# *Data-Driven Business Process Management:*

# *Advancing Process Data Quality and Process Improvement*

*Dissertation*

*zur Erlangung des Grades eines Doktors der Wirtschaftswissenschaft*

*der Rechts- und Wirtschaftswissenschaftlichen Fakultät*

*der Universität Bayreuth*

*Vorgelegt*

*von*


*Christopher van Dun*

*aus*

*Bamberg*

| | |
|---|---|
| Dekan: | Prof. Dr. Jörg Schlüchtermann |
| Erstberichterstatter: | Prof. Dr. Maximilian Röglinger |
| Zweitberichterstatter: | Prof. Dr. Moe Wynn |
| Datum der mündlichen Prüfung: | 04.04.2022 |

*“Everything is a process.”*

*Redewendung (nach Heraklit)*

---

*Mit dem Abschluss dieser Dissertation geht für mich eine sehr lange, vielseitige und lehrreiche universitäre Phase meines Lebens zu Ende. Unglaubliche elf Jahre habe ich in Vorlesungssälen, Seminarräumen, CIP-Pools und an verschiedenen Lehrstühlen geschwitzt, gehadert und zum Glück auch ab und an gejubelt. Ich blicke schon jetzt mit einem lachenden und einem weinenden Auge zurück – einerseits sehr erleichtert, dass ich es bis hierhin geschafft habe, andererseits traurig, dass hier das Kapitel nun abgeschlossen werden muss. Jetzt geht es aus der behüteten akademischen Umgebung in die „echte“ Welt hinaus und ich möchte die Gelegenheit dieses Meilensteins nutzen, um von ganzem Herzen Danke zu sagen:*

*Danke an meinen Doktorvater Max: Für die Chance, die du mir mit der Promotion bei dir gegeben hast. Für die allzeit direkte und hilfsbereite Betreuung, auch wenn es bei dir selbst gerade mal wieder sehr stressig war. Für die gute Laune, die du selbst mit schlechter Laune verbreitest. Und für die tollen Erfahrungen, die ich in den letzten Jahren machen durfte.*

*Danke an meine Familie, Irmi, Peter und Felix: Danke für all das, was ihr mir mitgegeben habt, und für all das, was ihr mir ermöglicht habt. Danke für eure uneingeschränkte Unterstützung auf meinem Weg – auch wenn ich nicht immer weiß, wohin er mich führt. Danke dafür, dass ihr immer da seid, dass ihr meine Blitzableiter seid, dass ihr mich in guten Zeiten ertragt und in schlechten Zeiten aufbaut. Einfach danke!*

*Danke an Elli: Danke für deinen unerschütterlichen Glauben an mich. Danke für deine Warmherzigkeit, deine bedingungslose Unterstützung und deine Nachsicht, wenn mich die Promotion mal wieder viel Zeit, Energie oder gute Laune gekostet hat. Danke, dass du mein Ruhepol und meine Energiequelle bist. Danke, dass ich bei dir ich sein kann. Danke für dich!*

*Danke an alle Freund*innen, Kolleg*innen und Wegbegleiter*innen in Augsburg und Bayreuth: Dafür, dass ihr die letzten Jahre zu einer so schönen Zeit gemacht habt. Dafür, dass ihr mir immer wieder gezeigt habt, dass die Arbeit nicht alles ist. Ich wünsche euch allen Erfolg der Welt auf eurem Weg!*

# Abstract

Business processes are at the core of every organisation's effort to deliver services and products to customers and, thus, achieve the organisation's goals. The discipline that deals with the design, analysis, execution, and improvement of such business processes is called business process management (BPM). Over the years, the BPM research discipline has created a large number of methods and tools to support practitioners in managing and improving their business processes. In recent years, the increasing abundance of process data available in organisational information systems and simultaneous progress in computational performance have paved the way for a new class of so-called data-driven BPM methods and tools, the most prominent of them being process mining. This cumulative doctoral thesis concentrates on two challenges related to data-driven BPM methods and tools that impede faster and more widespread adoption.

First, while data-driven methods and tools have found quick adoption in BPM lifecycle phases such as process discovery and process monitoring, the lifecycle phase of process improvement has so far been neglected. However, process improvement is considered to be the most value-adding BPM lifecycle phase since it is the necessary step to address existing issues in as-is processes or to adapt these processes to constantly changing environments and customer needs and expectations. Process improvement is often expensive, time-consuming, and labour-intensive, which is why there is a particular need to support process stakeholders in redesigning their processes.

Second, there is a need for high-quality process data in all phases of the BPM lifecycle. In practice, process data, e.g., in the form of event logs for process mining, is often far from the desired quality and process analysts spend the majority of their time on identifying, assessing, and remedying data quality issues. Thus, in the BPM community, the interest in exploring the roots of data quality problems and the related assurance of high-quality process data is rising. Hence, it is essential to have a means for detecting and quantifying process data quality.

Against this backdrop, this cumulative doctoral thesis comprises five research articles that present advances in process data quality management on the one hand and data-driven process improvement on the other hand. Taking on a design-oriented research paradigm and applying different qualitative and quantitative research methods, this thesis proposes several IT-enabled artifacts that support stakeholders in managing process data quality and improving business processes. The insights contained in this thesis are relevant for academia and practice as they provide both scientific perspectives and practical guidance.

Concerning process data quality management, research article #1 presents an approach for (semi-) automated and quality-informed event log extraction from process-agnostic relational databases. It applies metrics for data quality dimensions that are relevant to process mining in order to quantify the data quality of the source data in selected database tables and simultaneously allows users to extract

event logs in XES format from the database tables. Research article #2 presents an approach for detecting and quantifying timestamp data quality issues in events logs already present in XES format. The approach applies metrics for identifying timestamp imperfection patterns and allows users to interactively filter, repair, and annotate the event log.

Furthermore, this thesis provides several concrete approaches to data-driven business process improvement. First, it focuses on process improvement in itself and aims to create artifacts for supporting process improvement initiatives. Therefore, research article #3 provides a model based on generative adversarial networks to create new process designs. Specifically, it uses event logs and annotated information on process variants and process deviance to generate a new process model which provides suggestions for process improvement to the user. Second, this thesis targets data-driven decision support in business processes. In particular, research article #4 uses multi-criteria decision analysis to extend traditional vehicle routing problems in last-mile delivery with a customer-centric perspective. The customer-centric vehicle routing uses process and customer data and the concept of customer lifetime values to predict customer satisfaction and, thus, optimise delivery routes. Finally, research article #5 presents a modelling approach for IT availability risks in smart factory networks based on Petri nets. The modelling approach uses modular components of information systems and production machines to model, simulate, and analyse production processes.

The thesis concludes by pointing to limitations of the presented research articles as well as directions for future research. Overall, this thesis contributes to several important research streams in BPM while applying a broad range of qualitative and quantitative research methods such as simulation, normative analytical modelling, multi-criteria decision analysis, and interview studies within an overarching design science research paradigm. It builds upon and extends existing research on process data quality management and business process improvement.

# Table of Contents

# I. Introduction[1]

## 1 Motivation

Business processes are an organisation's way of delivering services and products to customers and, as such, they are key to organisational success (Dumas et al. 2018; Gross et al. 2019). They are defined as an event-driven temporal and logical sequence of task executions in which outputs are generated using available organisational resources (Becker et al. 2012; Davenport 1997; Hammer 2015). The discipline that deals with the design, analysis, execution, and improvement of such business processes is called business process management (BPM) (Hammer 2015; Reijers 2021; Weske 2019).

As a way of structuring BPM activities, the so-called BPM lifecycle (Figure 1) describes a continuous cycle of phases from process identification via discovery, analysis, redesign, implementation, to process monitoring (Dumas et al. 2018; Malinova et al. 2022). In the first phase (*process identification*), business processes relevant to a specific business problem are identified and delimited from the rest of the process architecture. During *process discovery*, the identified processes are documented in detail to generate an as-is process model. In the third phase (*process analysis*), issues in the identified and documented processes are inspected, evaluated, and prioritised. In the phase of *process redesign* (or *process improvement*), a to-be process model is created by using remedies to the identified issues or by capitalising on opportunities (Malinova et al. 2022). In *process implementation*, the remedies and changes are performed to change the as-is process to the to-be process. In the last phase (*process monitoring*), the new processes are analysed from an ex-post perspective to measure the performance of the implemented changes. The cycle is repeated over and over for all processes.

The continuous management and improvement of business processes is necessary in order to adapt processes to constantly changing environments and to keep up with customer needs and expectations in a fast-moving world (Gross et al. 2020; Kreuzer et al. 2020). The goal of every process improvement initiative is to significantly increase quality, productivity, customer satisfaction, and efficiency (Gross et al. 2019; Kreuzer et al. 2020; Vanwersch et al. 2015). Organisations must update their processes at an increasingly fast pace if they are to thrive in today's economy (Beverungen et al. 2021; Gross et al. 2020; Kreuzer et al. 2020).

[1] This section is partly comprised of content taken from the research articles in this thesis. To improve the readability of the text, I have omitted the standard labelling of these citations.

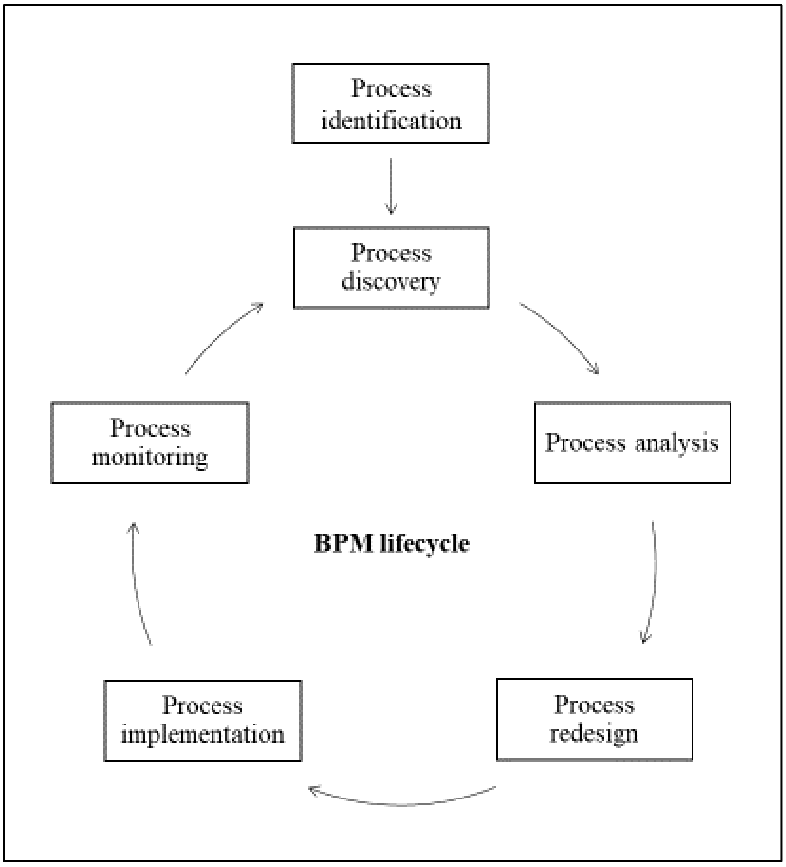



**Figure 1** - BPM lifecycle based on Dumas et al. (2018)

Academics and practitioners have developed a large amount of both BPM methods and IT capabilities for all phases of the BPM lifecycles in order to support organisations in improving their business processes (Kerpedzhiev et al. 2021; Rosemann and vom Brocke 2015; vom Brocke et al. 2021). For example, process identification is supported by reference models such as *ITIL* and *TOGAF* (Dumas et al. 2018) and BPM maturity frameworks (Pöppelbuß and Röglinger 2011; Rosemann and Bruin 2005; Tarhan et al. 2016). Process mining tools can be used in process discovery, process analysis, and process monitoring (van der Aalst et al. 2012; van der Aalst 2016). Redesign heuristics (Frank et al. 2020; Limam Mansar et al. 2009; Rosemann 2020) or structured approaches such as the process recombinator tool (Bernstein et al. 1999) or the business process design space (Gross et al. 2020) are used in process redesign. In the process implementation phase, executable process models can be created in BPM systems (Allweyer 2014; Karagiannis 1995).

In recent years, the increasing abundance of process data available in organisational information systems and simultaneous progress in computational performance have paved the way for a new class of so-called *data-driven* BPM methods and tools (Kerpedzhiev et al. 2021; Kratsch et al. 2017; Recker and Mendling 2016; van der Aalst 2013; van der Aalst et al. 2016). The most prominent development is process mining, a way of extracting knowledge on business processes from process data readily available in information systems (van der Aalst et al. 2012). Beyond, current and future BPM methods and tools will leverage the available data in manifold ways, e.g., for advanced process automation, adaptive process execution, or agile process improvement (Dumas et al. 2022; Kerpedzhiev et al. 2021). As an example, artificial intelligence (AI) techniques such as machine learning (ML) have already been effectively employed in several phases of the BPM lifecycle (Dumas et al. 2022) – e.g., in process structure discovery (Han et al. 2020), for process variant analysis (Taymouri et al. 2021), to identify causal dependencies (Bozorgi et al. 2020), and in predictive process monitoring (Heinrich et al. 2021; Kratsch et al. 2020).

One BPM lifecycle phase that has so far been treated with neglect when it comes to data-driven BPM support is process improvement. Process improvement is the necessary step to address existing issues in the as-is process, adapt the as-is process to constantly changing environments and customer needs and expectations, and take hold of emerging opportunities (Gross et al. 2020; Kreuzer et al. 2020). Therefore, this lifecycle phase holds the potential to significantly increase quality, productivity, customer satisfaction, and efficiency (Gross et al. 2019; Kreuzer et al. 2020; Vanwersch et al. 2015; Vanwersch et al. 2016). Unsurprisingly, the last decades have seen relevant research on process improvement methods (Limam Mansar and Reijers 2007; Reijers 2003). Yet, process improvement is often expensive, time-consuming, and labour-intensive (Al-Anqoudi et al. 2021; Gross et al. 2020; Huang et al. 2015; Limam Mansar et al. 2009), which is why there is a particular need to support process stakeholders in redesigning their processes.

However, traditional approaches to process improvement are often subjective and biased as they rely heavily on human intuition and creativity. They also often involve significant cognitive effort and can be too rigid (Gross et al. 2020; Limam Mansar et al. 2009; Röglinger et al. 2021). Rarely is the full solution space of process improvement explored. Computational approaches, such as the decision-making tool by Limam Mansar et al. (2009), commonly rely on previous projects or redesign heuristics with a low level of automation. Thus, such approaches often provide a narrower perspective on improvement as they redeploy knowledge from past improvement activities (Gross et al. 2020). Other approaches, such as evolutionary algorithms, require large amounts of structured data (Afflerbach et al. 2017; Fehrer et al. 2022). Therefore, the issue of data-driven business process improvement remains a key gap in research (Al-Anqoudi et al. 2021; Röglinger et al. 2021; van der Aalst 2013; Zuhaira and Ahmad 2020).

Regardless of the BPM lifecycle phase, all data-driven BPM approaches are reliant on data describing the context (Möhring et al. 2015), status quo (van der Aalst et al. 2012), or future (Poll et al. 2018) of one or more business processes. The most common type of process data and the starting point for most process mining techniques is an event log (van der Aalst et al. 2012). Event logs contain a sequential record of events executed in a business context at a given time "such that each event refers to an activity (i.e., a well-defined step in some process) and is related to a particular case (i.e., a process instance)." (van der Aalst et al. 2012; p. 174). Additional information about the resources executing the event or event-related data points can also be stored. Event logs and other types of process data are either (1) extracted and transformed from data stored in process-agnostic information systems used in businesses everywhere, such as enterprise resource planning (ERP), customer relationship management (CRM) systems or simple databases, or (2) readily available for extraction and use in so-called process-aware information systems (PAIS), such as BPM or workflow management systems. A PAIS is defined as "a software system that manages and executes operational processes involving people, applications, and/or information sources on the basis of process models" (Dumas et al. 2005; p. 7). Nowadays, the definition of PAIS is not necessarily restricted to traditional workflow management systems but includes all systems that entertain an explicit process notion and that are aware of the processes they support, such as large ERP systems like SAP and Oracle (van der Aalst 2013). However, most PAIS work with different data formats and some record events only implicitly, which impedes the standardised analysis of event logs (van der Aalst 2015). To solve this problem, the XES format has become the generally acknowledged de-facto standard for the interchange of event log data between different tools and application domains (Günther and Verbeek 2014; Verbeek et al. 2011).

