]. Fifth, the volume and veracity of data in internal logistics is

2212-8271 © 2019 The Authors. Published by Elsevier B.V.
Peer-review under responsibility of the scientific committee of the 12th CIRP Conference on Intelligent Computation in Manufacturing Engineering.
10.1016/j.procir.2019.02.116

significantly high. For example, in one production plant of a high model-mix assembly line in automotive industry, up to 5,000,000 transfer orders are processed every month [14]. Consequently, in order to apply process mining, the right data has to be identified, extracted, merged and interpreted. Within the area of process mining, domain ontologies are frequently used to support these steps. However, there is no domain ontology for internal logistics to support process mining yet.

The remainder of this paper is organized as follows. Section 2 presents ontology-based data preprocessing for process mining. Section 3 reviews the state of the art of logistics domain ontologies. Section 4 outlines the methodology to develop the domain ontology which has been applied in Section 5. Section 6 concludes this research and potential next steps.

2. Ontology-based preprocessing for process mining

In the following section, a short review of current practice to utilize ontologies for the preprocessing phase of process mining is outlined.

Process mining requires a set of data correlated to an event log. An event log refers to an activity, a particular case and a timestamp [5] and can be extended by additional information (e.g. size of an order) [12]. Consequently, the goal of the pre-processing phase is to extract normalized event logs, a job which involves activities such as data cleansing, feature selection, and merging of distinct data sources [13]. This requires the reduction of complex data models to an event log containing all relevant information. To do so, the use of ontologies to facilitate data preparation is found in the process mining and data mining area [15]. An ontology is an explicit specification of a conceptualization [16]. This set of objects, and the describable relationships among them, are reflected in the representational vocabulary with which a knowledge-based program represents knowledge [16].

For process mining, ontologies provide three benefits. First, a set of relevant objects and their properties enable an efficient identification of necessary information systems, underlying data tables and corresponding columns. Further on, the ontology helps to identify additional relevant information (e.g. master data) which is not directly recorded in the event data. Second, as ontologies also include existing relations between these relevant objects, the reduction of complex models to the event log can be supported. For example, Calvanese et al. (2016) provide a framework to link a domain ontology with raw data to create event logs [12]. Third, most researches have claimed that the process data elements in logs are label-based and lower in abstraction. To enhance the interpretation of results, the data elements in the event log are enriched with their concepts from domain ontologies [15]. For example, Jareevongpiboon & Janecek (2013) linked to product codes of leatherwear with its human-readable names [15].

Summarizing, the extendibility and reusability of ontologies helps to reduce time and interpret results [15]. In particular, for systems which do not produce event logs explicitly, the preprocessing is the most time-consuming and work intensive phase [13]. However, these fundamental approaches and methodologies to utilize ontologies for process mining require a domain ontology with the concepts, object properties and data properties [12, 13, 15]. Especially the quality of ontological data in the process mining projects is critical for interpreting events at the system level and constructing meaningful process models [13].

3. State of the art

To apply ontology-based preprocessing for process mining, a domain ontology of internal logistics is required. Thus, in the following section a review of existing ontologies within internal logistics and corresponding areas will be given.

3.1. Internal logistics ontologies

In the area of internal logistics, Negri et al. (2017) recently carried out a detailed review of existing ontologies [4]. In total, 26 logistics ontologies have been identified. We extended this research by the keywords of internal logistics ontology and production logistics ontology and identified 16, respectively 57 ontologies. In total, six duplicates have been removed and 67 ontologies remain. Additionally, we added two ontologies available in German. In contrast to Negri et al. (2017) who developed an ontology with the focus to include a classification of internal logistics resources [4], we aim on the process perspective for process mining. Subsequently, first we present the overall findings and afterwards we outline the most important ontologies in detail.

According to Negri et al. (2017), much of the work has been published in the area of manufacturing and production while not enough work has been carried out to internal logistics [4]. Out of the 26 logistics ontologies, only two keep the main focus on internal logistics and warehousing. In contrast, the other 24 ontologies focus on supply chain [4]. As the majority focusses on the inter-organizational perspective of logistics, they only often provide abstract concepts and a very small number of concepts and relationships [4]. Noteworthy, Negri et al. (2017) pointed out that the majority of supply chain ontologies have been developed for information and data integration between supply chain partners [4]. This emphasizes the usability of ontologies for the purpose outlined in Section 2.

Our research confirmed this perspective and we can add four more findings of existing ontologies for the use of process mining.

