3rd Conference on Production Systems and Logistics

Traceability System's Impact On Process Mining in Production

Markus Schreiber[1], Joachim Metternich[1]

[1] *Institute for Production Management, Technology and Machine Tools (PTW), Otto-Berndt Str. 2, 64287 Darmstadt, Germany*

Abstract

From the perspective of manufacturing companies, data handling is gaining more attention as it is becoming a strategic resource in digital ecosystems. Market forces such as rising amounts of product variants and decreasing batch sizes lead to higher complexity in manufacturing processes. Therefore, production management's demand for data-based process transparency is growing continuously as well as the number of companies turning to process mining to address these challenges. The increased use of process mining has uncovered many documented data quality issues that hamper output quality.

In response to data usage and quality problems, research in the field of Big Data has turned to sophisticated data value chains as a promising approach to optimize data usage. This paper presents the application of the data value chain concept on a manufacturing use case, delivering an assessment of traceability systems and their effect on data quality issues. This assessment reviews commonly known quality issues and investigates how traceability systems can influence and facilitate better data quality. The results support manufacturing companies in their use of traceability systems to improve the reliability of their process mining input data and, hence, their output performance indicators to meet the demand for more data-based process transparency.

Keywords

Data value chain; traceability system; process mining, data-based process transparency; data quality

1. Introduction

Industrial companies are challenged by rising complexity in their manufacturing processes. Good examples for complexity drivers are the constantly rising demand for more individualised products on the markets as well as short product lifecycles and delivery times. In response to competitors, companies create more product variants to stay attractive on their markets [2].

In the context of Industry 4.0, data-based transparency is needed to tackle complexity and support effective managers' decision-making [4]. To address process complexity, the use of process mining has become more popular amongst manufacturing companies. In a Deloitte study in 2020, out of 104 interviewed companies, 40% have stated to use process mining in production, aiming for process transparency and improvements as their top two reasons [6]. Still, the use of process mining is particularly challenging in production. Typically, production is characterized by numerous different processes that complicate identifying available data sources [8]. Available as well as reliable data are essential requirements for input datasets called "event logs" to conduct successful process mining analysis. Process mining projects tend to fail due various data quality issues such as missing and unreliable input data points that result in insufficient digital traces [9]. In his

DOI: https://doi.org/10.15488/12136

research, Jahn identifies data acquisition as the key factor for improving availability and reliability of input data and for gaining data-based transparency, the basis to optimizing production processes [10].

Within the vision of smart factories and smart products, automatic identification (autoID) technologies such as RFID are being used to generate data and gain transparency [13]. Although most manufacturing companies use these so called traceability systems due to legal obligations [14] or to inventory existing objects, the potential of data acquisition through autoID technologies is still not fully reached. Research in the field of traceability often focuses on tracking or tracing objects themselves, giving insights on how to consistently mark objects in production processes, find fitting technologies to track objects or products [14], or identify effort versus benefit levels that should be considered when tracking different product categories [16]. For successful production management, it has become crucial to focus on the data application perspective. A recent research project demonstrated several beneficial use cases originating from the use of a traceability system and its generated data [18]. In this context, companies still lack the knowledge to generate targeted feedback data of their processes using the traceability system and its ability to locate objects [4].

Traceability systems generate process data and can function as an important data supplier in production [16]. From a theoretical point of view, the combination of traceability as a data generating system and process mining as the tool for data analysis offers great potential for data-based process transparency [20]. However, researchers have not yet investigated whether and how traceability systems can avoid the occurrence of quality issues in input datasets. Based on a manufacturing use case, this paper aims to investigate the ability and impact of a traceability system to avoid common quality issues and improve the reliability of process mining outputs.

2. Approach

The paper is divided into two main sections. Section 3 addresses the conceptual development to identify data quality issues (QIs) that can potentially be affected by the traceability system in the manufacturing use case (section 3.3). This requires two tasks: Firstly, the explanation of the use case's data value creation process by introducing the data value chain (DVC) concept and the assignment of the use case to the phases of the DVC (section 3.1). Secondly, an overview about what kind of QIs exist and where the QIs occur in the DVC (section 3.2). Section 4 presents the analysis of the manufacturing use case. At first, the process mining input dataset (event log) and the obtained outputs based on the traceability data is introduced (section 4.1). Eventually, section 4.2 analyses the traceability system's impact to avoid the five most relevant QIs in the use case and to ensure reliable process mining outputs.