In the end, no matter how and where process data was collected and stored, reliable results of data-driven BPM approaches are contingent on high-quality data (van der Aalst et al. 2012; van der Aalst 2016). In practice, process data is often far from the desired quality (Bose et al. 2013; Suriadi et al. 2017b). Therefore, such data should not be naively used as input to data-driven BPM approaches without ensuring that it is of adequate quality (van der Aalst 2016). Data scientists spend up to eighty percent of their work on identifying, assessing, and remedying data quality issues (Wynn and Sadiq 2019). A recent survey among researchers and practitioners revealed that incomplete or inconsistent process data is one of the biggest challenges to process mining projects and that data quality-related support is still lacking in most process mining solutions (Wynn et al. 2021). Thus, the interest in exploring the roots of data quality problems and the related assurance of high-quality process data is rising (van der Aalst et al. 2017b; Wynn and Sadiq 2019). For example, previous research established data quality as a multidimensional concept and defined data quality dimensions (see Lee et al. 2002; Pipino et al. 2002; Wand and Wang 1996; Wang and Strong 1996), which are widely adopted for evaluating data quality as they reflect the "fitness for use" of data by data consumers (Wynn and Sadiq 2019). Focusing on process data quality in particular, van der Aalst et al. (2012) provide maturity levels for the suitability of different data sources for process mining. Regarding real-life event logs, Bose et al. (2013) and Emamjome et al. (2019) identified typical data quality issues and Suriadi et al. (2017b) proposed eleven common event log imperfection patterns. However, to the best of our knowledge, research that addresses the assessment of process data quality and its impact on data analysis results remains scarce (Andrews et al. 2019).

# 2 Research Objectives

Based on the status quo in research and practice, as described in the motivation above, the aim of this thesis is twofold: First, research on process data quality has so far mainly focused on general, i.e., domain-agnostic, data quality dimensions on the one hand and specific process data quality issues on the other hand. In between, there is a need for detecting data quality issues in process data and quantifying their impact on data-driven BPM techniques that take process data as input. This thesis, therefore, aims to contribute to this research stream by designing artifacts for detecting and quantifying process data quality issues as a basis for further action, such as the repair of quality issues or the establishment of provenance in process data analysis.

Second, data-driven approaches in the BPM lifecycle phase of process improvement have so far not lived up to expectations, compared to approaches employed in other lifecycle phases. This thesis, therefore, aims to provide several concrete artifacts that demonstrate the potential of process data as enabler for process improvement. For one thing, this thesis aims at supporting the actual execution of process improvement initiatives by demonstrating and evaluating a model that uses generative machine learning to create new and improved process designs based on event logs of the as-is process. Additionally, this thesis aims at creating data-driven decision support systems in business processes. It provides two concrete artifacts: Customer data and multi-criteria decision analysis are employed to improve customer-facing last mile delivery processes. Finally, a modelling and simulation approach based on Petri nets is used to analyse and mitigate IT availability risks in production processes within smart factory networks.

The overarching goal of this thesis is not only to provide theoretical contributions to existing research streams, but also to provide BPM stakeholders in practice with guidance, practical assistance, and decision support for dealing with process data quality issues on the one hand and data-driven process improvement on the other hand. Therefore, all research questions addressed in this thesis are answered using design-oriented research to create, instantiate, and evaluate IT artifacts that (semi-)automate previously manual tasks in the fields of BPM in general and business process intelligence in particular.

# 3 Structure of this Thesis and Embedding of the Research Papers

This cumulative doctoral thesis consists of five research articles that investigate (1) advances in process data quality and (2) advances in business process improvement. The research articles in both pillars answer their respective research questions using various qualitative and quantitative methods as well as conceptual and theoretical lenses in the broader context of Design Science Research (DSR). As a result, the thesis presents design artifacts addressing the problem of identifying and quantifying process data quality as well as the challenge of using data in order to improve existing business processes. Covering both theoretical and practical perspectives on process data, this thesis is relevant for both researchers and practitioners.

**I. Introduction**

**II. Advancing Process Data Quality Management**

#1 **Quality-Informed Semi-Automated Event Log Generation for Process Mining**
Andrews R, van Dun C, Wynn MT, Kratsch W, Röglinger M, ter Hofstede AHM

#2 **Towards Interactive Event Log Forensics: Detecting and Quantifying Timestamp Imperfections**
Fischer DA, Goel K, Andrews R, van Dun C, Wynn MT, Röglinger M

**III. Advancing Data-Driven Process Improvement**

#3 **ProcessGAN: Creating Process Design Options through Generative Machine Learning**
van Dun C, Moder L, Kratsch W, Röglinger M

#4 **Customer-Centric Vehicle Routing: Incorporating Customer Centricity into Last Mile Delivery**
van Dun C, Fehrer T, Kratsch W, Röglinger M

#5 **IT Availability Risks in Smart Factory Networks – Analyzing the Effects of IT Threats on Production Processes Using Petri Nets**
Berger S, van Dun C, Häckel B

**IV. Summary and Future Research**

**V. Publication Bibliography**

**Figure 2** - Structure of this thesis including all research articles

As depicted in Figure 2, the research articles either address the topic of process data quality or the topic of data-driven business process improvement. In Section II (including research articles #1 and #2), the thesis provides an overview of approaches for detecting, identifying, and quantifying data quality issues in process data. For this purpose, the thesis distinguishes between process data from process-agnostic source systems and process data already available in process-specific formats or from process-aware information systems. In particular, research article #1 presents an approach for (semi-) automated and quality-informed event log extraction from process-agnostic relational databases. It applies metrics for data quality dimensions that are relevant to process mining in order to quantify the data quality of the source data in selected database tables and simultaneously allows users to extract event logs in XES format from the database tables. Research article #2 presents an approach for detecting and quantifying timestamp data quality issues in events logs already present in XES format. The approach applies metrics for identifying timestamp imperfection patterns as proposed by Suriadi et al. (2017b) and allows users to interactively filter, repair, and annotate the event log.

In Section III (including research articles #3, #4, and #5), the thesis demonstrates several advances in data-driven business process improvement. Research article #3 provides a model based on generative adversarial networks, i.e., a generative machine learning technique based on opposing neural networks, to create new process designs. Specifically, it uses event logs and annotated information on process variants and (positive and negative) process deviance as input and outputs a new process model which significantly differs from the as-is process and provides suggestions for process improvement to the user. Research article #4 focuses on last mile delivery processes, e.g., for meals or parcels, and uses multi-criteria decision analysis to extend traditional vehicle routing problems with a customer-centric perspective. In particular, the customer-centric vehicle routing uses process and historical customer data and the concept of customer lifetime values to predict customer satisfaction and, thus, optimise delivery routes. Finally, research article #5 presents a modelling approach for IT availability risks in smart factories based on Petri nets. The modelling approach uses modular components of information systems and production machines to model, simulate, and analyse propagation effects and cascading failures in production processes within smart factory networks.

Section IV concludes this thesis by providing a summary of the key insights, limitations of this work, and directions for future research. Section V lists the publication bibliography. The Appendix in Section VI compiles additional information on all research articles (VI.1), my individual contributions (VI.2), and the research articles themselves (VI.3 to VI.7).

# II. Advancing Process Data Quality Management[2]

Process mining is a set of data-driven BPM techniques used to analyse business processes (van der Aalst 2016). With process mining, organisations can leverage process data to gain insights into business process performance, conformance of processes to existing process models, and improvement opportunities (Suriadi et al. 2017b). In doing so, process mining supports crucial BPM lifecycle phases (van der Aalst et al. 2012; van der Aalst 2016), facilitates evidence-based process improvement (Partington et al. 2015), and strategic decision making (Mans et al. 2013).

While process mining has recently seen an extraordinary rate of adoption in practice across a wide range of industries (Reinkemeyer 2020), research has kept up and has generated substantial output over the last decade – resulting in advances in both technical and organisational areas (Martin et al. 2021; R'bigui and Cho 2017; Thiede et al. 2018). In the process, research has mainly focused on the application of process mining in specific use cases (e.g., Andrews et al. 2018a; Andrews et al. 2018b; Emamjome et al. 2019) and the development and refinement of process mining algorithms (R'bigui and Cho 2017). Several new research streams have identified further application areas such as predictive process monitoring (Kratsch et al. 2020; Teinemaa et al. 2019) or robotic process mining (Leno et al. 2021). Having said this, most recent process mining approaches are predicated on the existence of high-quality process data in the form of event logs, without describing how such logs can be generated and how the quality of the log can be assured (e.g., Evermann et al. 2017; Kratsch et al. 2017; Martin et al. 2017; Suriadi et al. 2017a; Wynn et al. 2017).

Reliable process mining results are, however, contingent on high-quality process data (Andrews et al. 2019; van der Aalst et al. 2012; van der Aalst 2016). Naively using process data for process mining without ensuring that it is of adequate quality may therefore lead to inaccurate results or questions left unanswered (Bose et al. 2013; van der Aalst 2016). Many existing studies show that process data in the form of event logs is often far from the desired quality as it tends to be noisy, incomplete, and imprecise (Andrews et al. 2018a; Andrews et al. 2018b; Bose et al. 2013; Dixit et al. 2018; Suriadi et al. 2017b). As a result, process analysts might spend up to eighty percent of their work on manually detecting and remedying data quality issues (Nooijen et al. 2012; Wynn and Sadiq 2019).

[2] This section is partly comprised of content taken from the research articles in this thesis. To improve the readability of the text, I have omitted the standard labelling of these citations.

To alleviate this problem, the BPM community has begun to explore the roots of data quality problems in process data and the related assurance of high-quality process data (Dixit et al. 2018; Suriadi et al. 2017b; van der Aalst et al. 2017a; Wynn and Sadiq 2019). As a first step, it is essential to have a means for detecting and quantifying process data quality (Wynn and Sadiq 2019). While some basic solutions have been devised (e.g., Fox et al. 2018; Kurniati et al. 2018), a generalised approach for the detection and quantification of process data quality does not yet exist. To address this need, Section II presents two IT-enabled approaches that complement each other and support process analysts in detecting, identifying, and quantifying process data quality issues in different data sources.

As a shared basis, both approaches build on existing research on the concept of data quality. Juran and Godfrey (1999) define quality as either “fitness for use” or “fitness for purpose”, and Wynn and Sadiq (2019) build on this definition to define data quality as “fitness for use” of said data by data consumers. Process data should, therefore, be able to answer the user’s (i.e., process analyst’s) questions. The assessment of this “fitness for use” is not a trivial task: In most publications, data quality is described as a multidimensional concept (Wand and Wang 1996). Thus, assessing data quality always involves examining and weighting not one, but multiple dimensions. Previous research has established several sets of data quality dimensions (e.g., Batini and Scannapieco 2016; Lee et al. 2002; Pipino et al. 2002; Redman and Blanton 1997; Scannapieco et al. 2005; Stvilia et al. 2007; Wand and Wang 1996; Wang and Strong 1996). However, these dimensions are not specific to process data, and most publications do not provide concrete, domain-specific metrics for measuring the data quality dimensions. Metrics represent a formal way of measuring and quantifying data quality dimensions (Görz and Kaiser 2012; Heinrich et al. 2018; Heinrich and Klier 2015; Pipino et al. 2002) that can be computed with a certain degree of automation (Even and Shankaranarayanan 2007; Kaiser et al. 2007). The two approaches for detecting and quantifying data quality issues differ in the type of data source being examined: Section II.1 (research article #1) presents an approach for detecting and quantifying process data quality in source systems, while Section II.2 (research article #2) presents an approach for analysing data already available in process-specific formats or from process-aware information systems.

# 1 Process Data Quality in Process-Agnostic Source Data

There is only a handful of approaches available to support process analysts in extracting event logs in a systematic manner, including ProM plugins (Günther and van der Aalst 2006; Verbeek et al. 2011), object-centric (Li et al. 2018), redo log-based (van der Aalst 2015), or ontology-based approaches (Calvanese et al. 2016; Calvanese et al. 2017). Each has limitations, and none address data quality issues of the source data.

To fill this gap, research article #1 presents RDB2Log, a semi-automated, quality-informed approach to event log extraction from relational databases. RDB2Log takes, as input, a relational data set and generates an assessment of its data quality based on data quality dimensions. Using this assessment and the database key constraints, the data's suitability for process mining can be evaluated. RDB2Log supports the mapping of data columns to event log attributes and subsequential extraction of the event log.

RDB2Log was developed following the precepts of the DSR methodology (Hevner et al. 2004; Peffers et al. 2007). DSR is an accepted research paradigm in IS research and "involves a rigorous process to design artifacts to solve observed problems, to make research contributions, to evaluate the designs, and to communicate the results to appropriate audiences" (Hevner et al. 2004; Peffers et al. 2007). Such artifacts can be constructs, models, methods, and instantiations (Hevner et al. 2004; March and Smith 1995). Therefore, RDB2Log is a valid design artifact following the definition by March and Smith (1995).

The DSR methodology by Peffers et al. (2007) proposes an iterative research process with feedback loops: (1) problem identification, (2) definition of solution objectives, (3) design and development, (4) demonstration, (5) evaluation, and (6) communication. In complying with the design-evaluate-construct-evaluate pattern advocated in Sonnenberg and vom Brocke (2012), the construction of RDB2Log did not traverse DSR phases strictly iteratively, but switched between the design and develop as well as the demonstration and evaluation phases.

The schematic architecture of RDB2Log is illustrated in Figure 3. The approach functions in three distinct steps: (1) database relationship assessment, (2) data quality assessment, and (3) attribute assignment.

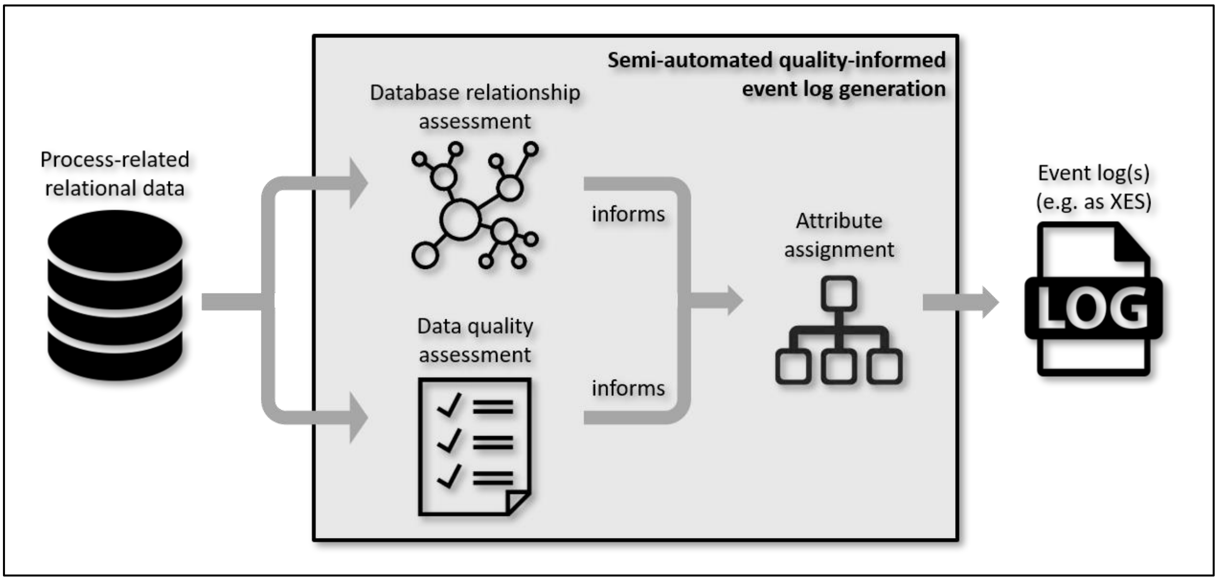



**Figure 3** - RDB2Log - quality-informed event log generation

In step (1), the structure of the source data is analysed to verify the suitability for event log extraction. RDB2Log caters to process data generated by process-agnostic information systems that is stored in relational form and that needs to be transformed into an event log for process mining. For this purpose, the source data must meet certain criteria such as existing Primary Keys and Foreign Keys as per the concept of relational databases. Additionally, the data must contain timestamp attributes which form the basis of the event ordering in the output event log.