First, most existing ontologies do not integrate both the process and resource perspective. In particular, on the resource perspective, existing ontologies focus on selected aspects. As an example, Negri et al. (2017) set the main focus on transport and storage resources while Kowalski & Quink (2013) focus on packaging [17]. Thus, existing relations are often neglected. Also, other data properties (e.g. value of the part) are not integrated while there exists ontologies addressing these issues.

Second, as the ontologies have a different purpose, they are modelled with a different level of detail and abstraction. On the one hand there exist ontologies with a high level of detail and an explicit formalization [4, 17]. For example, Kowalski

& Quink (2013) describe packaging, packaging goods (parts), unit loads, underlying standards including instances (e.g. DIN ISO, DIN EN). On the other hand there are ontologies which only include a minimal set of objects, object properties and data properties. While Libert et al. (2010) integrate material flow, information flow and resources [18], the level of detail is low. For example, the modelled material flow objects only contain buffer, store and transfer. Other relevant material flow activities such as picking, sorting and distributing are not included.

Third, the there is no standardized taxonomy for equivalent objects, object properties and data properties. For example, Merdan et al. (2008) use the term "pallet" [19] while other ontologies refer to the term "unit load" (or "load unit"). Even further, two useful ontologies are only available in German.

Fourth, existing concepts of upper ontologies are often neglected. Expect from Negri et al. (2017), which is based on the *Manufacturing Systems Ontology*, upper ontologies (such as *Process Specification Language* [20] or *Know-Ont* [21] are not included. In consequence, every ontology describes the process and its activities in a different way. In particular, this is challenging if the names and concepts are not compatible with the process mining perspective.

3.2. Shortcomings

Ontologies provide support for the ontology-based preprocessing for process mining. While there exist various methods to support the preprocessing, the quality and purpose of the domain ontology is critical [13]. Existing ontologies in the domain of internal logistics are still not complete due to: (1) the partial focus of the ontologies without an integrated view of process and resource, (2) the different level of detail and abstraction, (3) the missing standardization of taxonomy and (4) the missing integration of existing upper ontologies.

Against the background of process mining and its aim to optimize processes, not all relevant information are integrated. Especially on the evaluation of the relevance of individual parts (e.g. value) and its relation within the supply chain (e.g. supplier name, demand) there is still need for improvement.

3.3. Objectives

Therefore, an applicable ontology of the domain internal logistics to support process mining has to be developed. For this reason, existing internal logistics ontologies have to be merged into an integrated ontology. The objective of the ontology is to create a representation of internal logistics with its main activities (material and information flow), the required resources and its relevant properties (both object and data properties) to enable process mining. Having a shared set of objects, object properties and data properties, process mining methods of preprocessing can be efficiently applied. The key contribution of this paper is (1) to compare objects, object properties and data properties of existing ontologies, (2) to merge the ontologies and to (3) extend missing information regarding the purpose of process mining.

4. Methodology

Creating and merging ontologies has been widely addressed in research. As the interest of ontologies has been increased within last years, there is a large variety of existing methodologies with different characteristics available.

As described in Section 2, there exist numerous ontologies in the domain of internal logistics. Subsequently, the integration of existing ontologies is a major requirement. We applied the well-established methodology of Noy & McGuinness (2001) due to three advantages: (1) reuse of existing ontologies, (2) a detailed documentation and (3) its simplicity. The methodology consists of seven steps [23]. First, the domain and scope of the ontology is determined (c.f. Section 2). Second, the existing ontologies are reviewed in order to re-use them (c.f. Section 3). Third, important terms of the ontology are enumerated. In the fourth step, the objects (classes) and its hierarchy are defined. Next, in the fifth step the object properties (relations) and in the sixth steps the data properties (slots) are extended. These steps are described using examples and results within Section 5. The last step to create instances is dropped in the context of process mining, as the instances are stored in the databases of information systems [15].

To overcome the challenges while merging ontologies, there exist different approaches. However, these approaches are far from being standards [24]. We decided to use the PROMPT methodology [25]. In contrast to FCA-Merge which is application-based using instances [26], PROMPT has been designed to integrate formalized ontologies. Beneficial is that objects, object properties and data properties are maintained and the effort is comparatively low.

5. Internal logistics ontology for process mining

Using the proposed methodology (c.f. Section 4), the internal logistics ontology for process mining is developed. This section is organized as follows: First, we summarize the steps of creating and merging the ontology using the methodology. Second, the ontology is outlined on a top level and selected objects are explained in detail.