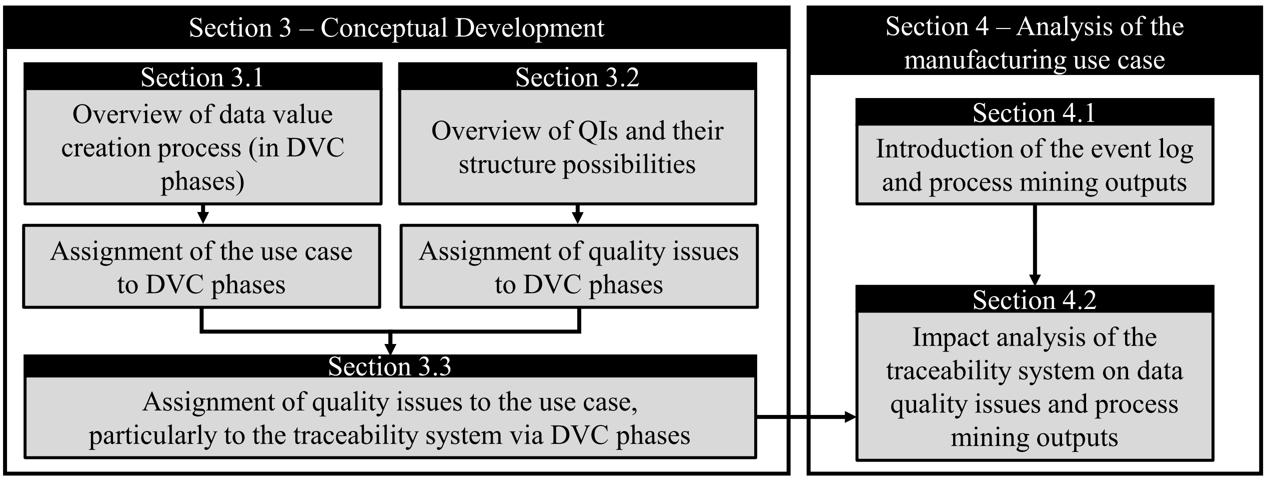

Figure 1: Approach of the conducted research

3. Conceptual development

3.1 Process of data value creation in the manufacturing use case

To enable the process of data value creation in companies the data value chain (DVC) concept is used. It represents a promising approach to improve the handling of data management. It originates from Big Data research and helps decision makers to adopt the data perspective on business processes in order to optimize data usage [12]. Generally, a DVC considers strategically important, value-creating activities [13] and integrates all data-effecting steps, starting with the generation and acquisition of data and ending with the possibility of decision-making based on data outputs [14,15]. Research shows that the representation of DVCs in literature differ in regard to the number of phases in the chain and the individually intended functions of each phase depending on the area of application [12,13,14,15,16].

To assess the impact of the traceability system on process mining in the manufacturing use case, the DVC is used to structure all relevant elements according to their role and function in the process of data value creation. These elements including the "configured traceability system", the "available traceability data", the "complete use case dataset" including traceability and more sensor data, the "transformation to event log" (filtered input dataset), "the process mining analysis" and finally "the process mining output". Figure 1 illustrates the application of the DVC concept on the manufacturing use case analysed in this paper. It suggests six phases that can be derived from the analysed sources considering the identified reoccurring patterns and functions explained in the grey boxes of each phase.