In step (2), the data quality of the whole database and of each of the attributes (i.e., the columns of the relational data source) is assessed. In sum, RDB2Log incorporates 12 data quality dimensions that have been deemed quantifiable and relevant for process mining. These are currency, accuracy, consistency, completeness, privacy, objectivity, sufficiency, conciseness, integrity, precision, informativeness, and uniqueness. Each of the 12 quality dimensions is informed by one or more metrics. A detailed description of all dimensions and metrics can be found in the research article. As a result, each attribute of the source database is assessed against several quality dimensions. Depending on the attribute's data type (e.g., integer, string, or timestamp) and on user preferences, the weights of the dimensions and respective metrics can be adapted. The research article provides a basis for interpreting the respective dimension values.

In step (3), users assign event log roles to selected database attributes. RDB2Log provides guidance on suitable database attributes in the form of acceptance criteria based on quality levels and on the structure of the database. As a way of mapping database attributes to event log attributes, the research article proposes the notion of event constructors, i.e., for each selected event representation in the source data, a constructor defining the necessary attributes (case identifier, activity label, timestamp, and optional case/event data) is created. The final set of event log rows is then extracted.

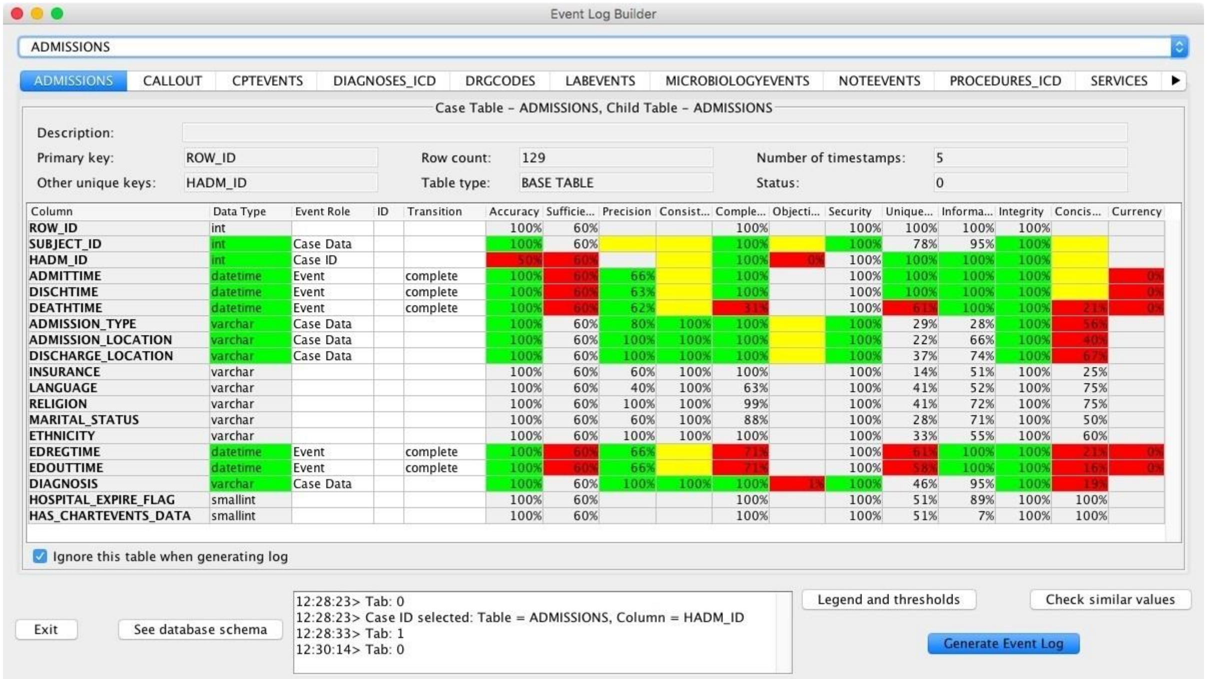



**Figure 4** - Example screenshot of the instantiated software prototype

RDB2Log has been evaluated based on the DSR evaluation framework proposed by Sonnenberg and vom Brocke (2012). This DSR evaluation framework comprises four activities (EVAL1 to EVAL4). EVAL1 aims to justify the research problem. It also requires deriving design objectives from existing knowledge. EVAL2 evaluates the artifact's design specification by discussing its features against competing artifacts as well as challenging its understandability and real-world fidelity with process mining experts from industry and academia. The results indicate that no other approach addresses the design objectives as comprehensively as RDB2Log and that RDB2Log is valid and fills a gap in literature. EVAL3 strives for validated instantiations. Thus, RDB2Log has been implemented as a software prototype (Figure 4) and applied to two data sets in a laboratory setting. In accordance with Sonnenberg and vom Brocke (2012), the approach is validated regarding feasibility and suitability. The evaluation can give a rough indication of RDB2Log’s ease-of-use and robustness. EVAL4 requires validating the instantiation's applicability in naturalistic settings. Thus, the prototype is applied to real-world data of a medium-sized manufacturing company and the results are evaluated and discussed in collaboration with the company.

In sum, research article #1 proposes RDB2Log, an approach towards semi-automatically generating quality-informed event logs from relational data. RDB2Log requires as input a relational data set and supports the user in selecting event log attributes from the available data columns by providing information on data quality and data constraints. As output, an event log is provided.

# 2 Process Data Quality in Process-Aware Event Logs

In times of maturing BPM, growing process orientation, and adoption of process mining in organisations, process data is available in PAIS (such as elaborate ERP or CRM systems) more often than not. However, research that addresses the (semi-automated) quality assessment of process data – even when it is already available in an event log format – remains scarce (Andrews et al. 2019; Suriadi et al. 2017b).

Research article #2 intends to bridge this gap in research specifically for timestamp-related data quality issues. Timestamps are at the core of many process mining use cases (Dixit et al. 2018; Gschwandtner et al. 2012; van der Aalst et al. 2012). In most cases, precise timestamps are essential for reproducing the correct ordering of activities in order to obtain accurate process models (Dixit et al. 2018; Gschwandtner et al. 2012). In contrast, inaccurate or granular timestamps often lead to convoluted process models that may result in erroneous analyses (Dixit et al. 2018). The article builds on earlier work on automating the detection and quantification of timestamp imperfections (Fischer et al. 2020) and focuses on the following research question: *How can timestamp-related data quality issues in event logs be detected and quantified?* As such, it is a first step in quantifying event log data quality.

To address the research question, research article #2 also adopts the DSR paradigm (Gregor and Hevner 2013; Peffers et al. 2007) to build and instantiate a user-guided and semi-automated approach to assess the quality of timestamps in event logs across two axes: four levels of abstraction (event, activity, trace, log) and four quality dimensions (accuracy, completeness, consistency, uniqueness). As an operationalisation of the four quality dimensions, 15 timestamp-related quality metrics are defined and computed using the software prototype implemented in the academic process mining framework ProM. Additionally, the approach provides starting points for user-interactive and domain-specific data quality detection and quantification. The approach can detect common timestamp-related issues and measure the quality of timestamp information in event logs.

The design specification conceived in the design and development phase of the DSR project consists of a timestamp quality framework, timestamp-related quality metrics, and interactivity functions. This design specification is backed by a comprehensive literature search which was conducted to identify data quality dimensions and issues related to timestamps. Four particularly relevant data quality dimensions (accuracy, completeness, consistency, and uniqueness) and five data quality issues at the event, activity, trace, and log level were identified. In regard to these data quality issues, the research article develops 15 data quality metrics in total. The metrics are the quantifiable measures for data quality issues spanning the four data quality dimensions and four levels of log abstraction. Thus, the metrics are positioned according to two axes: four data quality dimensions and four levels of log abstraction resulting in a timestamp quality assessment framework (Table 1). The metrics are derived from either existing detection approaches, modification to existing detection approaches, or newly designed detection approaches based on insights from the literature. The metrics are designed to fit event logs in standardised formats like XES. For the calculation of a value for a whole dimension or abstraction level, the weighted average of the allocated metrics is used. The respective weights can be adapted depending on user preferences and domain-specific constraints.

**Table 1** - Timestamp quality assessment framework

<table>
<tr><th rowspan="2"></th><th colspan="4">Timestamp Quality</th></tr>
<tr><th>$QD_1$: Accuracy</th><th>$QD_2$: Completeness</th><th>$QD_3$: Consistency</th><th>$QD_4$: Uniqueness</th></tr>
<tr><td rowspan="2">Log Level</td><td rowspan="2"></td><td rowspan="2">$M_5$: Missing Trace[c]</td><td>M9: Mixed Granu-larity of the Log[c]</td><td rowspan="2">$M_{13}$: Duplicates within Log[c]</td></tr>
<tr><td>M10: Format[b]</td></tr>
<tr><td rowspan="2">Trace Level</td><td>$M_1$: Infrequent Activity Ordering[a]</td><td rowspan="2">$M_6$: Missing Activity[b]</td><td rowspan="2">$M_{11}$: Mixed Granu-larity of Traces[a]</td><td rowspan="2">$M_{14}$: Duplicates within Trace[b]</td></tr>
<tr><td>$M_2$: Overlapping Activities per Resource[a]</td></tr>
<tr><td>Activity Level</td><td></td><td>$M_7$: Missing Event[c]</td><td>$M_{12}$: Mixed Granu-larity of Activities[c]</td><td>$M_{15}$: Duplicates within Activity[c]</td></tr>
<tr><td rowspan="2">Event Level</td><td>$M_3$: Future Entry[c]</td><td rowspan="2">$M_8$: Missing Timestamp[c]</td><td rowspan="2"></td><td rowspan="2"></td></tr>
<tr><td>M4: Granularity[c]</td></tr>
</table>

□: metric can be allocated; □: no metric can be allocated

a: pre-existing detection approach used; b: modification of pre-existing detection approach; c: new development

The approach for detecting and quantifying timestamp imperfections has been implemented as an open-source software prototype as part of the ProM framework (Figure 5). In order to complement automated data quality measurements with domain and use case knowledge from the user, the instantiation provides multiple components for the integration of user input: The metric weights are highly configurable, so that users can tailor the quantification towards specific domains and use cases. Furthermore, the software provides a detailed view of each identified data quality issue and allows users to select or deselect data quality issues in order to avoid false positives. And lastly, the user can add the quality information to the metadata of the event log under consideration as a way of storing it for further use.

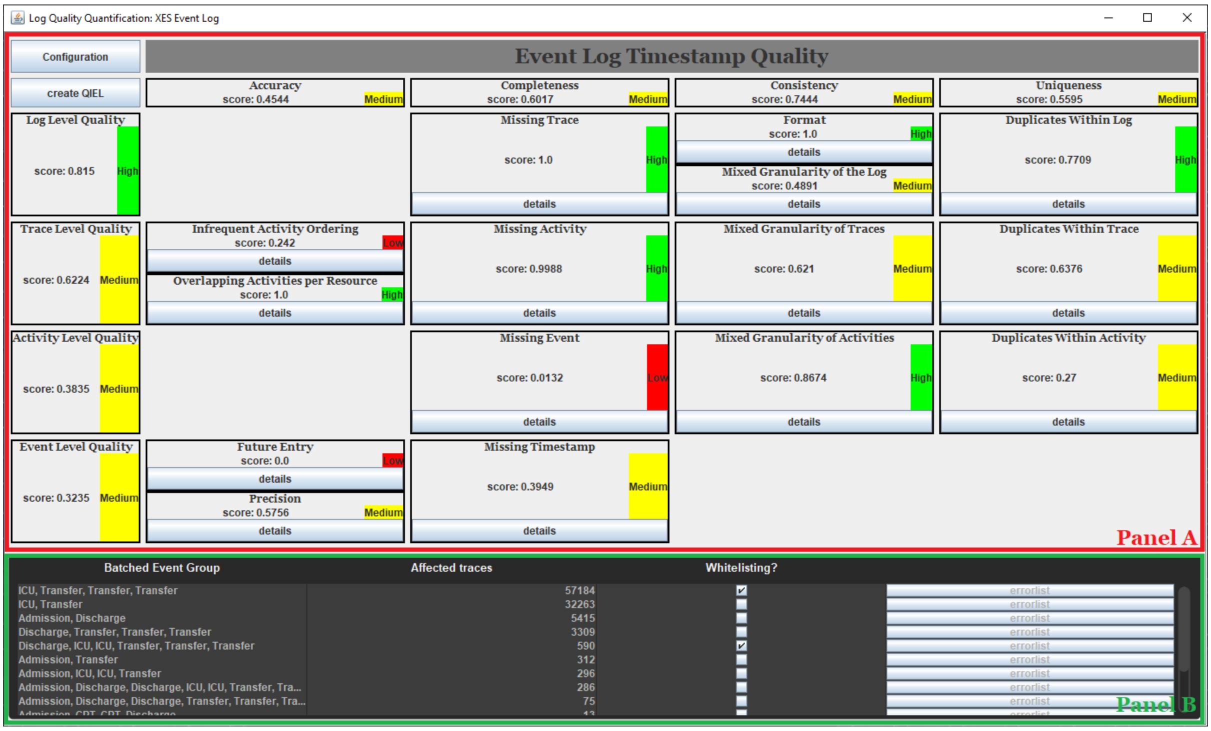



**Figure 5** – Example screenshot of the quality quantification (Panel A) and a detailed view of quality issues (Panel B)

Following the DSR evaluation framework by Sonnenberg and vom Brocke (2012), the approach is evaluated in four phases: In EVAL1, the research gap is justified and design objectives are derived. EVAL2 strives for validated design specifications by discussing the approach's features against competing artifacts. By also comparing its features with the derived design objectives, the article underpins the approach's significant value-add to existing literature. EVAL2 concludes that the approach sufficiently caters to all design objectives and thus adds to the prescriptive body of knowledge related to data quality in process mining. In EVAL3, the implemented instantiation of the approach is used in experiments with experts from research and practice using real-world event logs in order to refine the timestamp quality assessment framework and the contained metrics. As for EVAL4, the author team conducted a survey study with process mining experts from academia and industry to validate the perceived ease-of-use and usefulness of the approach and its implementation. These evaluation criteria are based on the well-known Technology Acceptance Model (TAM; Davis 1989).

In a nutshell, the presented approach supports process stakeholders in determining the suitability of an event log for process mining. It also assists data scientists in interactively identifying and assessing data quality issues in event logs. The implementation as part of the ProM framework allows for interoperability with other tools and methods in the process mining toolchain. Finally, the approach paves the way for future research on detecting and quantifying quality issues of further event log components (e.g., activity labels).

In sum, Section II presents artifacts that support BPM stakeholders (such as process owners and process analysts) in detecting data quality issues in process data from various data sources and stored in various formats. The artifacts seek to address relevant problem classes with useful solutions. Specifically, research article #1 provides an approach for managing process data quality and extracting process data from process-agnostic source systems while research article #2 provides an approach for managing data quality of process data already stored as an event log. Both presented artifacts address different steps in the data preparation phase of a data-driven BPM project and, thus, complement each other well.

# III. Advancing Data-Driven Process Improvement[3]

In the BPM lifecycle phase of *process improvement* (or *process redesign*), a to-be process model is created by using remedies to existing issues in the as-is process or by capitalising on opportunities (Baumbach et al. 2020; Malinova et al. 2022). Process improvement has the potential to significantly increase quality, productivity, customer satisfaction, and efficiency (Gross et al. 2019; Kreuzer et al. 2020; Vanwersch et al. 2015; Vanwersch et al. 2016) and, therefore, entails significant economic value (Vanwersch et al. 2015; Zellner 2011).