The first step of the methodology is the determination of the domain and the scope of the ontology. Based on the objectives of this paper, the following four competency questions have been formulated:

- Which logistics activities are related to the material flow and information flow?
- Which resources are required to fulfill these activities?
- How can the transition of objects be described regarding the time and location?
- Which information are required to characterize parts and processes on the level of process mining?

For this purpose, 42 relevant ontologies out of the 67 within the domain of internal logistics have been evaluated. To reuse the existing ontologies, we (1) derived the main focus of the ontologies (e.g. process, resource) and afterwards we evaluated the quality and quantity of (2) objects, (3) object

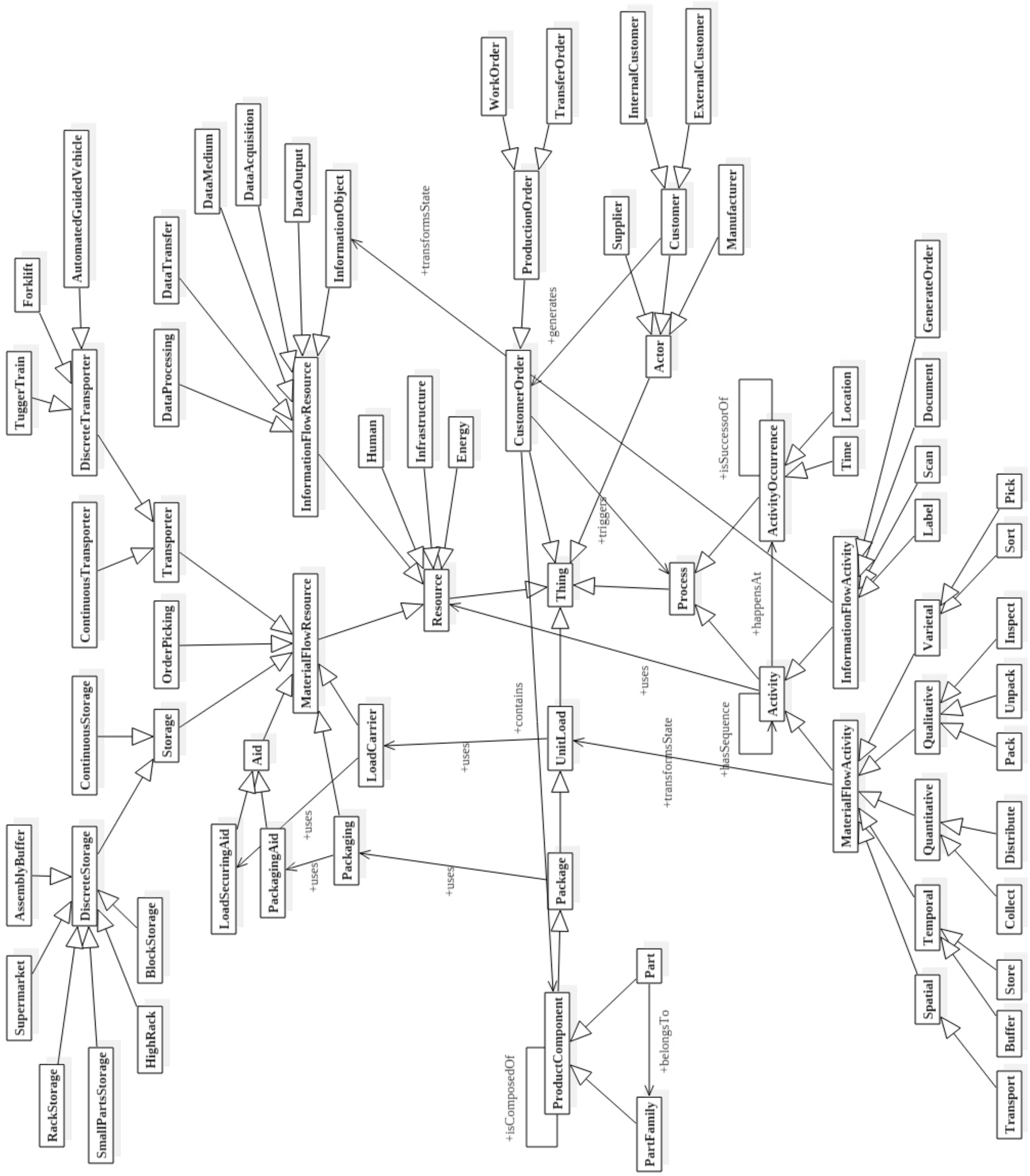

Fig. 1. Internal logistics ontology.

properties and (4) data properties. In sum, the main focus of existing ontologies is (1) resource (e.g. storage, transfer), (2) process (e.g. activities) and (3) actors (e.g. customer). Also load unit, product, part, logistics operations and activities, location and time are named frequently. However, as the terms used in most ontologies are not very well aligned with the existing upper ontologies (except for [4]) we adapted (1) *Process Specification Language* [20], (2) *Know-Ont* [21] and (3) *Unified Foundational Ontology* [22].