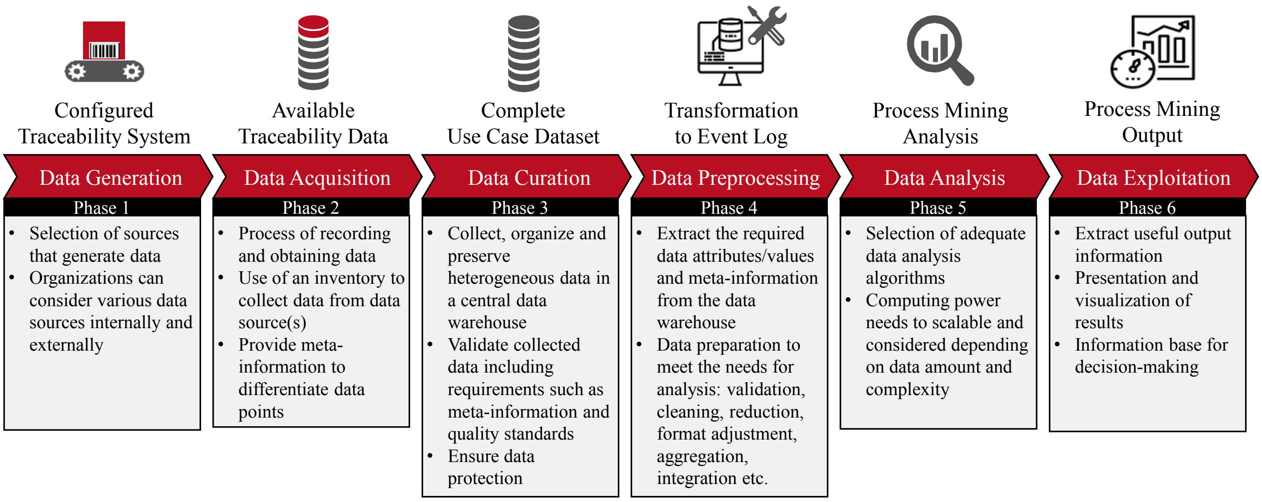

Figure 2: Data value creation process based in the manufacturing use case

3.2 Literature review of typical data quality issues

The high quality of data is the key to interpretable and trustworthy data analytics and the basis for meaningful outputs. The process mining manifesto confirms this interdependency and stresses the need for high-quality event logs (representing the input data) in the context of process mining [3]. Typically, a company's decision to use an analytics tool such as process mining is made without paying attention to the possibly poor quality of the available event log which results in poor quality outputs – a dynamic often characterized by the term "garbage in – garbage out" [5].

The poor quality of event logs is a known problem amongst companies. A study by Suriadi et al. identifies several imperfection patterns from their experience with over 20 Australian industry datasets which confirm the severity of data QI in process data and their potential impact on process mining analysis [1]. Another study by the Meta Group revealed that 41% of the data exploration projects fail, mainly due to insufficient data quality, leading to misinformed decisions [7].

The goal of the conducted literature review is to identify typical QIs that can possibly be affected by the traceability system. The following criteria were applied in the literature search:

- A: Which literature source provide a collection of data QI?
- B: Do these QIs refer specifically to process mining?
- C: Do these QIs refer generally to data analysis?
- D: Are QIs structured in categories/dimensions based on commonalities?
- E: Are QIs structured by their location of origin?

Table 1: Literature Review

Sou-rce	[1]	[3]	[5]	[7]	[11]	[12]	[15]	[17]	[19]	[21]	[22]	[23]	[24]	[25]	[26]	[27]	[28]	[29]	[30]	[31]	Σ
A	•		•	•							•	•	•	•	•		•			•	**10**
B	•	•	•			•			•								•				**6**
C				•	•		•	•		•	•	•	•	•	•	•		•	•	•	**14**
D	•		•	•	•			•		•	•	•	•	•	•	•	•	•	•	•	**16**
E					•		•		•						•						**4**

Criteria A of Table 1 shows that there are several literature sources that provide collections of QIs. In literature sources related to process mining (criteria B) and general data analysis (criteria C), the amount of identified QIs is so large that a useful structure is required in order to tackle them systematically. Most studies (criteria D) provide a structure of QIs based on categories or dimensions. The general benefit of this structure is the fact that related QIs are gathered into the same group. However, this structure implicates a major disadvantage, as it hampers the search for the origin of the QIs. In contrast, the literature sources based on criteria E organize data QIs along the phases of the DVC as introduced in Figure 2. This approach has two main advantages, firstly it allows the identification of the QIs root cause by locating their places of origin [26], secondly, the structured QIs along the phases of the DVC can be used to connect the use case and its traceability system as it is linked to the DVC as well. It represents the basis to assess the traceability system's capability to avoid the occurrence of QIs.

Due to the large number of identified QIs, Singh et al. provide a study that summarizes a broad collection of data QIs acquired from extensive research in that field [26]. Based on the collected QIs, they suggest four groups of root causes for QIs that can be assigned to their places of origin in the phases of the DVC. The identified groups are the following [26]:

- **Group 1: QI based on data sources** – A leading cause for data QI is to obtain the wrong or poor data. On the one hand, every individual data source needs to be configured thoroughly to provide the data needed. On the other hand, various different data sources are likely to be inconsistent and cause difficulties in subsequent phases of the DVC.
- **Group 2: QI based on data profiling** – Once data sources are selected, the data profiling of every source system (e.g. traceability system, ERP, CRM, Web, etc.) needs to be examined to avoid negative impacts on data quality. The profiling is a fundamental step in which every individual source system as well as the gathered data of all source systems in a central data warehouse ensure data integrity and consistency for later analysis.
- **Group 3: QI based on data staging and ETL** (extraction, transformation, loading) – In this phase QI occur firstly in the central data warehouse when the data and metadata from all source systems is audited and validated and, secondly, in the pre-processing phase when a dataset is extracted, transformed and loaded for the following data analysis.
- **Group 4: QI based on data modeling** – If no major QI is detected up to this point and the available dataset demonstrates high quality, the data modeling itself can cause QI for two main reasons. The first

occurs when the dataset is not successfully transformed to fulfil the input specifications for the intended data analysis. The second can occur when the selection of the data analytics application is inadequate and obtains no or useless results.