Generally, *process improvement* describes the act of systematically changing existing processes or developing new processes (Kettinger et al. 1997). Related notions such as *process redesign*, *process reengineering*, or *process optimisation* are used synonymously to a degree (Grisold et al. 2021; Malinova et al. 2022). However, each carries their own connotation: For example, reengineering is often connected to radical changes in a process (Hammer and Champy 1993), while redesign is often used in connection with incremental (i.e., less radical) approaches, such as changing existing processes to decrease time and cost, or to increase quality and flexibility (Mansar and Reijers 2005). This thesis follows Malinova et al. (2022) and uses the term *process improvement* in a broad sense, including all acts of changing existing or developing new processes.

Since the 1990's, a large number of methods for process improvements have been developed (Gross et al. 2020; Limam Mansar et al. 2009; Malinova et al. 2022; vom Brocke et al. 2021). They have been highlighted as a critical success factor (Rosemann and vom Brocke 2015) and can be classified regarding their ambition, nature, and perspective (Dumas et al. 2018; Recker 2012). The ambition of a process improvement method lies on a spectrum between incremental and radical (i.e., between redesign and reengineering), the nature of a method is creative or analytical, and the perspective is inward-looking or outward-looking. Based on a definition by Gross et al. (2020), process improvement methods include problem-based approaches (e.g., Bortolotti and Romano 2012; Kwak and Anbari 2006), imitation-based approaches (e.g., König et al. 2019; Setiawan and Sadiq 2013), pattern-based approaches (e.g., Frank et al. 2020; Mansar and Reijers 2005; Rosemann 2020), interaction-based approaches (e.g., Rosemann 2018), customer-based approaches (e.g., Bettencourt et al. 2013), and several others.

[3] This section is partly comprised of content taken from the research articles in this thesis. To improve the readability of the text, I have omitted the standard labelling of these citations.

In recent years, the increasing availability of process data in organisational information systems has opened up opportunities for data-driven BPM methods and tools (Kerpedzhiev et al. 2021; Kratsch et al. 2017; Netjes et al. 2006; Recker and Mendling 2016; van der Aalst 2013; van der Aalst et al. 2016). This fundamental paradigm shift leads to two complementary data-driven opportunities with respect to process improvement: Process data can now be used either as an *enabler of improvement initiatives* themselves or as an *enabler of data-driven decision support* in improved business processes. More precisely, process data can be used (1) as a tool in the phase of process improvement (e.g., as a source of improvement ideas), or (2) as the basis for improved process (e.g., through data-driven automation or predictive process monitoring).

Regarding (1) *process data as enabler of process improvement initiatives*, researchers have begun to identify opportunities to leverage data for process improvement. Traditional process improvement methods are often associated with mostly manual, creative work, and high investments in terms of cost and time (Al-Anqoudi et al. 2021; Gross et al. 2020; Huang et al. 2015; Limam Mansar et al. 2009). Thus, Röglinger et al. (2021) distinguish between manual and automated, data-driven process improvement methods and position the latter as a key challenge for future BPM research. In the meantime, some early concepts for data-driven process improvement have been developed: For example, Afflerbach et al. (2017) deploy evolutionary algorithms based on structured data on process activities and properties to generate improved processes. Truong and Le (2016) rely on data mining techniques, such as rule-based classification, to determine opportunities for task elimination or resequencing. Niedermann et al. (2010) apply best practice patterns via process matching based on operational and event data, utilising several data mining methods. Zemni et al. (2016) serve as representatives of the emerging compositional methods (Reijers 2021) as they provide a systematic merge mechanism for process fragments based on a path matrix. Lastly, Borgianni et al. (2015) present an algorithmic decision support model to indicate value bottlenecks applying process value analysis and Monte Carlo simulation. Nevertheless, these existing data-driven approaches still require much cognitive effort and manual input from the user or large amounts of structured and pre-processed data, thus providing either only low levels of automation or a narrow view of process improvement alternatives.

In addition, advances in AI and, more specifically, in ML have launched an area of research on AI-based innovation. Automated innovation approaches based on generative machine learning and computational creativity have become increasingly popular (Heinrich et al. 2021; Kratsch et al. 2020; Taymouri et al. 2020). Machine learning has already been effectively applied in several phases of the BPM lifecycle (e.g., Dumas et al. 2022; Han et al. 2020; Heinrich et al. 2021; Kratsch et al. 2020; Taymouri et al. 2020; Taymouri et al. 2021). Thus, it is only reasonable to assume that such developments in AI also stand to make process improvement less dependent on human creativity (Röglinger et al. 2021).

Thus, creating (semi-)automated, ML-based, and data-driven approaches for process improvement is a worthy goal of BPM research. For this purpose, Section III.1 (research article #3) presents such a process improvement approach based on generative adversarial networks, i.e., a generative machine learning technique with opposing neural networks, to create new process designs.

Regarding (2) *process data as enabler of data-driven decision support*, researchers and practitioners alike have already leveraged data for designing effective and efficient business processes. Heinrich et al. (2021) use process data to dynamically predict next events of running processes, while Kratsch et al. (2020) predict outcomes of such running process instances. Another example is robotic process mining which allows users to dynamically automate repetitive routines with robotic process automation (Leno et al. 2021). The goal of this thesis is to advance this stream of research by adding to the host of concrete process improvements based on data-driven decision support. Thus, Section III.2 (research articles #4 and #5) provides examples for data-driven process improvements, specifically in last mile delivery processes and smart factory production processes.

## 1 ML-Based Business Process Improvement

Röglinger et al. (2021) propose further research on automated process improvement systems. To follow this recommendation and address the lack of computational support for process improvement, research article #3 poses the research question: *How can business processes be improved through generative machine learning?*

To answer this question, the article builds and evaluates ProcessGAN, a novel approach for business process improvement based on generative adversarial networks (GANs) that supports the generation of new process designs using process data. It consists of a GAN generating new process models based on event log data, thereby conducting pioneering work at the intersection of process improvement and generative machine learning. GANs have so far been used in other domains such as art and design (e.g., Elgammal et al. 2017; Sbai et al. 2019). The article's goal is to evaluate the suitability of GANs for automated process improvement.

For this purpose, the article follows the DSR paradigm (Hevner et al. 2004; March and Smith

1995) and adheres to the six-step DSR methodology proposed by Peffers et al. (2007) which structures the development and prototypical implementation of ProcessGAN by iteratively combining building and evaluation activities, thereby focusing on ProcessGAN's functional core. As DSR aims to develop IT artefacts that solve organisational problems, ProcessGAN is a valid design artifact following the definition by March and Smith (1995).

As shown in Figure 6, ProcessGAN comprises three stages: (1) input and data pre-processing, (2) the GAN model for automated process improvement, and (3) output and data postprocessing. First, event log traces are transformed into encoded sequences. In training, the GAN learns how to generate new process traces. The generated sequences are then converted into a process model for evaluation and further use within the BPM toolchain.

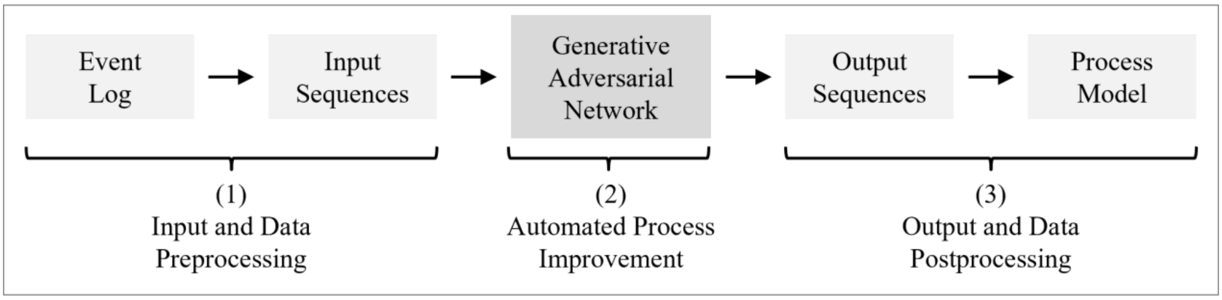



**Figure 6** - Schematic architecture of ProcessGAN

In (1), the input to ProcessGAN is prepared. ProcessGAN uses process data in the form of event logs as input. Event logs contain information about activities, timestamps, and attributes from process instances (Dumas and Mendling 2019; van der Aalst et al. 2012) and are sensible as input since event logs are readily available (van der Aalst 2013) and GANs require a large amount of training data (Chollet 2018). The data is encoded and standardised to satisfy GAN requirements. Furthermore, information on positively and negatively deviant process instances is added to the encoded log for the purpose of identifying desirable and undesirable process instances (Delias 2017; König et al. 2019).

In (2), the GAN is initialised and trained. In setting up the GAN, the article follows the architecture provided in Taymouri et al. (2020). GANs consist of two long short-term memory (LSTM) neural networks, called discriminator and generator. LSTM networks are adaptions of conventional recurrent neural networks (RNN) that are well suited to sequential data such as process data (LeCun et al. 2015; Sengupta et al. 2020). The discriminator and generator are trained based on several different and opposing incentives, thereby trying to distinguish between real and artificially generated as well as between undesirable and positively deviant process instances. An overview of this training procedure is given in Figure 7.

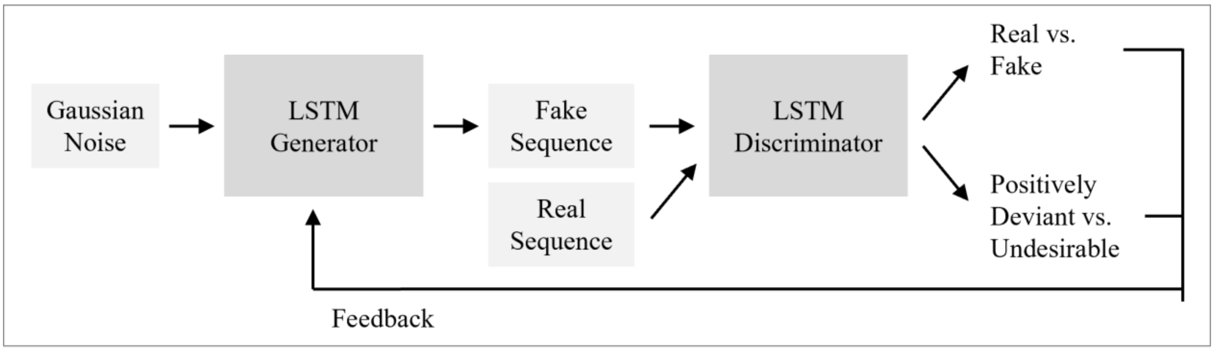



**Figure 7** - ProcessGAN training procedure

In (3), once the training is completed, the generator can create new batches of artificial process instances as output that are decoded and then transferred into a process model. The training can lead to results of varying quality, and the user is, therefore, required to provide input at several stages of ProcessGAN in order to mitigate quality issues: In the first step, obviously erroneous models or sub-sequences are identified and rejected. In the second step, the feasibility of the generated models is assessed by the user. Lastly, the third step involves an economic assessment based on estimated process performance criteria weighted by the user. Overall, this analysis procedure does not lead to a strict recommendation but enables informed decision-making based on suggestions for new (sub-)process designs.

To evaluate ProcessGAN, the article conducts four evaluation activities (EVAL1-EVAL4) as proposed by Sonnenberg and vom Brocke (2012). Specifically, the research gap is justified, and design objectives are derived in Sections 1 and 2, thereby addressing EVAL1. For EVAL2, a feature comparison and competing artifact analysis is performed. While ProcessGAN does not yet enable fully automated process improvement, it nevertheless offers the possibility of semi-automatically creating new process designs, thereby sufficiently addressing the design objectives and outperforming most other existing approaches. In EVAL3, a prototypical implementation of the artifact as a software prototype is presented and examined using four publicly-available datasets. Preliminary results show that ProcessGAN can create process models that deviate from undesirable variants but also resemble the original process, demonstrating the approach's general feasibility and applicability for automated process improvement. The artifact's applicability and usefulness in practice is established in EVAL4 by applying the prototype to real-world data from a multinational company and discussing the results with stakeholders from the company.

In sum, research article #3 proposes ProcessGAN, an ML-enabled and data-driven approach to process improvement based on event logs. The article's contribution is twofold: By investigating the potential of generative machine learning for process improvement and by designing ProcessGAN, it adds to prescriptive knowledge on process improvement. Second, the article uses the observations taken from the design and implementation of ProcessGAN to infer implications for the class of systems comprising ProcessGAN which it calls automated process improvement systems (APIS). APISs support and automate process improvement and represent an emerging subcategory within PAISs that has not yet been explored in a structured way. As one type of APISs, ProcessGAN not only serves as a proof-of-concept but also includes design knowledge that can be applied to the whole system class. Hence, the implications represent learnings from the conception of ProcessGAN that should be transferrable to other types of APISs and are meant as a stimulant for further research.

## 2 Data-Driven Decision Support within Business Processes

Besides putting forward an approach for using process data as an enabler of business process improvement as described in Section III.1, this thesis aims to provide guidance on how to enable real-time data-driven decision making in business processes. Thus, this section presents two examples of data-driven decision support in two different application domains.

First, research article #4 investigates last mile delivery businesses. In essence, last mile delivery is a customer-facing business process (Frank et al. 2020; Winkelhaus and Grosse 2020) with the aim of delivering products from one or more depots to customers in multiple locations with a limited number of couriers. The backbone for executing such processes is the vehicle routing problem (VRP) which is employed to minimise delivery efforts in a complex environment (Ahmadi-Javid et al. 2018; Xiang et al. 2008). While this efficiency-driven view of last mile delivery processes minimises costs, the long-term customer-centric perspective is often neglected. Customer satisfaction in last mile delivery processes is largely determined by delivery times (Barkaoui et al. 2015; Sivaramkumar et al. 2017; Zhang et al. 2012). Neglecting this factor may lead to diminishing customer satisfaction in demographics with unfavourable characteristics such as comparatively long distance from the delivery depot or high traffic density on the route. Eventually, decreased customer satisfaction may cause disaffection and churn of valuable customers (Galbraith 2005; Vakulenko et al. 2019).

One way for delivery businesses to avoid such unintended consequences is to add a customer-centric perspective to the delivery process. Recently, the BPM community has attempted to account for customer centricity in the design and improvement of business processes (Frank et al. 2020; Kreuzer et al. 2020; Trkman et al. 2015). Related literature has also gained insights into the trade-off between customer centricity and efficiency in business processes (Afflerbach and Frank 2016; Frank et al. 2020). Thus, it is beneficial for businesses to establish a balance between customer-centric and efficiency-driven perspectives in last mile delivery processes (van den Hemel and Rademakers 2016). Building on this, research article #4 poses the following research question: *How can last mile delivery be enhanced by incorporating long-term customer centricity?*

To answer this question and building on earlier work on customer-centric last mile delivery by van Dun et al. (2020), the article adopts the DSR paradigm proposed by Gregor and Hevner (2013). In the design and development phase, it employs normative analytical modelling (Meredith et al. 1989) and multi-criteria decision analysis (Cohon 2004; Marttunen et al. 2017) to build and evaluate a decision model called *Customer-Centric Vehicle Routing* (C2VR). The C2VR is based on prescriptive knowledge from logistics and operations research as well as descriptive knowledge on customer centricity in BPM.

Incorporating C2VR into typical dynamic vehicle routing problems enables last mile delivery businesses to strike a balance between short-term efficiency (i.e., minimising costs) and long-term customer centricity. Like any other VRP variant, the C2VR offers a solution to the underlying routing problem. However, it extends traditional VRPs by adding a customer-centric perspective to the optimisation problem, the benefit being that it enhances customer satisfaction and increases customer lifetime values by shortening delivery times for disadvantaged customers without neglecting routing efficiency.

As a data-driven process improvement, the C2VR uses the definition of the underlying dynamic VRP and data on the customer base as input. In dynamic VRPs, the orders for delivery are being revealed at the moment of placement and can, therefore, not be predicted accurately beforehand. Given this input, the C2VR computes the solution to the underlying VRP with the highest value contribution to the last mile delivery business. This value contribution is assessed using an objective function that contrasts a solution's loss in efficiency compared to the optimal solution with the respective impact on the involved customers' customer lifetime values, also compared to the optimal solution. To valuate this impact on customer lifetime values, we follow Gupta and Lehmann (2003) who define the customer lifetime value as the discounted expected cashflow generated from a customer. The customer lifetime value of a last mile delivery customer is, therefore, influenced by their future order probability which, in turn, is predicated on customer satisfaction. Every solution to the given vehicle routing problem influences customer satisfaction based on the respective delivery time of the customer's order (Barkaoui et al. 2015; Vakulenko et al. 2018; Vakulenko et al. 2019). This influence is modelled using time windows (Cheng et al. 1995; Zhang et al. 2012). An overview of the C2VR's conceptual architecture with inputs and outputs is shown in Figure 8.