Based on these findings, the important terms and objects are defined. The developed internal logistics ontology is mainly based on 14 ontologies. Libert et al. (2010) provide a set of relevant objects with the process, the logistics activities (e.g. transfer, store) and the execution of orders using transfer orders [18]. Negri et al. (2017) show a valuable formalization of resources (transporters, storages, unit loads) within internal logistics. In particular, the object properties and data properties are described in detail [4]. Using these results, we iteratively merged the (1) packaging objects [17], (2) supply chain actors *Onto-SCM* [27] and (3) parts and products using *ONTO-PDM* [28] and [29]. During the steps of merging, both the different level of detail and abstraction and the missing standardization of taxonomy have been a challenge. The result of the entire ontology on a top level is shown in Figure 1. On the top level (subclass of *Thing*), we formulated the objects *Process*, *Resource*, *UnitLoad, Actor and CustomerOrder*.

The *Process* is triggered by a *CustomerOrder*, which is decomposed to a *ProductionOrder* with a daily demand of parts and the *WorkOrder*. Latter is the execution on shop-floor level, e.g. to trigger a single *Part* transfer using a transfer order from one *Location* to another according a *Path*. Both the *Supplier*, the *Manufacturer* and the *Customer* are defined as *Actor*. The object *Process* represents a set of correlated activities (c.f. Figure 2). As subclasses we defined the *Activity* and *ActivityOccurrence* according the Process Specification Language [20]. The *Activity* has a sequence and happens at an occurrence. The subclass of the *Activity* is divided into the *MaterialFlowActivity* and the *InformationFlowActivity*. To model the subclasses we used the fundamental definitions of logistics [30, 31]. A *MaterialFlowActivity* transforms the physical state of a *UnitLoad*.

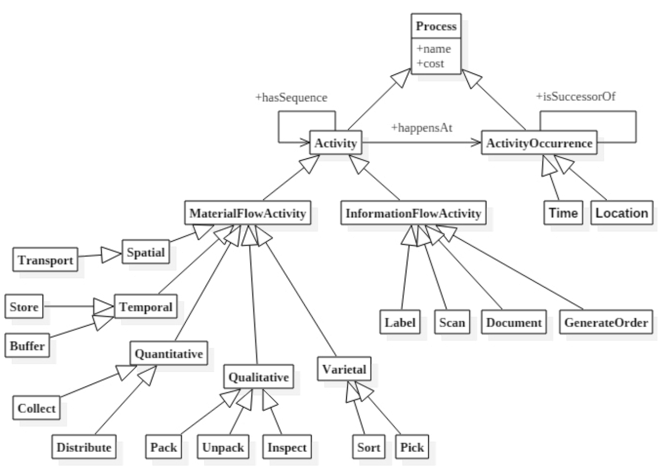

Fig. 2. Extract of the Process object.

Any *InformationFlowActivity* transforms the *InformationObject* which is stored in the information system. Subsequently, this is the key information for process mining. The *ActivityOccurrence* contains the *Time* and *Location*. The *Location* is related to a *Path* with the subclasses *SimplePath* and *CombinedPath* [4] and linked to a *Storage*. Using this set of information, the flow of a *UnitLoad* across the storages and time can be mined.

The *Resource* object describes the relevant resources to support the execution of logistics activities (c.f. Figure 3). Similar to the definition of the activities, we defined a *MaterialFlowResource*, *InformationFlowResource*, *Human*, *Infrastructure* and *Energy*. For example, the *Storage* can be a continuous or discrete storage (e.g. high rack).

The object *UnitLoad* (c.f. Figure 4) is defined according VDI 3968 and contains one or many packages (*Package)* and is transformed by material flow activities (*MaterialFlowActivity*). The *Package* combines both the *Packaging* (e.g. bin) and any good (e.g. variant or sub-assembly). Thus, we defined a *ProductComponent*, a *ProductFamily* and a *Part* inspired by [28, 29]. For internal logistics, the *ProductComponent* is described using a unique part number and its name. For enabling process mining, we integrated additional relevant data properties (e.g. price).