3.3 Assignment of quality issues to the traceability system

The following step intends to determine which of the four identified groups of QI can be affected by the traceability system. Therefore, the groups of QIs mentioned above need to be linked to the DVC in the use case. Through the assignment of the identified groups of QIs and the manufacturing use case (see section 3.1) to the phases of the DVC, it is possible to link them as illustrated in Figure 2. This approach allows to break down the large amount of QI and assess a potentially positive impact of traceability systems on data quality and thus improve the process mining output.

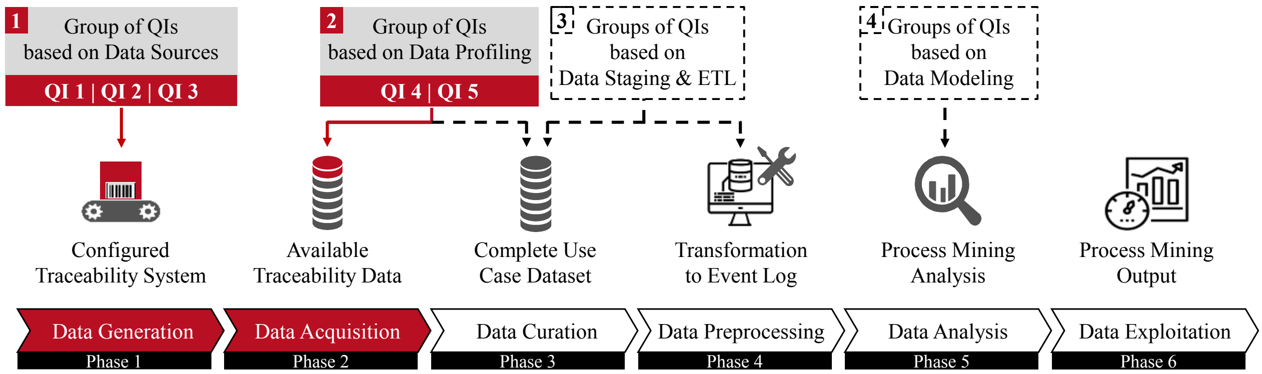

Figure 3: Link of QI groups to traceability system via DVC

Figure 3 demonstrates the result of connecting the groups of QIs as well as the elements of data value creation process in the use case to the DVC phases. As the research in this paper investigates the traceability system's impact on QIs, group 3 "QI based on data staging & ETL" and group 4 "QI based on data modelling" are not further considered in the analysis (shown black and white in Figure 3). There is a large number of identified single QI based on data sources (group 1) and data profiling (group 2) [26]. For handling purposes, the five most relevant QI (QI 1 – QI 5) for this use case are selected to explain the positive impact of the traceability system in connection to the conducted process mining analysis in section 4.2.

4. Analysis of the manufacturing use case

4.1 Introduction of the event log and the process mining outputs

The dataset in the use case was generated in the transfer project called "ArePron" (www.arepron.com). It represents a discrete production network involving parallel machine resources and consists of traceability data as well as sensor-based machine data (pressure consumption, electrical power consumption etc.). To investigate the traceability system's ability to prevent the emergence of QIs and to contribute to reliable process mining results, the machine sensor data is filtered and the traceability data providing process information remains to be used as input data for the process mining analysis.

The application of process mining requires a dataset as input data that contains at least a "case ID" (process trace) including "events" (process activities) and a "time stamp" for each event. Every case must be provided in a separate line [1]. The extracted traceability data from the dataset in this use case is shown in Table 2. Every individual "component No." functions as case ID, while the "machine name" and "process" as event, and "start scan" and "end scan" as time stamp. The event log is created in pre-processing (phase 4 of DVC), which mainly consists of format adjustments of the original dataset, so that any event (machining process) is given in a separate line.

Table 2: Use case event log with traceability data

Component No.	Machine name	Process	Start Scan	End Scan	[…]*
1042	Kasto	Sawing	08.03.2020 09:30	08.03.2020 09:38	
1042	OP10	Turning	08.03.2020 10:15	08.03.2020 10:42	
1042	OP20	Milling	08.03.2020 11:03	08.03.2020 11:23	
1042	[…]**				
1043	Kasto	Sawing	08.03.2020 09:40	08.03.2020 09:52	
1043	HaasST10	Turning	08.03.2020 10:31	08.03.2020 10:58	
1043	HaasMM2	Milling	08.03.2020 11:23	08.03.2020 11:48	
1043	[…]**				
1044	Kasto	Sawing	08.03.2020 11:00	08.03.2020 11:14	
1044	OP10	Turning	08.03.2020 11:27	08.03.2020 11:56	
1044	OP20	Milling	08.03.2020 12:29	08.03.2020 12:50	
1044	[…]**				