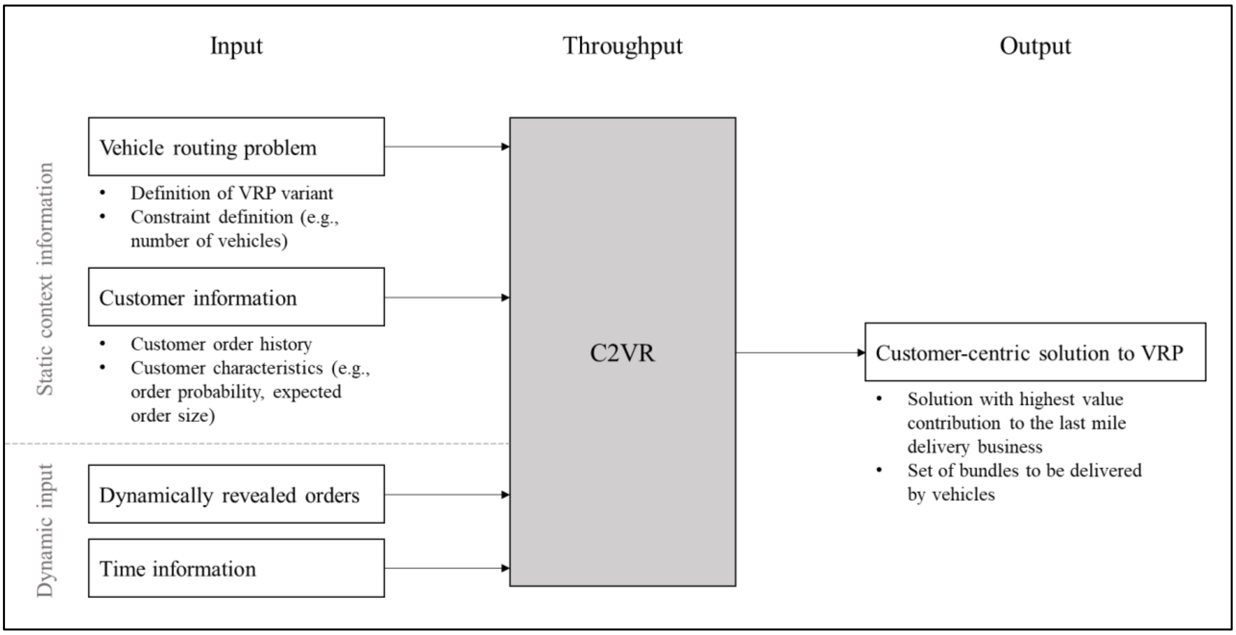



**Figure 8** - Conceptual architecture for C2VR

The decision model is evaluated following the DSR evaluation framework proposed by Sonnenberg and vom Brocke (2012). EVAL1 is used to establish that the research problem in question is real and in need of remedy. The article motivates the problem and derives design objectives. In EVAL2, the article compares the design specification of the C2VR decision model to the characteristics of competing artifacts based on the design objectives derived from relevant literature. This comparison indicates that the design specification of C2VR is fit for the purpose of solving the identified problem, whereas no other approach addresses the design objectives as comprehensively as C2VR. In EVAL3, the article presents an implementation of C2VR as a software prototype to be used with a generic version of the dynamic VRP (Berbeglia et al. 2010). Experiments with simulated scenarios provide evidence of C2VR's applicability and functionality. For EVAL4, the instantiated decision model is applied in a real-world case study with pseudonymised data collected from a German platform-to-consumer delivery service to provide evidence of its practical applicability and usefulness.

To summarise, research article #4 presents the data-driven process improvement C2VR, a decision model for incorporating a long-term customer-centric perspective into typically short-term and efficiency-driven last mile delivery processes. C2VR contributes to the body of prescriptive knowledge concerning customer centricity in last mile delivery. Furthermore, the decision model contributes to the body of knowledge on data-driven BPM in the process execution phase, and it does so by means of a sample application of adaptive process execution (Kerpedzhiev et al. 2021).

As a second example, research article #5 investigates production processes in smart factory networks (SFNs). The high degree of openness and cross-linking of IT systems and physical production components in complex SFNs and new production concepts like just-in-time increase the probability and damage potential of disturbances and errors (Broy et al. 2012; Tupa et al. 2017). Thus, production processes in SFNs are more vulnerable to IT security risks than their counterparts in traditional factories (Häckel et al. 2018; Smith et al. 2007; Tupa et al. 2017; Yoon et al. 2012). Specifically, IT availability threats (e.g., caused by unintentional errors or intentional IT attacks) within the information network of SFNs affect the availability of the production processes (Amiri et al. 2014; Broy et al. 2012), and threat propagation within the SFN can lead to entire system breakdowns (Kang et al. 2015; Smith et al. 2007).

From a BPM perspective, organisations require appropriate tools to manage such availability risks and associated effects on their production processes (Hallikas et al. 2004). In a first step, data on the architecture and state of SFNs can be used to model SFNs including the information network and respective production processes. Based on this, the identification of weak spots and the derivation of countermeasures becomes feasible. This is essential in order to achieve secure and robust SFN layouts but also to predict threat occurrence and propagation during process execution. Thus, since it is critical for organisations' success and survival to foresee, analyse, and counteract availability risks within their SFNs, the research article poses the following research question: *How can smart factory networks be modelled in order to analyse IT availability risks and associated effects on production?*

To answer this question, the article builds on previous research projects (Berger et al. 2019) and adopts the DSR Methodology as per Peffers et al. (2007) to design and evaluate a modelling approach for IT availability risks in SFNs. Drawing on a broad literature review to derive formal and functional modelling requirements, the article employs Petri Nets (PNs; Petri 1966) in the DSR design and development phase to provide models of modular information and production components. These modular components represent resources involved in the SFN production processes which are relevant for analysing availability risks. The modelling approach enables the detailed modelling, simulation, and analysis of attack and error propagation in IT networks and associated effects on the production processes.

The article draws on existing research to define SFNs as the combination of an information network and a production network containing one or more inter-connected production processes (e.g., Brettel et al. 2014; Lasi et al. 2014; Osterrieder et al. 2020; Radziwon et al. 2014; Xu et al. 2018). Figure 9 illustrates the composition of a schematic SFN: The information network consists of connected information components, while the production network is built of production machines. Information components do not represent loose entities of operating software but are highly connected and build hierarchies to enable flexible and dynamic production processes. By providing information to the production network, SFNs enable the production of customised products, just in time and down to lot size. Within the typical production network, a product has to undergo several production steps before being completed. Thereby, each machine conducts one production step. One production step can be conducted by multiple production machines. Hence, the machines are arranged either sequentially or in parallel, i.e., forking the control flow. Machines can be either smart (i.e., they require product-specific information to process customised products) or simple (i.e., they treat every product equally without using additional information).

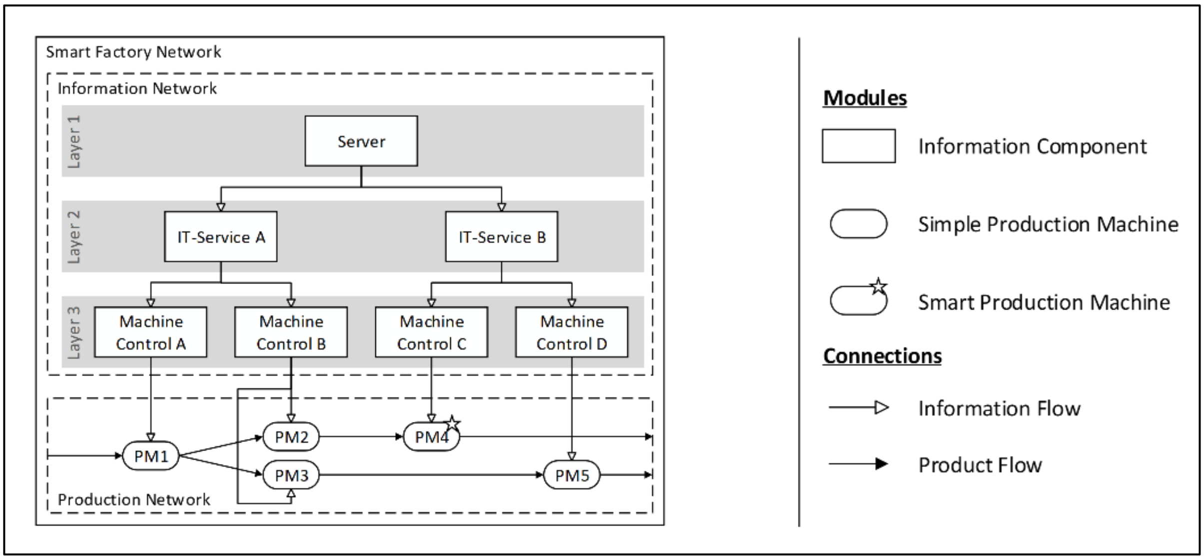
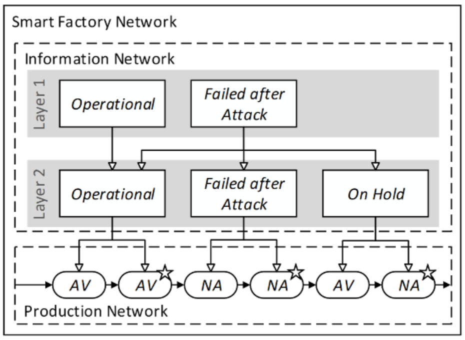



**Figure 9** - Exemplary Layout of a Smart Factory Network

Each component (i.e., information component or production machine) is modelled as a finite state machine using PNs and several extensions (De La Mota et al. 2017; Jensen 1991; Valk 1981). To model the availability of a component, the article adopts the conceptualisation by Miehle et al. (2019) who define four states of information components: the functional state 'operational' (OP), the semi-functional state 'on hold' (OH), and the non-functional states' failed after error' (FE) and 'failed after attack' (FA). Production machines can be in the state 'available'

(AV) or 'unavailable' (NA). Smart and simple production machines are available when the connected information component is functional, i.e., in OP, and not available when the information component is non-functional, i.e., in FE or FA. The OH state represents a functional information component that is connected to a non-functional information component on a higher level. Components in OH do not provide information to smart production machines, making them not available, whereas simple machines remain available as they do not require information. An overview of these dependencies is given in Figure 10. The effects of occurring IT attacks and errors as well as possible propagation effects within the network on the availability of the production processes can now be modelled and analysed in detail using stochastic simulation.




| | State of Connected Information Component | | | |
|---|---|---|---|---|
| | Operational | On Hold | Failed after Error | Failed after Attack |
| Simple Production Machine | AV | AV | NA | NA |
| Smart Production Machine | AV | NA | NA | NA |

AV = available; NA = not available

**Figure 10** - States of Production Machines depending on Information Components

To demonstrate and evaluate the modelling approach, the article follows the DSR evaluation framework proposed by Sonnenberg and vom Brocke (2012). For EVAL1, the article justifies the research problem and derives design objectives from existing knowledge of the field. In EVAL2, the article discusses the artifact via feature comparison against the design objectives and competing artifacts. From this comparison, it follows that there is no artefact in existing research that covers all derived design objectives, and that the proposed modelling approach is the first that offers an integrated view of both information and production networks and sufficiently fulfils all formulated design objectives. Additionally, the understandability and real-world fidelity of the design specification is validated with focus groups and an interview with industry experts. For EVAL3, the article presents an implementation of the modelling approach as a software prototype and applies its functionality to laboratory scenarios to evaluate its applicability in artificial settings. Finally, EVAL4 requires validating the instantiation in naturalistic settings. To evaluate the applicability and usefulness of the modelling approach, the prototype is applied in a real-world case study with a German manufacturer in the mechanical engineering sector. The results are discussed with stakeholders from the company who find the modelling approach useful and applicable in practice.

In sum, the article presents an approach for modelling, simulating, and analysing IT availability threats in production processes within SFNs. Thus, it adds to the theoretical and practical body of knowledge on production processes in SFNs. Furthermore, the modelling approach represents an example of a data-driven decision support in the context of SFN production processes, using data on SFN architectures and properties to make SFN production processes more robust and secure.

To recapitulate, Section III presents data-driven artifacts that support BPM stakeholders in improving business processes. In particular, it distinguishes between process data as an enabler of process improvement itself and as an enabler of data-driven decision making in business processes. On the one hand, research article #3 proposes an artifact for data-driven and ML-enabled process improvement. And on the other hand, research articles #4 and #5 present two specific approaches for using process or customer data as a basis for building decision support systems to support the execution of business processes, specifically in last mile delivery and production processes.

# IV. Summary and Future Research[4]

## 1 Summary

In light of the ever-increasing availability of process data produced and stored in business information systems, BPM is being supported by a growing number of so-called data-driven methods and tools. Although many methods and tools, e.g., process mining, have already been adopted in practice, organisations still face several challenges in realising their full potential (Wynn et al. 2021). In particular, both research and practice demand support when it comes to managing process data quality or employing process data for further purposes such as process improvement. With the presented research articles, this thesis contributes to advancing the fields of process data quality management and data-driven business process improvement. First, this thesis investigates ways to manage data quality issues both in process-agnostic data sources and in process-aware information systems and data formats. Second, this thesis presents an approach for automated process improvement based on generative machine learning. Finally, this thesis also demonstrates two IT-enabled process improvement artifacts designed to use data in two different types of business processes, specifically logistics and production processes.

Concerning the first topic of process data quality, Section II presents two IT-enabled approaches that complement each other and support process analysts in detecting, identifying, and quantifying process data quality issues in different data sources. Research article #1 examines how process data can be extracted from source systems like relational databases while simultaneously answering the question if this process data is fit-for-use in process mining. For this purpose, an approach is developed that uses established data quality dimensions and metrics to quantify data quality issues in a given relational database. Based on the structure of the relational data and on the data quality quantification, it provides suggestions for how to extract parts of the database in the form of an event log ready for process mining. The approach builds on justificatory knowledge on data quality measurement in the form of dimensions and metrics, and database and event log structures such as key constraints. The implementation provides users with an integrated log generation tool replacing the laborious, time-consuming, and error-prone manual process. The approach contributes to descriptive knowledge on process data quality management by presenting a framework of data quality dimensions that are relevant for data-driven BPM. Additionally, the approach adds to prescriptive knowledge on process data

[4] This section is partly comprised of content taken from the research articles in this thesis. To improve the readability of the text, I have omitted the standard labelling of these citations.

quality and log extraction by providing a structured way of measuring source data quality and using it in log extraction.

Research article #2 complements the approach for process-agnostic data sources by answering the research question of how to manage process data quality in event data. Specifically, it provides an approach for detecting and quantifying timestamp-related quality issues on the basis of a set of data quality dimensions and metrics since timestamps are essential for reproducing the correct ordering of activities in order to obtain accurate process models. Similar to the approach from research article #1, the approach is based on justificatory knowledge on data quality measurement in the form of dimensions and metrics. Additionally, it draws from knowledge on process data structures and human-computer interaction to provide an interactive approach to log analysis and adjustment. The approach extends the contributions of research article #1 to both descriptive and prescriptive knowledge on process data quality management.

Regarding the second topic, Section III presents three concrete examples of data-driven process improvement with different methods and tools and in different application domains. Research article #3 examines the opportunities created by advances in AI for data-driven process improvement. Considering recent progress in the fields of AI and computational creativity, it examines how processes can be improved with the help of generative machine learning. The article builds and evaluates ProcessGAN, a novel approach for business process improvement based on generative adversarial networks. ProcessGAN is presented as a proof-of-concept to facilitate future research and application, and as a step towards further automation of process improvement. The evaluation demonstrates that ProcessGAN not only creates new process models but is also able to unlock the human creative potential in commonly overlooked areas by providing stimuli. ProcessGAN is positioned as an instantiation of a new class of systems called APIS. APISs support and automate process improvement and represent an emerging subcategory within PAISs that has not yet been explored in a structured way. Future research into APISs is promoted by a set of implications for APISs derived from the development of ProcessGAN. The article's main contribution is twofold, concerning both operational principles and architectures, and a situated implementation (Gregor and Hevner 2013).