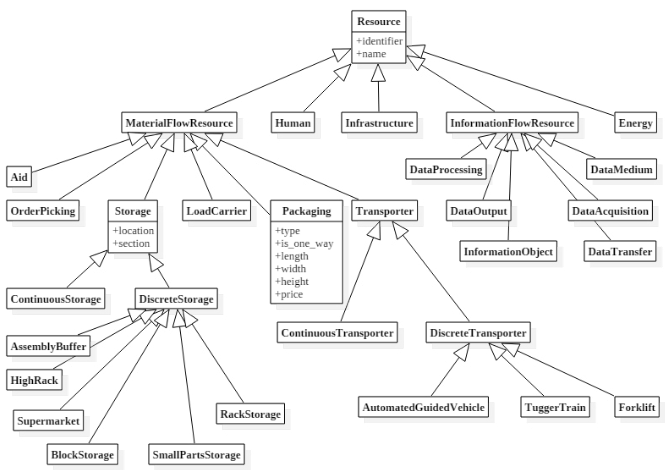

Fig. 3. Extract of the Resource object.

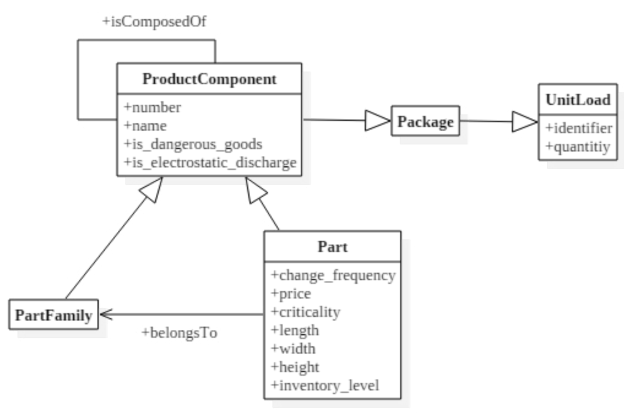

Fig. 4. Extract of the UnitLoad object.

6. Conclusion and Outlook

The step of preprocessing data for process mining is challenging in practice. The data is often not correlated as events and significant effort is required to create event logs. Main reasons are (1) large variety of complex information systems, (2) numerous tables and relations (3) different levels of abstraction, (4) vocabulary used in the IT implementation differs from domain knowledge and (5) volume and veracity of data in internal logistics is significantly high.

While there exist various methods to support the preprocessing, the quality and purpose of the domain ontology is critical. Existing ontologies in the domain of internal logistics are still not complete due to: (1) the partial focus of the ontologies without an integrated view of process and resource, (2) the different level of detail and abstraction, (3) the missing standardization of taxonomy and (4) the missing integration of existing upper ontologies.

This paper presents an internal logistics ontology based on existing ontologies with a focus on the preprocessing of data for process mining. The main contribution of this paper is a systematic approach to re-use existing ontologies and to develop an internal logistics ontology for preprocessing within process mining. We analyzed 42 relevant ontologies in the area of internal logistics to derive the main objects of process, resource, actors, load unit, product and part.

However, there is still need for research. As we only developed the ontology for high model-mix assembly line production application scenario, further case studies are required. Further on, semantics of values in information systems differ from company to company and the integration of application ontologies can be beneficial. This could also provide support for further automation of the preprocessing steps.

Acknowledgements

The Bayerische Motoren Werke Group (BMW Group) funds this research. We extend our sincere thanks to BMW for the generous support of the work described in this research.

References

[1] Timm G, Herbert K. Comprehensive logistics. Berlin, Heidelberg: Springer Berlin Heidelberg; 2009.

[2] Boysen N, Emde S, Hoeck M, Kauderer M. Part logistics in the automotive industry: Decision problems, literature review and research agenda. European Journal of Operational Research 2015; 242(1):107-120.

[3] Sali M, Sahin E. Line feeding optimization for Just in Time assembly lines: An application to the automotive industry. International Journal of Production Economics 2016; 174:54-67.

[4] Negri E, Perotti S, Fumagalli L, Marchet G, Garetti M. Modelling internal logistics systems through ontologies. Computers in Industry 2017; 88:19-34.

[5] Van Der Aalst W, et al. Process mining manifesto. International Conference on Business Process Management 2010; 99:169-194.