*Filtered data points (pressure consumption, electrial power consumption etc.) irrelevant for process analysis
**Following manufacturing steps in the same structure

The following represents a selection of exemplary process mining outputs shown in Figure 4 that are obtained from the traceability data:

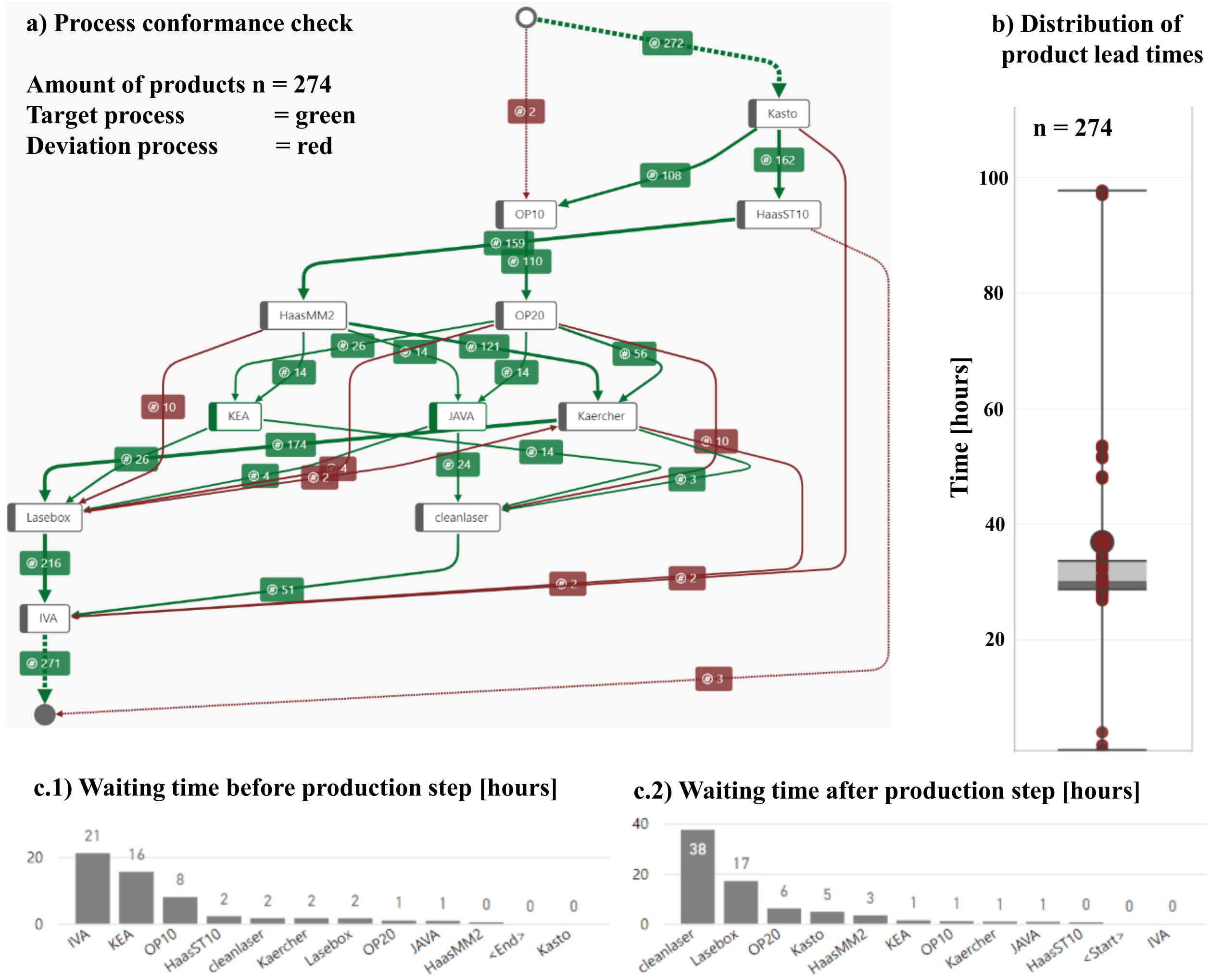

Figure 4: Exemplary process mining outputs

a) The standard process mining outputs process discovery and process conformance check are obtained. The conformance check in Figure 4 shows the amount of products that have taken the target process in green and those that deviated from their target process in red.
b) The available time stamp (either start or end time stamp) enable various output options regarding process lead time analyses, such as average lead time of all products, average lead time per production path, etc. The exemplary output distribution analysis of individual product lead times is shown in a boxplot in Figure 4. Each red point represents one product; the bigger the point, the more products were product in this lead time.
c) The availability of two time stamps (start and end time of each production step) allows for the analyses of the products' waiting times before and after every production step. Those indicators can hint at problems within production and can, for example, identify bottlenecks in the production network.