Research article #4 highlights the importance of customer centricity in customer-facing business processes. In particular, it examines the problem of customer-centric process designs for last-mile delivery processes. To this end, it uses multi-criteria decision analysis and normative analytical modelling to create C2VR, a decision model that incorporates a customer-centric perspective on the last-mile delivery process. C2VR extends existing efficiency-driven vehicle routing problems and takes customer satisfaction into account. C2VR uses prescriptive justificatory knowledge gained from logistics and operations research. To provide balanced decision support, C2VR draws on descriptive knowledge of customer centricity in BPM and CRM. C2VR contributes to the body of prescriptive knowledge concerning customer centricity in last mile delivery. The decision model is, to the best of our knowledge, the first model and instantiated application that solves a dynamic vehicle routing problem in the context of last mile delivery while navigating a cost-effective balance between efficiency and customer-centricity. Furthermore, this decision model contributes to the body of knowledge on data-driven BPM in the process execution phase, and it does so by means of a sample application of adaptive process execution (Kerpedzhiev et al. 2021).

Finally, research article #5 addresses the topic of IT availability risks in smart factory networks. Specifically, it demonstrates improvement potential for production processes in SFNs by providing a modelling approach for analysing the effects of IT threats on these production processes. Based on Petri nets, the article presents modular SFN components for modelling SFN architectures and for simulating stochastic attack and error propagation. The modelling language supports the analysis and comparison of different SFN and production process architectures regarding spreading effects, availability of information and production components, and associated effects on productivity of the underlying production processes. The approach enables and serves as a foundation for decision support on SFN layouts from a risk perspective and the derivation of IT security mitigation measures in both research and practice. This work adds to the theoretical body of knowledge on SFNs and smart production processes by extending the understanding of IT availability risks in SFNs. The modelling approach also provides a practical contribution by enabling the modelling, simulation, and analysis of individual components, interdependencies, production processes, and the entire SFN under attack and error occurrence and propagation.

## 2 Limitations and Future Research

Like any research endeavour, this doctoral thesis is beset with limitations and questions left unanswered that warrant further research. This subsection provides an aggregated overview of the thesis' limitations while detailed descriptions of the limitations of each of the research articles are addressed in the individual articles (see Appendix VI.3 to VI.7). It also provides directions and concrete ideas for further research into process data quality and data-driven process improvement.

First, regarding process data quality, this thesis takes a design-oriented view on process data quality management approaches. Both discussed articles present artifacts for decision support in the data preparation stage of process mining projects. A central limitation to such approaches can be found in the correct balance between automation on the one hand and manual input provided by the user on the other hand. While automated approaches facilitate the user's tasks, they so far lack the ability to consider the context of the process data in its application domain and might therefore over- or underestimate data quality issues. Furthermore, both presented approaches consider only a defined subset of existing data quality dimensions and respective metrics. Additionally, research article #1 only demonstrates two specific event constructors for log extraction from relational databases. Based on these limitations, future research should continue to refine such (semi-)automated approaches for process data quality measurement to make them more independent of user input and more robust in different application domains. In the area of process data quality management, the next reasonable steps in research include the design of approaches for process data repair and process data provenance on a broad scale. Lastly, from a process mining perspective, it is imperative to further investigate the impact of data quality on the quality of process mining results.

Second, this thesis again takes a design-oriented perspective to create advances in data-driven process improvement. All three discussed articles come with limitations in the area of practical applicability: While all three presented artifacts have been evaluated in practical settings, the articles do not yet provide insights into the behaviour and usefulness of the artifacts in real operation. The robustness and generalisability of the artifacts in various different contexts is yet to be understood. Regarding research article #3 specifically, research on generative machine learning for process improvement is in its infancy and all results are to be treated with caution. All of this warrants further research: The applicability and usefulness of the described artifacts would benefit from extending their functionality in different ways, as described in the specific articles. In addition, research article #4 calls for further research into customer value and customer satisfaction in logistics processes, while research article #5 suggests the investigation of further IT security dimensions besides availability. Finally, research article #3 provides a foundation for the exploration of the system class of APIS, e.g., in the form of reference architectures or further exemplary instantiations.

In sum, this thesis contributes to the existing bodies of knowledge on process data quality management and data-driven process improvement, respectively. Next to existing traditional approaches to BPM, organisations will have to establish data-driven BPM capabilities if they are to remain competitive. These capabilities go beyond what is shown in this thesis: In the digital age, organisations must be able to leverage data both at design and at run time of business processes – for descriptive, but also for prescriptive and predictive purposes.

In its final form, this development will eventually lead to a completely new understanding of BPM which uses advanced data analytics and AI methods to achieve continuous process improvement. Researchers have coined this new understanding *augmented BPM* (Dumas 2021). Augmented BPM will not only answer questions on the historical, current, and future state of organisations' business processes but will instead be able to guide and execute business processes with more autonomy and natural interaction than ever before. Future BPM systems may be able to (1) automatically detect and implement opportunities for process improvement at design time, and (2) autonomously drive and adapt process execution at run time. To facilitate this change, future BPM research should go beyond the current state of research on descriptive, predictive, and prescriptive BPM methods and tools and further examine the necessary prerequisites for augmented BPM. These prerequisites include not only the necessary technological advances in data analytics and AI but also the capabilities necessary for adopting augmented BPM in practice.

At the same time, organisations should not lose track of this development towards augmented BPM. While descriptive and predictive BPM has matured (e.g., in the form of process mining) and the importance of prescriptive BPM in academic literature is rapidly growing (as shown in this thesis), many organisations are still doing groundwork. First and foremost, organisations still struggle with process data collection and preparation as well as the alignment of basic data-driven BPM initiatives to the corporate strategy. Such organisations are at risk of being left behind in a race to take advantage of the emerging opportunities provided by data-driven – and, later on, augmented – BPM. Therefore, organisations are well-advised to start laying the foundation for data-driven BPM.

In this spirit, I hope that the results of this thesis help to facilitate further research in and the practical adoption of data-driven and augmented BPM and, thus, support organisations in continually improving and transforming their business processes in order to thrive in today's fast-moving world.

# V. Publication Bibliography


Afflerbach P, Frank L (2016) Customer Experience Versus Process Efficiency: Towards an Analytical Framework About Ambidextrous BPM. In: 37th Intl Conference on Information Systems, Dublin

Afflerbach P, Hohendorf M, Manderscheid J (2017) Design it like Darwin - A value-based application of evolutionary algorithms for proper and unambiguous business process redesign. Inf Syst Front 19:1101–1121. https://doi.org/10.1007/s10796-016-9715-1

Ahmadi-Javid A, Amiri E, Meskar M (2018) A Profit-Maximization Location-Routing-Pricing Problem: A Branch-and-Price Algorithm. European Journal of Operational Research 271:866–881. https://doi.org/10.1016/j.ejor.2018.02.020

Al-Anqoudi Y, Al-Hamdani A, Al-Badawi M, Hedjam R (2021) Using Machine Learning in Business Process Re-Engineering. BDCC 5:61. https://doi.org/10.3390/bdcc5040061

Allweyer T (2014) BPMS: Einführung in Business Process Management-Systeme. Books on Demand, Norderstedt

Amiri A, Cavusoglu H, Benbasat I (2014) When is IT Unavailability a Strategic Risk?: A Study in the Context of Cloud Computing. In: ICIS

Andrews R, Wynn M, Hofstede AHM, Xu J, Horton K, Taylor P, Plunkett-cole S (2018a) Exposing Impediments to Insurance Claims Processing. In: BPM Cases, pp 275–290

Andrews R, Suriadi S, Wynn M, Hofstede AHM ter, Rothwell S (2018b) Improving Patient Flows at St. Andrew’s War Memorial Hospital’s Emergency Department Through Process Mining. In: BPM Cases, pp 311–333

Andrews R, Wynn MT, Valmuur K, Hofstede AHM ter, Bosley E, Elcock M, Rashford S (2019) Leveraging Data Quality to Better Prepare for Process Mining: an Approach Illustrated Through Analysing Road Trauma Pre-Hospital Retrieval and Transport Processes in Queensland. Int. J. Environ. Res. Publ. Health 16. https://doi.org/10.3390/ijerph16071138

Barkaoui M, Berger J, Boukhtouta A (2015) Customer satisfaction in dynamic vehicle routing problem with time windows. Applied Soft Computing 35:423–432. https://doi.org/10.1016/j.asoc.2015.06.035

Batini C, Scannapieco M (2016) Data quality dimensions. In: Data and information quality: Dimensions, principles and techniques. Springer International Publishing, Cham, pp 21–51

Baumbach S, Oberländer AM, Röglinger M, Rosemann M (2020) Dynamic capabilities for opportunity exploration: insights from an explorative case study. International Journal of Entrepreneurial Venturing 12:575–616

Becker J, Kugeler M, Rosemann M (eds) (2012) Prozessmanagement: Ein Leitfaden zur prozessorientierten Organisationsgestaltung. Springer Gabler, Berlin, Heidelberg

Berbeglia G, Cordeau J-F, Laporte G (2010) Dynamic pickup and delivery problems. European Journal of Operational Research 202:8–15. https://doi.org/10.1016/j.ejor.2009.04.024

Berger S, Bogenreuther M, Häckel B, Niesel O (2019) Modelling Availability Risks of IT Threats in Smart Factory Networks: A Modular Petri Net Approach. In: Proceedings of the 27th European Conference on Information Systems (ECIS), Stockholm & Uppsala, Sweden

Bernstein A, Klein M, Malone T (1999) The process recombinator: a tool for generating new business process ideas. ICIS 1999 Proceedings:178–192

Bettencourt LA, Brown SW, Sirianni NJ (2013) The secret to true service innovation. Business Horizons 56:13–22. https://doi.org/10.1016/j.bushor.2012.09.001

Beverungen D, Buijs JC, Becker J, Di Ciccio C, van der Aalst WMP, Bartelheimer C, vom Brocke J, Comuzzi M, Kraume K, Leopold H (2021) Seven paradoxes of business process management in a hyper-connected world. Bus Inf Syst Eng 63:145–156

Borgianni Y, Cascini G, Rotini F (2015) Business Process Reengineering driven by customer value: a support for undertaking decisions under uncertainty conditions. Computers in Industry 68:132–147. https://doi.org/10.1016/j.compind.2015.01.001

Bortolotti T, Romano P (2012) 'Lean first, then automate': a framework for process improvement in pure service companies. A case study. Production Planning & Control 23:513–522. https://doi.org/10.1080/09537287.2011.640040

Bose R, Mans R, van der Aalst W (2013) Wanna Improve Process Mining Results? It‘s High Time We Consider Data Quality Issues Seriously. In: 2013 IEEE Symposium on Computational Intelligence and Data Mining (CIDM), pp 127–134

Bozorgi ZD, Teinemaa I, Dumas M, La Rosa M, Polyvyanyy A (2020) Process Mining Meets Causal Machine Learning: Discovering Causal Rules from Event Logs. In: 2nd International Conference on Process Mining (ICPM). IEEE, pp 129–136

Brettel M, Friederichsen N, Keller M, Rosenberg M (2014) How Virtualization, Decentralization and Network Building Change the Manufacturing Landscape: An Industry 4.0 Perspective. Intl Journal of Information & Communication Engineering 8:1–8. https://doi.org/10.5281/zenodo.1336426

Broy M, Cengarle MV, Geisberger E (2012) Cyber-Physical Systems: Imminent Challenges. In: Calinescu R, Garlan D (eds) Large-scale complex IT systems: Development, operation and management, vol 7539. Springer, Berlin, pp 1–28

Calvanese D, Montali M, Syamsiyah A, van der Aalst W (2016) Ontology-Driven Extraction of Event Logs from Relational Databases. In: International Conference on Business Process Management, pp 140–153

Calvanese D, Kalayci TE, Montali M, Tinella S (2017) Ontology-based data access for extracting event logs from legacy data: The onprom tool and methodology. In: International Conference on Business Information Systems, pp 220–236

Cheng R, Gen M, Tozawa T (1995) Vehicle Routing Problem with Fuzzy Due-Time Using Genetig Algorithms. Journal of Japan Society for Fuzzy Theory and Systems 7:1050–1061

Chollet F (2018) Deep learning with Python. Safari Tech Books Online. Manning, Shelter Island, NY

Cohon JL (2004) Multiobjective Programming and Planning. Courier Corporation

Davenport TH (1997) Process innovation: Reengineering work through information technology, 11th edn. Harvard Business School Press, Boston, Mass.

Davis FD (1989) Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. MIS Quarterly 13:319. https://doi.org/10.2307/249008

De La Mota IF, Guasch A, Angel Piera M, Mujica Mota M (eds) (2017) Robust Modelling and Simulation. Springer International Publishing, Cham. https://doi.org/10.1007/978-3-319-53321-6

Delias P (2017) A positive deviance approach to eliminate wastes in business processes. IMDS 117:1323–1339. https://doi.org/10.1108/IMDS-09-2016-0393

Dixit PM, Suriadi S, Andrews R, Wynn MT, Hofstede AHM ter, Buijs, Joos C. A. M., van der Aalst WMP (2018) Detection and interactive repair of event ordering imperfection in process logs. In: Krogstie J, Reijers HA (eds) Advanced Information Systems Engineering. Springer International Publishing, Cham, pp 274–290

Dumas M (2021) From Process Mining to Augmented Business Process Management. https://www.linkedin.com/pulse/from-process-mining-augmented-business-management-marlon-dumas/. Accessed 8 February 2022

Dumas M, Mendling J (2019) Business Process Event Logs and Visualization. In: Sakr S, Zomaya AY (eds) Encyclopedia of Big Data Technologies. Springer, Cham, pp 398–409

Dumas M, van der Aalst W, Hofstede AH ter (2005) Process-Aware Information Systems: Bridging People and Software Through Process Technology, 1st edn. John Wiley & Sons, New York, NY

Dumas M, La Rosa M, Mendling J, Reijers HA (2018) Fundamentals of Business Process Management. Springer Berlin Heidelberg, Berlin, Heidelberg. https://doi.org/10.1007/978-3-662-56509-4

Dumas M, Fournier F, Limonad L, Marrella A, Montali M, Rehse J-R, Accorsi R, Calvanese D, Giacomo G de, Fahland D, Gal A, La Rosa M, Völzer H, Weber I (2022) Augmented Business Process Management Systems: A Research Manifesto. https://arxiv.org/pdf/2201.12855

Elgammal A, Liu B, Elhoseiny M, Mazzone M (2017) CAN: Creative Adversarial Networks, Generating "Art" by Learning About Styles and Deviating from Style Norms. http://arxiv.org/pdf/1706.07068v1

Emamjome F, Andrews R, Hofstede AHM ter (2019) A Case Study Lens on Process Mining in Practice. In: Panetto H, Debruyne C, Hepp M, Lewis D, Ardagna CA, Meersman R (eds) OTM 2019 Conferences. Springer International Publishing, Cham, pp 127–145

Even A, Shankaranarayanan G (2007) Utility-Driven Assessment of Data Quality. ACM SIGMIS Database: the DATABASE for Advances in Information Systems 38:75–93

Evermann J, Rehse J-R, Fettke P (2017) Predicting Process Behaviour Using Deep Learning. Decision Support Systems 100:129–140

Fehrer T, Fischer DA, Leemans SJ, Röglinger M, Wynn MT (2022) An assisted approach to business process redesign. DSS:113749. https://doi.org/10.1016/j.dss.2022.113749

Fischer DA, Goel K, Andrews R, van Dun CGJ, Wynn MT, Röglinger M (2020) Enhancing Event Log Quality: detecting and Quantifying Timestamp Imperfections. In: Fahland D, Ghidini C, Becker J, Dumas M (eds) Business Process Management. Springer International Publishing, Cham, pp 309–326

Fox F, Aggarwal VR, Whelton H, Johnson O (2018) A Data Quality Framework for Process Mining of Electronic Health Record Data. In: 2018 IEEE International Conference on Healthcare Informatics (ICHI), pp 12–21