[6] Rozinat A, de Jong I, Gunther C, Van Der Aalst W. Process mining applied to the test process of wafer scanners in ASML. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews) 2009; 39(4):474-479.[7] Park J, Lee D, Zhu J. An integrated approach for ship block manufacturing process performance evaluation: Case from a Korean shipbuilding company. International Journal of Production Economics 2014; 156:214-222.

[8] Tu T, Song M. Analysis and prediction cost of manufacturing process based on process mining. Industrial Engineering, Management Science and Application (ICIMSA) 2016;1-5.

[9] Lee S, Kim B, Huh M, Cho S, Park S, Lee D. Mining transportation logs for understanding the after-assembly block manufacturing process in the shipbuilding industry. Expert Systems with Applications 2013; 40(1), 83-95.

[10] Er M, Astuti H M, Wardhani I R K. Material movement analysis for warehouse business process improvement with process mining: a case study. Asia-Pacific Conference on Business Process Management 2015; 219:115-127.

[11] Wang Y, Caron F, Vanthienen J, Huang L, Guo Y. Acquiring logistics process intelligence: Methodology and an application for a Chinese bulk port. Expert Systems with Applications 2014; 41(1):195-209.

[12] Calvanese D, Montali M, Syamsiyah A, Van Der Aalst W. Ontology-driven extraction of event logs from relational databases. International Conference on Business Process Management 2015; 140-153.

[13] Ingvaldsen J, Gulla J. Preprocessing support for large scale process mining of SAP transactions. International Conference on Business process management 2007; 4928:30-41.

[14] Knoll D, Prüglmeier M, Reinhart G. Material flow analysis using transport orders. wt-online 2017; 3:129-133.

[15] Jareevongpiboon W, Janecek P. Ontological approach to enhance results of business process mining and analysis. Business Process Management Journal 2013; 19(3):459-476.

[16] Gruber T. Toward principles for the design of ontologies used for knowledge sharing?. International Journal of Human-Computer Studies 1995; 43(5-6):907-928.

[17] Kowalski M, Quink N. Erstellung einer Ontologie zum Themenkomplex Verpackungen in der Logistik mithilfe des Ontologie-Editors Protégé. OrGoLo-Projektbericht 2013; 16.

[18] Libert S, Chisu R, Keutner K. Eine Ontologie für das Internet der Dinge. Internet der Dinge in der Intralogistik. Berlin, Heidelberg: Springer Berlin Heidelberg; 2010.

[19] Merdan M, Koppensteiner G, Hegny I, Favre-Bulle B. Application of an ontology in a transport domain. Industrial Technology 2008; 1-6.

[20] Bock C, Gruninger M. PSL: A semantic domain for flow models. Software & Systems Modeling 2005; 4(2):209-231.

[21] Kumar H, Park P. Know-ont: a knowledge ontology for an enterprise in an industrial domain. International Journal of Database Theory and Application 2010; 3(1):23-32.

[22] Guizzardi G, Falbo R, Guizzardi R. Grounding Software Domain Ontologies in the Unified Foundational Ontology (UFO): The case of the ODE Software Process Ontology.CIbSE 2008; 127-140.

[23] Noy N, McGuinness D. Ontology development 101: A guide to creating your first ontology. 2001.

[24] Kalfoglou Y, Schorlemmer M. Ontology Mapping: The State of the Art. The Journal of Semantic Interoperability and Integration 2005; 18:1-31.

[25] Noy N, Musen M. The PROMPT suite: interactive tools for ontology merging and mapping. International Journal of Human-Computer Studies 2003; 59(6):983-1024.

[26] Stumme G, Maedche A. FCA-Merge: Bottom-up merging of ontologies. IJCAI 2001; 1:225-230.

[27] Ye Y, Yang D, Jiang Z, Tong L. Ontology-based semantic models for supply chain management. The International Journal of Advanced Manufacturing Technology 2008; 37(11-12):1250-1260.

[28] Panetto H, Dassisti M, Tursi A. ONTO-PDM: Product-driven ONTOlogy for Product Data Management interoperability within manufacturing process environment. Advanced Engineering Informatics 2012; 26(2):334-348.

[29] Bruno G, Antonelli D, Villa A. A reference ontology to support product lifecycle management. Procedia CIRP 2015; 33:41-46.

[30] Günthner W, Boppert J. Lean Logistics. Berlin, Heidelberg: Springer Berlin Heidelberg; 2013.

[31] Knössl T. Gestaltung schlanker Prozessschnittstellen in automobilen Logistikketten.Diss, Technical University of Munich; 2015.