4.2 Impact analysis of the traceability system

To assess the traceability system's capability to avoid QIs and to ensure reliable process mining outputs, the impact analysis is conducted individually for every QI. At first, the analysis names the individual QI and outlines its relevance, then it explains the traceability system's impact on the QI and eventually on the process mining output (shown in Figure 4):

- **QI 1:** Inadequate selection of data sources do not provide needed input data
  Relevance of QI 1: The goal in the project is to analyse the performance of the production network and to receive transparency about the products' path within the production network, their lead time through the production process, wasteful waiting times etc. To receive the right process mining outputs, the selection of data sources is the most crucial success factor contributing to the project goal by improving the availability and reliability of input data and thus by gaining data-based transparency [10]. In case the planning and selection of needed source systems and their data points is neglected or has not been conducted at all, this QI is likely to result in missing analysis outputs. The effort to handle this QI is high, as the implemented hardware (source system, sensors etc.) needs to be changed and re-implemented.
  Impact of traceability system on QI 1: Generally, a traceability system's configuration determines the process data being captured, and what process mining outputs can be obtained. In this use case, the traceability system functions as the only source system to generate the process data shown in Table 2 and is responsible to ensure the performance evaluation by enabling the creation of the intended outputs. The traceability data required and generated are the "component ID" to capture the product, the "machine ID" to identify the taken paths through the production network, one "time stamp before the start" and another "time stamp after the end of a production process" (see Table 2).
  Impact of traceability system on process mining: The "component ID" functions as the case ID of the event log and represents the products trace through the production network. The "machine ID" functions as the event and determines the actual production stations passed by the product. The "time stamps" of component ID and machine ID captured in the production network help to order the different steps through production. As case ID, event and time stamp are the minimum requirement to create an event log, the exemplary process mining outputs such as process discovery, conformance check, lead times analysis or waiting time analysis shown in Figure 4 would not have been obtained without the traceability system.
- **QI 2:** Missing data values in data source system
  Relevance of QI 2: Depending on the implemented source system in each production case, the likeliness of missing some relevant data points in operational practice is high and therefore an important QI to be considered.
  Impact of traceability system on QI 2: Incomplete data generation with missing values is more likely to appear in a manually working traceability system than in a highly automated one. The ability to avoid missing data depends highly on the automation level of the traceability system and the operator's

reliability in the case of a high manual level. In this use case a manual traceability system with hand scanners is implemented. An important measure for the successful and complete data acquisition is to instruct all users of the traceability system on its correct handling. Relying on the manual handling of the traceability system and a planned acquisition of 3.288 data points in 274 cases (sum of manufactured products), only 7 data points were missing. That's an error ratio of about 0.21%.

Impact of traceability system on process mining: Considering the error ratio of 0.21% in the event log, the process mining analysis and output is practically not affected by the few missing values.

- **QI 3:** Misspelled data in data source system

  Relevance of QI 3: Misspelled data points can occur especially when implemented source systems require manual system inputs by operators. Large amounts of misspelled data in a dataset may cause major efforts in the pre-processing phase when detected, otherwise subsequent data analysis become obsolete.

  Impact of traceability system on QI 3: Traceability systems offer technological possibilities such as optical labels (e.g. data matrix code) or RFID tags that save identification numbers and transfer those when captured via optical scanner or RFID receiver to a source system. In this use case, a data matrix code is used to provide the component ID and the machine ID so that misspelled data cannot occur during data acquisition.

  Impact of traceability system on process mining: Using the technological options to save relevant data in optical codes or RFID tags, no manual inputs into the traceability software are required. As result, there is no misspelled data available in the event log.

- **QI 4:** Insufficient data profiling of data sources such as lack of data validation routines at source system

  Relevance of QI 4: As Figure 3 demonstrates, the first opportunity to perform data profiling is possible at phase 2 (data acquisition) at the source system, such as the traceability system in this use case. The second opportunity is at phase 3 where a combined dataset is formed out of several possible source systems in a data warehouse. Validation routines represent data capturing rules that support the acquisition of the right data as needed. When applied in phase 2 and 3, the risk of crating QI is counteracted.

  Impact of traceability system on QI 4: In the use case, the traceability system is used for data profiling to ensure the high quality of the generated data. Therefore, data validation routines are embedded in the traceability system. This is even more important when the traceability system is operated manually and errors in the data generation phase are more likely to happen. For instance, the system captures only data points if the specified scan sequence is followed. For the traceability system to capture every individual production step of a product as valid data point, the machine ID must be scanned firstly, the component ID secondly. Additionally, both scans need to be performed within 10 seconds. This way the accidental capture of data can be avoided.