Frank L, Poll R, Röglinger M, Rupprecht L (2020) Design heuristics for customer-centric business processes. Business Process Mgmt Journal 26:1283–1305. https://doi.org/10.1108/BPMJ-06-2019-0257

Galbraith JR (2005) Designing the Customer-Centric Organization: A Guide to Strategy, Structure, and Process. Jossey-Bass, San Francisco

Görz Q, Kaiser M (2012) An Indicator Function for Insufficient Data Quality - A Contribution to Data Accuracy. In: Mediterranean Conference on Information Systems, pp 169–184

Gregor S, Hevner AR (2013) Positioning and Presenting Design Science Research for Maximum Impact. MIS Quarterly 37:337–355. https://doi.org/10.25300/MISQ/2013/37.2.01

Grisold T, Groß S, Stelzl K, vom Brocke J, Mendling J, Röglinger M, Rosemann M (2021) The Five Diamond Method for Explorative Business Process Management. Bus Inf Syst Eng. https://doi.org/10.1007/s12599-021-00703-1

Gross S, Malinova M, Mendling J (2019) Navigating Through the Maze of Business Process Change Methods. In: Bui T (ed) 52nd HICSS

Gross S, Stelzl K, Grisold T, Mendling J, Röglinger M, vom Brocke J (2020) The Business Process Design Space for exploring process redesign alternatives. Business Process Mgmt Journal. https://doi.org/10.1108/BPMJ-03-2020-0116

Gschwandtner T, Gärtner J, Aigner W, Miksch S (2012) A taxonomy of dirty time-oriented data. In: Quirchmayr G, Basl J, You I, Xu L, Weippl E (eds) Multidisciplinary Research and Practice for Information Systems. Springer, Berlin, Heidelberg, pp 58–72

Günther CW, van der Aalst W (2006) A Generic Import Framework For Process Event Logs. In: International Conference on Business Process Management, pp 81–92

Günther CW, Verbeek HMW (2014) XES - standard definition. BPM reports. BPMcenter.org

Gupta S, Lehmann DR (2003) Customers as assets. Journal of Interactive Marketing 17:9–24. https://doi.org/10.1002/dir.10045

Häckel B, Hänsch F, Hertel M, Übelhör J (2018) Assessing IT availability risks in smart factory networks. Business Research. https://doi.org/10.1007/s40685-018-0071-5

Hallikas J, Karvonen I, Pulkkinen U, Virolainen V-M, Tuominen M (2004) Risk management processes in supplier networks. International Journal of Production Economics 90:47–58. https://doi.org/10.1016/j.ijpe.2004.02.007

Hammer M (2015) What is Business Process Management? In: vom Brocke J, Rosemann M (eds) Handbook on Business Process Management 1. Springer, Berlin, pp 3–16

Hammer M, Champy J (1993) Business process reengineering. London: Nicholas Brealey 444:730–755

Han X, Hu L, Dang Y, Agarwal S, Mei L, Li S, Zhou X (2020) Automatic Business Process Structure Discovery using Ordered Neurons LSTM: A Preliminary Study. http://arxiv.org/pdf/2001.01243v1

Heinrich B, Klier M (2015) Metric-Based Data Quality Assessment: Developing and Evaluating a Probability-Based Currency Metric. Decision Support Systems 72:82–96

Heinrich B, Hristova D, Klier M, Schiller A, Szubartowicz M (2018) Requirements for Data Quality Metrics. JDIQ 9:1–32

Heinrich K, Zschech P, Janiesch C, Bonin M (2021) Process data properties matter: Introducing gated convolutional neural networks (GCNN) and key-value-predict attention networks (KVP) for next event prediction with deep learning. Decision Support Systems 143:113494. https://doi.org/10.1016/j.dss.2021.113494

Hevner AR, March ST, Park J, Ram S (2004) Design Science in Information Systems Research. MIS Quarterly 28:75–105. https://doi.org/10.2307/25148625

Huang SY, Lee C-H, Chiu A-A, Yen DC (2015) How business process reengineering affects information technology investment and employee performance under different performance measurement. Inf Syst Front 17:1133–1144. https://doi.org/10.1007/s10796-014-9487-4

Jensen K (1991) Coloured Petri Nets: A High Level Language for System Design and Analysis. In: Jensen (ed) High-level Petri nets, pp 44–119

Juran JM, Godfrey AB (1999) Juran's Quality Handbook. 0968-4875. https://doi.org/10.1108/09684879310045286

Kaiser M, Klier M, Heinrich B (2007) How to Measure Data Quality? A Metric-Based Approach. In: ICIS 2007 Proceedings, pp 1–15

Kang CM, Hong YS, Huh WT, Kang W (2015) Risk Propagation Through a Platform: The Failure Risk Perspective on Platform Sharing. IEEE Transactions on Engineering Management 62:372–383. https://doi.org/10.1109/TEM.2015.2427844

Karagiannis D (1995) BPMS. SIGOIS Bull. 16:10–13. https://doi.org/10.1145/209891.209894

Kerpedzhiev GD, König UM, Röglinger M, Rosemann M (2021) An Exploration into Future Business Process Management Capabilities in View of Digitalization. Bus Inf Syst Eng 63:83–96. https://doi.org/10.1007/s12599-020-00637-0

Kettinger WJ, Teng JTC, Guha S (1997) Business Process Change: A Study of Methodologies, Techniques, and Tools. MIS Quarterly 21:55. https://doi.org/10.2307/249742

König UM, Linhart A, Röglinger M (2019) Why do business processes deviate? Results from a Delphi study. Bus Res 12:425–453. https://doi.org/10.1007/s40685-018-0076-0

Kratsch W, Manderscheid J, Reißner D, Röglinger M (2017) Data-driven Process Prioritization in Process Networks. DSS 100:27–40. https://doi.org/10.1016/j.dss.2017.02.011

Kratsch W, Manderscheid J, Röglinger M, Seyfried J (2020) Machine Learning in Business Process Monitoring: A Comparison of Deep Learning and Classical Approaches Used for Outcome Prediction. Bus Inf Syst Eng. https://doi.org/10.1007/s12599-020-00645-0

Kreuzer T, Röglinger M, Rupprecht L (2020) Customer-centric prioritization of process improvement projects. Decision Support Systems 133:113286. https://doi.org/10.1016/j.dss.2020.113286

Kurniati AP, Rojas E, Hogg D, Hall G, Johnson OA (2018) The Assessment of Data Quality Issues for Process Mining in Healthcare using MIMIC III. Health Informatics Journal 25:1878–1893

Kwak YH, Anbari FT (2006) Benefits, obstacles, and future of six sigma approach. Technovation 26:708–715. https://doi.org/10.1016/j.technovation.2004.10.003

Lasi H, Fettke P, Kemper H-G, Feld T, Hoffmann M (2014) Industry 4.0. Business & Information Systems Engineering 6:239–242. https://doi.org/10.1007/s12599-014-0334-4

LeCun Y, Bengio Y, Hinton G (2015) Deep learning. Nature 521:436–444. https://doi.org/10.1038/nature14539

Lee YW, Strong DM, Kahn BK, Wang RY (2002) AIMQ: A methodology for information quality assessment. Information and Management 40:133–146. https://doi.org/10.1016/S0378-7206(02)00043-5

Leno V, Polyvyanyy A, Dumas M, La Rosa M, Maggi FM (2021) Robotic Process Mining: Vision and Challenges. Bus Inf Syst Eng 63:301–314. https://doi.org/10.1007/s12599-020-00641-4

Li G, Murillas EGL de, Carvalho RM de, van der Aalst W (2018) Extracting Object-Centric Event Logs to Support Process Mining on Databases. In: International Conference on Advanced Information Systems Engineering. Springer, pp 182–199

Limam Mansar S, Reijers HA (2007) Best practices in business process redesign: use and impact. BPMJ 13:193–213. https://doi.org/10.1108/14637150710740455

Limam Mansar S, Reijers HA, Ounnar F (2009) Development of a decision-making strategy to improve the efficiency of BPR. Expert Systems With Applications 36:3248–3262. https://doi.org/10.1016/j.eswa.2008.01.008

Malinova M, Gross S, Mendling J (2022) A study into the contingencies of process improvement methods. Information Systems 104:101880. https://doi.org/10.1016/j.is.2021.101880

Mans RS, Reijers HA, Berends H, Bandara W, Prince R (2013) Business Process Mining Success. In: ECIS 2013

Mansar SL, Reijers HA (2005) Best practices in business process redesign: validation of a redesign framework. Computers in Industry 56:457–471. https://doi.org/10.1016/j.compind.2005.01.001

March ST, Smith GF (1995) Design and natural science research on information technology. Decision Support Systems 15:251–266

Martin N, Swennen M, Depaire B, Jans M, an Caris, Vanhoof K (2017) Retrieving Batch Organisation of Work Insights From Event Logs. Decis. Support Syst. 100:119–128. https://doi.org/10.1016/j.dss.2017.02.012

Martin N, Fischer DA, Kerpedzhiev GD, Goel K, Leemans SJJ, Röglinger M, van der Aalst WMP, Dumas M, La Rosa M, Wynn MT (2021) Opportunities and Challenges for Process Mining in Organizations: Results of a Delphi Study. Bus Inf Syst Eng 63:511–527. https://doi.org/10.1007/s12599-021-00720-0

Marttunen M, Lienert J, Belton V (2017) Structuring problems for Multi-Criteria Decision Analysis in practice: A literature review of method combinations. European Journal of Operational Research 263:1–17. https://doi.org/10.1016/j.ejor.2017.04.041

Meredith JR, Raturi A, Amoako-Gyampah K, Kaplan B (1989) Alternative research paradigms in operations. Journal of Operations Management 8:297–326. https://doi.org/10.1016/0272-6963(89)90033-8

Miehle D, Häckel B, Pfosser S, Übelhör J (2019) Modeling IT Availability Risks in Smart Factories: a Stochastic Petri Nets Approach. Business & Information Systems Engineering

Möhring M, Schmidt R, Härting R-C, Bär F, Zimmermann A (2015) Classification Framework for Context Data from Business Processes. In: Fournier F, Mendling J (eds) Business Process Management Workshops, vol 202. Springer International Publishing, Cham, pp 440–445

Netjes M, Vanderfeesten I, Reijers HA (2006) "Intelligent" Tools for Workflow Process Redesign: A Research Agenda. In: Hutchison D, Kanade T, Kittler J, Kleinberg JM, Mattern F, Mitchell JC, Naor M, Nierstrasz O, Pandu Rangan C, Steffen B, Sudan M, Terzopoulos D, Tygar D, Vardi MY, Weikum G, Bussler CJ, Haller A (eds) Business Process Management Workshops, vol 3812. Springer Berlin Heidelberg, Berlin, Heidelberg, pp 444–453

Niedermann F, Radeschutz S, Mitschang B (2010) Design-Time Process Optimization through Optimization Patterns and Process Model Matching. In: 2010 IEEE 12th Conference on Commerce and Enterprise Computing. IEEE, pp 48–55

Nooijen E, van Dongen B, Fahland D (2012) Automatic Discovery of Data-centric and Artifact-centric Processes. In: International Conference on Business Process Management, pp 316–327

Osterrieder P, Budde L, Friedli T (2020) The smart factory as a key construct of industry 4.0: A systematic literature review. International Journal of Production Economics 221:107476. https://doi.org/10.1016/j.ijpe.2019.08.011

Partington A, Wynn M, Suriadi S, Ouyang C, Karnon J (2015) Process Mining for Clinical Processes. ACM Transactions on Management Information Systems 5:1–18. https://doi.org/10.1145/2629446

Peffers K, Tuunanen T, Rothenberger MA, Chatterjee S (2007) A Design Science Research Methodology for Information Systems Research. Journal of Management Information Systems 24:45–77. https://doi.org/10.2753/MIS0742-1222240302

Petri CA (1966) Communication with Automata. Diploma Thesis, Technical University of Darmstadt

Pipino LL, Lee YW, Wang RY (2002) Data quality assessment. Commun. ACM 45:211–218. https://doi.org/10.1145/505248.506010

Poll R, Polyvyanyy A, Rosemann M, Röglinger M, Rupprecht L (2018) Process Forecasting: Towards Proactive Business Process Management. In: Weske M, Montali M, Weber I, vom Brocke J (eds) Business Process Management, vol 11080. Springer International Publishing, Cham, pp 496–512

Pöppelbuß J, Röglinger M (2011) What makes a useful maturity model? A framework of general design principles for maturity models and its demonstration in business process management. ECIS

R’bigui H, Cho C (2017) The State-of-the-Art of Business Process Mining Challenges. International Journal of Business Process Integration and Management 8:285–303

Radziwon A, Bilberg A, Bogers M, Madsen ES (2014) The Smart Factory: Exploring Adaptive and Flexible Manufacturing Solutions. Procedia Engineering 69:1184–1190. https://doi.org/10.1016/j.proeng.2014.03.108

Recker J (2012) From Product Innovation to Organizational Innovation – and what that has to do with Business Process Management. BPTrends (Class Notes: BPM Research and Education)

Recker J, Mendling J (2016) The State of the Art of Business Process Management Research as Published in the BPM Conference. Bus Inf Syst Eng 58:55–72. https://doi.org/10.1007/s12599-015-0411-3

Redman TC, Blanton A (1997) Data Quality for the Information Age. Artech House

Reijers HA (ed) (2003) Design and Control of Workflow Processes: Business Process Management for the Service Industry. Springer eBook Collection Computer Science, vol 2617. Springer-Verlag Berlin Heidelberg, Berlin, Heidelberg. https://doi.org/10.1007/3-540-36615-6

Reijers HA (2021) Business Process Management: The evolution of a discipline. Computers in Industry 126. https://doi.org/10.1016/j.compind.2021.103404

Reinkemeyer L (ed) (2020) Process mining in action: Principles, use cases and outlook. Springer, Cham

Röglinger M, van Dun C, Fehrer T, Fischer DA, Moder L, Kratsch W (2021) Automated Process (Re-)Design. In: Intl Workshop on BPM Problems '21 co-located with BPM 2021

Rosemann M (2018) The NESTT: Rapid Process Redesign at Queensland University of Technology. In: vom Brocke J, Mendling J (eds) Business Process Management Cases. Springer International Publishing, Cham, pp 169–185

Rosemann M (2020) Explorative Process Design Patterns. In: Fahland D, Ghidini C, Becker J, Dumas M (eds) Business Process Management, vol 12168. Springer International Publishing, Cham, pp 349–367

Rosemann M, Bruin Td (2005) Towards a Business Process Management Maturity Model. In: ECIS

Rosemann M, vom Brocke J (2015) The Six Core Elements of Business Process Management. In: vom Brocke J, Rosemann M (eds) Handbook on Business Process Management 1. Springer, Berlin, pp 105–122

Sbai O, Elhoseiny M, Bordes A, LeCun Y, Couprie C (2019) DesIGN: Design Inspiration from Generative Networks. In: Leal-Taixé L, Roth S (eds) Computer Vision – ECCV 2018 Workshops, vol 11131. Springer International, Cham, pp 37–44

Scannapieco M, Missier P, Batini C (2005) Data Quality at a Glance. Datenbank-Spektrum

Sengupta S, Basak S, Saikia P, Paul S, Tsalavoutis V, Atiah F, Ravi V, Peters A (2020) A review of deep learning with special emphasis on architectures, applications and recent trends. Knowledge-Based Systems 194. https://doi.org/10.1016/j.knosys.2020.105596

Setiawan MA, Sadiq S (2013) A Methodology for Improving Business Process Performance through Positive Deviance. International Journal of Information System Modeling and Design 4:1–22. https://doi.org/10.4018/jismd.2013040101

Sivaramkumar V, Thansekhar, Saravanan R, Miruna Joe Amali S (2017) Multi-objective vehicle routing problem with time windows: Improving customer satisfaction by considering gap time. Journal of Engineering Manufacture 231:1248–1263. https://doi.org/10.1177/0954405415586608

Smith GE, Watson KJ, Baker WH, Pokorski II JA (2007) A critical balance: collaboration and security in the IT-enabled supply chain. Intl Journal of Production Research 45:2595–2613. https://doi.org/10.1080/00207540601020544