  Impact of traceability system on process mining: The used validation routines add significantly to the generation of a high quality event log. Component IDs or machine IDs are not obtained as individual data points so that the case ID and the corresponding event are always saved together. This way, outputs such as discovery, conformance check, lead times etc. are not distorted.

- **QI 5:** Hand coded data profiling is likely to be incomplete or results in an inappropriate selection of automated profiling rules

  Relevance of QI 5: QI 4 and QI 5 are related. QI 4 represents the conceptual level of what validation rules are useful to preserve the same data quality. QI 5 refers to the technical implementation of the validation routines via coding. Potentially, the selected validation routines embedded as rules in the system can declare correct data points as invalid and do not capture them.

  Impact of traceability system on QI 5: To ensure the coded rules in the traceability system contribute positively towards data quality and data completeness, tests with possible errors have been conducted to analyse if the implemented rules in the system function as expected and do not cause new QI. Moreover, the traceability system is designed to support operators by giving feedback if the intended scanning process is performed correctly and valid data is generated successfully. This gives system users the chance to verify if the system captures the correct data.

Impact of traceability system on process mining: The result of embedding coded rules for the automated validation of generated data in the traceability system is a reliable event log providing useful outputs as shown in Figure 4.

The analysis of the manufacturing use case demonstrates the high impact of traceability systems enabling the process mining analysis by generating the required process data. It outlines that the traceability system has the capability to improve the data handling by avoiding or at least minimizing the risk of QIs to occur and hence ensuring the reliability of the obtained outputs.

Summary and Outlook

This paper investigates the impact of traceability systems on data quality issues (QIs) and process mining results, based on a manufacturing use case in a production network. First, the connection between traceability and process mining is explained through the data value chain (DVC) concept in six phases. A thorough literature review results in the identification of four groups of QIs that are distinguishable by the location in which they occur along the DVC. Considering the application of the DVC phases on the use case, there are two (out of four) groups of QIs, "QIs based on data sources" and "QIs based on data profiling" that can be assigned to the traceability system and hence be positively affected by it. The investigation of five concrete QIs out of the two groups confirms that traceability systems can avoid QIs and improve the number and reliability of process mining outputs.

Traceability systems have a great potential to provide process data that enable transparency through process mining analysis in production. Due to growing complexity and more frequent use of process mining in production, traceability systems are not only relevant for commonly known purposes such as recall campaigns, but also take on an important role in today's data-based production management. As a supplier of valuable process data, they have the capability to enable transparency through process mining in production, firstly by providing the selected data points needed and secondly by its ability to prevent the occurrence of QIs.

Future research in the field of traceability needs to develop a scientific approach that allow companies the target use of traceability systems as a data supplier in production. This approach needs to address the question of how to configure a traceability system in order to maximises the number of process mining outputs and hence, the gained data-based transparency. At the same time, the traceability system's ability to avoid the potential occurrence of QIs needs to be considered in this approach, so that it contributes to reliable and high quality results of process mining analyses.

References

[1] Suriadi, S., Andrews, R., Hofstede, A. ter, Wynn, M.T., 2017. Event log imperfection patterns for process mining: Towards a systematic approach to cleaning event logs. Information Systems 64.

[2] Gottmann, J., 2019. Produktionscontrolling. Springer Fachmedien Wiesbaden, Wiesbaden.

[3] Van der Aalst et al., 2012. Process Mining Manifesto. International conference on business process management.

[4] G. Schuh, R. Anderl, R. Dumitrescu:A. Krüger:M. ten Hompel, 2020. Der Industrie 4.0 Maturity Index in der betrieblichen Anwendung – aktuelle Herausforderungen, Fallbeispiele und Entwicklungstrends (acatech Kooperation), München.

[5] Bose, R.J.C., Mans, R.S., van der Aalst, W.M. Wanna improve process mining results?

[6] Galic, G., Wolf, M., Salzmann, O., Unger, T., 2021. Delivering Value with Process Analytics: Process Mining adoption and success factors.

[7] Oliveira, P., Rodrigues, F., Henriques, P.R., 2005. A Formal Definition of DQ Problems.doc.

[8] Flack, C., Dreher, S., Birk, A., Wilhelm, Y., 2020. Process Mining in der Produktion: Spezifische Herausforderungen bei der Anwendung 115 (11).

[9] Reinkemeyer, L., 2020. How to Get Started, in: Reinkemeyer, L. (Ed.), Process Mining in Action. Springer International Publishing, Cham.

[10] Jahn, M., 2017. Industrie 4.0 konkret. Springer Fachmedien Wiesbaden, Wiesbaden.

[11] El Alaoui, I., Gahi, Y., Messoussi, R., 2019. Big Data Quality Metrics for Sentiment Analysis Approaches, in: Proceedings of the 2019 International Conference on Big Data Engineering. BDE 2019: 2019 International Conference on Big Data Engineering, Hong Kong Hong Kong. 11 06 2019 13 06 2019. Association for Computing Machinery, New York,NY,United States.

[12] Mocnik, F.-B., Zipf, A., Fan, H., 2017. Data Quality and Fitness for Purpose.

[13] Bitkom-Gremium - AK Big Data, 2015. Leitlinien für den Big-Data-Einsatz. https://www.bitkom.org/sites/default/files/file/import/150901-Bitkom-Positionspapier-Big-Data-Leitlinien.pdf. Accessed 13 November 2021.

[14] Wank, A., 2019. Methodik zur Wertstromintegration einer aktiven Bauteilrückverfolgung in die diskrete Variantenfertigung. Shaker, Herzogenrath.

[15] Rajpurohit, A., 2013. Big data for business managers — Bridging the gap between potential and value, in: 2013 IEEE International Conference on Big Data. Silicon Valley, California, USA, 6 - 9 October 2013. 2013 IEEE International Conference on Big Data, Silicon Valley, CA, USA. 06.10.2013 - 09.10.2013. IEEE, Piscataway, NJ.

[16] ZVEI. ZVEI-Traceability-Initiative "Traceability-Levels für Produktkategorien".

[17] Batini, C., Cappiello, C., Francalanci, C., Maurino, A., 2009. Methodologies for data quality assessment and improvement. ACM Comput. Surv. 41 (3).

[18] Urnauer, C., Schreiber, M., Bausch, P., Metternich, J., 2021. Anwendungen aktiver Traceability-Systeme: Datennutzung in der digitalisierten Produktion. ZWF - Zeitschrift für wirtschaftlichen Fabrikbetrieb 116 (3).

[19] Ceravolo, P., Azzini, A., Angelini, M., Catarci, T., Cudré-Mauroux, P., Damiani, E., Mazak, A., van Keulen, M., Jarrar, M., Santucci, G., Sattler, K.-U., Scannapieco, M., Wimmer, M., Wrembel, R., Zaraket, F., 2018. Big Data Semantics.

[20] Schreiber, M., Bausch, Phillip, Best, Julian, Metternich, J., 2020. Datenanalyse in Produktionsprozessen: Potenziale und Herausforderungen des Process-Mining-Einsatzes in Theorie und betrieblicher Praxis. ZWF - Zeitschrift für wirtschaftlichen Fabrikbetrieb 115 (5).

[21] Ehrlinger, L., Rusz, E., Wöß, W., 2019. A Survey of Data Quality Measurement and Monitoring Tools.

[22] Rahm, E., Do, H.H., 2000. Data Cleaning: Problems and Current Approaches.

[23] Karkouch, A., Mousannif, H., Al Moatassime, H., Noel, T., 2016. Data quality in internet of things: A state-of-the-art survey. Journal of Network and Computer Applications 73.

[24] Laranjeiro, N., Soydemir, S.N., Bernardino, J., 2015. A Survey on Data Quality: Classifying Poor Data.

[25] Ge, M., Helfert, M. A Review Of Information Quality Research.

[26] Ranjit Singh, Dr. Kawaljeet Singh, 2010. A Descriptive Classification of Causes of Data Quality Problems in Data Warehousing.

[27] Strong, D.M., Lee, Y.W., Wang, R.Y., 1997. Data quality in context. Commun. ACM 40 (5).

[28] Verhulst, 2016. Evaluating quality of event data within event logs an extensible framework.

[29] Wang, R.Y., Reddy, M.P., Kon, H.B., 1995. Toward quality data: An attribute-based approach. Decision Support Systems 13 (3-4).

[30] Wang, R.Y., Strong, D.M., 1996. Beyond Accuracy: What Data Quality Means to Data Consumers. Journal of Management Information Systems 12 (4).

[31] Woodall, P., Oberhofer, M., Borek, A., 2014. A classification of data quality assessment and improvement methods. IJIQ 3 (4).

Biography

Markus Schreiber, M. Sc. has been a researcher and PhD student at the Institute of Production Management, Technology and Machine Tools (PTW) at the Darmstadt University of Technology since 2018. His research interests include the traceability of components and equipment, as well as the optimization of production processes.

Prof. Dr.-Ing. Joachim Metternich has been the head of the Institute for Production Management, Technology and Machine Tools (PTW) at the Darmstadt University of Technology since 2012. In addition, Prof. Metternich is the spokesman for the Center of Excellence for Medium-Sized Businesses in Darmstadt and president of the International Association of Learning Factories.