Sonnenberg C, vom Brocke J (2012) Evaluations in the Science of the Artificial - Reconsidering the Build-Evaluate Pattern in Design Science Research. In: Peffers K, Rothenberger M, Kuechler B (eds) Design Science Research in Information Systems. Advances in Theory and Practice 2012. Springer, Berlin, Heidelberg, pp 381–397

Stvilia B, Gasser L, Twidale MB, Smith LC (2007) A framework for information quality assessment. J. Am. Soc. Inf. Sci. Technol. 58:1720–1733. https://doi.org/10.1002/asi.20652

Suriadi S, Wynn MT, Xu J, van der Aalst W, Hofstede AHM ter (2017a) Discovering work prioritisation patterns from event logs. Decision Support Systems 100:77–92

Suriadi S, Andrews R, Hofstede AHM ter, Wynn MT (2017b) Event log imperfection patterns for process mining - towards a systematic approach to cleaning event logs. Inf. Syst. 64:132–150. https://doi.org/10.1016/j.is.2016.07.011

Tarhan A, Turetken O, Reijers HA (2016) Business process maturity models: A systematic literature review. Information and Software Technology 75:122–134. https://doi.org/10.1016/j.infsof.2016.01.010

Taymouri F, La Rosa M, Erfani S, Bozorgi ZD, Verenich I (2020) Predictive Business Process Monitoring via Generative Adversarial Nets: The Case of Next Event Prediction. In: Fahland D, Ghidini C, Becker J, Dumas M (eds) Business Process Management, vol 12168. Springer International Publishing, Cham, pp 237–256

Taymouri F, La Rosa M, Dumas M, Maggi FM (2021) Business process variant analysis: Survey and classification. Knowledge-Based Systems 211. https://doi.org/10.1016/j.knosys.2020.106557

Teinemaa I, Dumas M, La Rosa M, Maggi FM (2019) Outcome-Oriented Predictive Process Monitoring. ACM Trans. Knowl. Discov. Data 13:1–57. https://doi.org/10.1145/3301300

Thiede M, Fuerstenau D, Bezerra Barquet AP (2018) How is process mining technology used by organizations? A systematic literature review of empirical studies. Business Process Mgmt Journal 24:900–922. https://doi.org/10.1108/BPMJ-06-2017-0148

Trkman P, Mertens W, Viaene S, Gemmel P (2015) From business process management to customer process management. BPMJ 21:250–266. https://doi.org/10.1108/BPMJ-02-2014-0010

Truong T-M, Le L-S (2016) On Business Process Redesign and Configuration: Leveraging Data Mining Classification & Outliers and Artifact-Centric Process Modeling. In: 2016 IEEE ACOMP, pp 59–66

Tupa J, Simota J, Steiner F (2017) Aspects of Risk Management Implementation for Industry 4.0. Procedia Manufacturing 11:1223–1230. https://doi.org/10.1016/j.promfg.2017.07.248

Vakulenko Y, Hellström D, Hjort K (2018) What's in the parcel locker? Exploring customer value in e-commerce last mile delivery. Journal of Business Research 88:421–427. https://doi.org/10.1016/j.jbusres.2017.11.033

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

Valk R (1981) Generalizations of Petri nets. In: Gruska C (ed) Mathematical foundations of computer science, vol 118, pp 140–155 van den Hemel C, Rademakers MF (2016) Building Customer-centric Organizations: Shaping Factors and Barriers. Journal of Creating Value 2:211–230. https://doi.org/10.1177/2394964316647822

van der Aalst W (2013) Business Process Management: A Comprehensive Survey. ISRN Software Engineering 2013:1–37. https://doi.org/10.1155/2013/507984

van der Aalst W (2015) Extracting Event Data from Databases to Unleash Process Mining. In: BPM - Driving Innovation in a Digital World. Springer, Cham, pp 105–128

van der Aalst W, Adriansyah A, Medeiros AKA de, Arcieri F, Baier T, Blickle T, Bose JC, van den Brand P, Brandtjen R, Buijs J, Burattin A, Carmona J, Castellanos M, Claes J, Cook J, Costantini N, Curbera F, Damiani E, Leoni M de, Delias P, van Dongen BF, Dumas M, Dustdar S, Fahland D, Ferreira DR, Gaaloul W, van Geffen F, Goel S, Günther C, Guzzo A, Harmon P, ter Hofstede A, Hoogland J, Ingvaldsen JE, Kato K, Kuhn R, Kumar A, La Rosa M, Maggi F, Malerba D, Mans RS, Manuel A, McCreesh M, Mello P, Mendling J, Montali M, Motahari-Nezhad HR, zur Muehlen M, Munoz-Gama J, Pontieri L, Ribeiro J, Rozinat A, Seguel Pérez H, Seguel Pérez R, Sepúlveda M, Sinur J, Soffer P, Song M, Sperduti A, Stilo G, Stoel C, Swenson K, Talamo M, Tan W, Turner C, Vanthienen J, Varvaressos G, Verbeek E, Verdonk M, Vigo R, Wang J, Weber B, Weidlich M, Weijters T, Wen L, Westergaard M, Wynn M (2012) Process Mining Manifesto. In: Daniel F, Barkaoui K, Dustdar S (eds) Business Process Management Workshops, vol 99. Springer Berlin Heidelberg, Berlin, Heidelberg, pp 169–194

van der Aalst W, Artale A, Montali M, Tritini S (2017a) Object-Centric Behavioral Constraints: Integrating Data and Declarative Process Modelling. In: Description Logics, pp 1–12

van der Aalst WMP (2016) Process mining: data science in action, 2nd edn. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-662-49851-4

van der Aalst WMP, La Rosa M, Santoro FM (2016) Business Process Management. Bus Inf Syst Eng 58:1–6. https://doi.org/10.1007/s12599-015-0409-x

van der Aalst WMP, Bichler M, Heinzl A (2017b) Responsible Data Science. Bus. Inf. Syst. Eng. 59:311–313. https://doi.org/10.1007/s12599-017-0487-z van Dun C, Fehrer T, Kratsch W, Wolf N (2020) Customers Like It Hot and Fast : Incorporating Customer Effects into the Meal Delivery Process. In: 28th European Conference on Information Systems (ECIS), Marrakesh, Marocco

Vanwersch RJB, Vanderfeesten I, Rietzschel E, Reijers HA (2015) Improving Business Processes: Does Anybody have an Idea? In: Motahari-Nezhad HR, Recker J, Weidlich M (eds) Business Process Management, vol 9253. Springer International, Cham, pp 3–18

Vanwersch RJB, Shahzad K, Vanderfeesten I, Vanhaecht K, Grefen P, Pintelon L, Mendling J, van Merode GG, Reijers HA (2016) A Critical Evaluation and Framework of Business Process Improvement Methods. Bus Inf Syst Eng 58:43–53. https://doi.org/10.1007/s12599-015-0417-x

Verbeek HMW, Buijs, Joos C. A. M., van Dongen BF, van der Aalst WMP (2011) XES, XESame, and ProM 6. In: Soffer P, Proper E (eds) Information Systems Evolution. Springer, Berlin, Heidelberg, pp 60–75

vom Brocke J, Baier M-S, Schmiedel T, Stelzl K, Röglinger M, Wehking C (2021) Context-Aware Business Process Management. Bus Inf Syst Eng 63:533–550. https://doi.org/10.1007/s12599-021-00685-0

Wand Y, Wang RY (1996) Anchoring data quality dimensions in ontological foundations. Commun. ACM 39:86–95. https://doi.org/10.1145/240455.240479

Wang RY, Strong DM (1996) Beyond accuracy: what data quality means to data consumers. J. Manag. Inf. Syst. 12:5–33. https://doi.org/10.1080/07421222.1996.11518099

Weske M (2019) Business Process Management: Concepts, Languages, Architectures, 3rd edn. Springer Berlin Heidelberg, Berlin, Heidelberg

Winkelhaus S, Grosse EH (2020) Logistics 4.0: a systematic review towards a new logistics system. International Journal of Production Research 58:18–43. https://doi.org/10.1080/00207543.2019.1612964

Wynn MT, Sadiq S (2019) Responsible Process Mining - A Data Quality Perspective. In: Hildebrandt T, van Dongen BF, Röglinger M, Mendling J (eds) Business Process Management. Springer International Publishing, Cham, pp 10–15

Wynn MT, Poppe E, Xu J, Hofstede AHM ter, others (2017) ProcessProfiler3D: A Visualisation Framework for Log-based Process Performance Comparison. Decision Support Systems 100:93–108

Wynn MT, Lebherz J, van der Aalst WM, Accorsi R, Di Ciccio C, Jayarathna L, Verbeek H (2021) Rethinking the Input for Process Mining: Insights from the XES Survey and Workshop. In: Xixi Lu, Jorge Munoz-Gama (eds) Process Mining Workshops - ICPM 2021 International Workshops, Eindhoven, Netherlands, October 31-November 5, 2021, Revised Selected Papers. Springer

Xiang Z, Chu C, Chen H (2008) The study of a dynamic dial-a-ride problem under time-dependent and stochastic environments. EJOR 185:534–551. https://doi.org/10.1016/j.ejor.2007.01.007

Xu LD, Xu EL, Li L (2018) Industry 4.0: state of the art and future trends. International Journal of Production Research 56:2941–2962. https://doi.org/10.1080/00207543.2018.1444806

Yoon J-S, Shin S-J, Suh S-H (2012) A conceptual framework for the ubiquitous factory. International Journal of Production Research 50:2174–2189. https://doi.org/10.1080/00207543.2011.562563

Zellner G (2011) A structured evaluation of business process improvement approaches. Business Process Mgmt Journal 17:203–237. https://doi.org/10.1108/14637151111122329

Zemni MA, Mammar A, Hadj-Alouane NB (2016) An automated approach for merging business process fragments. Computers in Industry 82:104–118. https://doi.org/10.1016/j.compind.2016.05.002

Zhang J, Wang W, Zhao Y, Cattani C (2012) Multiobjective Quantum Evolutionary Algorithm for the Vehicle Routing Problem with Customer Satisfaction. Mathematical Problems in Engineering 2012:1–19. https://doi.org/10.1155/2012/879614

Zuhaira B, Ahmad N (2020) Business process modeling, implementation, analysis, and management: the case of business process management tools. Business Process Mgmt Journal 27:145–183. https://doi.org/10.1108/BPMJ-06-2018-0168

# VI. Appendix

## 1 Index of Research Articles

**Research Article #1: Quality-Informed Semi-Automated Event Log Generation for Process Mining**

Andrews R, van Dun C, Wynn MT, Kratsch W, Röglinger M, ter Hofstede AHM (2020) Quality-Informed Semi-Automated Event Log Generation for Process Mining. In: *Decision Support Systems*. 132.

(VHB-JOURQUAL 3: Category B)

**Research Article #2: Towards Interactive Event Log Forensics: Detecting and Quantifying Timestamp Imperfections**

Fischer DA, Goel K, Andrews R, van Dun C, Wynn MT, Röglinger M (2022) Towards Interactive Event Log Forensics: Detecting and Quantifying Timestamp Imperfections. Accepted (minor revisions) in: *Information Systems*.

(VHB-JOURQUAL 3: Category B)

Earlier version published in *Proceedings of the 18th International Conference on Business Process Management (BPM), 2020.*

**Research Article #3: ProcessGAN: Creating Process Design Options through Generative Machine Learning**

van Dun C, Moder L, Kratsch W, Röglinger M (2022) ProcessGAN: Creating Process Design Options through Generative Machine Learning. Submitted to: *Decision Support Systems*.

(VHB-JOURQUAL 3: Category B)

**Research Article #4: Customer-Centric Vehicle Routing: Incorporating Customer Centricity into Last Mile Delivery**

van Dun C, Fehrer T, Kratsch W, Röglinger M (2022) Customer-Centric Vehicle Routing: Incorporating Customer Centricity into Last Mile Delivery Processes. Submitted to: *Electronic Markets*.

Earlier version published in *Proceedings of the 28th European Conference on Information Systems (ECIS), 2020.*

**Research Article #5: IT Availability Risks in Smart Factory Networks – Analyzing the Effects of IT Threats on Production Processes Using Petri Nets**

Berger S, van Dun C, Häckel B (2022) IT Availability Risks in Smart Factory Networks – Analyzing the Effects of IT Threats on Production Processes Using Petri Nets. In: *Information Systems Frontiers*.

(VHB-JOURQUAL 3: Category B)

Earlier version published in *Proceedings of the 27th Conference on European Conference on Information Systems (ECIS), 2019*.

## 2 Individual Contribution to the Included Research Articles

In this cumulative thesis, five research articles build the main body of this work. All research articles were developed in teams with multiple co-authors. Thus, this section details the respective research settings and highlights my individual contribution to each research article.

Research article #1 (Andrews et al. 2020) was developed together with five co-authors. I took a key role in conducting the research project and developing the main artifact of the article based on input from my co-authors. Additionally, I developed the artifact instantiation as a software prototype. Moreover, I was primarily responsible for the underlying literature work, the data collection, preparation, and analysis, and the application and evaluation of the artifact. I also took a key role in revising the article for re-submission. In sum, I was involved in each part of the project.

Research article #2 (Fischer et al. 2022) was developed together with five co-authors. I contributed to this article by co-initiating and co-developing the entire research project. Moreover, I participated in research discussions and provided feedback on the paper's content and structure. In particular, I engaged in the further development of the research idea, the synthesis and presentation of the research results, as well as textual elaboration. Additionally, I co-created the instantiation of the developed artifact as a software prototype. I also took a key role in revising the article for re-submission. Throughout, I had a key role in all parts of the research project.

Research article #3 (van Dun et al. 2022) was developed together with three co-authors. I contributed to this article by co-initiating and co-developing the entire research project. I was mainly responsible for developing the research method and identifying the gap in existing research literature. Moreover, I participated in research discussions and provided feedback on the paper's content and structure. In particular, I engaged in the development of the central artifact of the article. Additionally, I co-created the instantiation of the developed artifact as a software prototype. I also took a key role in textual elaboration and preparing the article for submission. Thus, I held a key role in all parts of the research project.

Research article #4 (van Dun et al. 2022) was developed with three co-authors. Being the leading author, I had the main role in initiating the research project and contributing by co-developing and driving the entire research project from start to finish. I was primarily responsible for carving out the research question, putting together the underlying literature work, developing the decision model for customer-centric last mile delivery processes, and for conducting the evaluation. Although the research article represents my work to a large extent, the three co-authors were involved in all parts of the project and helped to advance our contribution.

Research article #5 (Berger et al. 2022) was developed together with two co-authors. All authors jointly created the Petri Net modelling approach which is central to the article. I was primarily responsible for the underlying literature work and for structuring and implementing the research method based on Design Science Research. Additionally, my contribution included the instantiation of the modelling approach as a software prototype and the creation of a simulation model for evaluation purposes. Moreover, I contributed to the synthesis and presentation of the research results as well as to textual elaboration. I also took a key role in revising the article for re-submission. Thus, my co-authorship is reflected in the entire research project.

# 3 Research Article #1: Quality-Informed Semi-Automated Event Log Generation for Process Mining

Authors: Andrews R, van Dun C, Wynn MT, Kratsch W, Röglinger M, ter Hofstede AHM

Published in: Decision Support Systems

Abstract: Process mining, as with any form of data analysis, relies heavily on the quality of input data to generate accurate and reliable results. A fit-for-purpose event log nearly always requires time-consuming, manual pre-processing to extract events from source data, with data quality dependent on the analyst's domain knowledge and skills. Despite much being written about data quality in general, a generalisable framework for analysing event data quality issues when extracting logs for process mining remains unrealised. Following the DSR paradigm, we present RDB2Log, a quality-aware, semi-automated approach for extracting event logs from relational data. We validated RDB2Log's design against design objectives extracted from literature and competing artifacts, evaluated its design and performance with process mining experts, implemented a prototype with a defined set of quality metrics, and applied it in laboratory settings and in a real-world case study. The evaluation shows that RDB2Log is understandable, of relevance in current research, and supports process mining in practice.

Keywords: Process mining, Data quality, Event log, Log extraction

# 4 Research Article #2: Towards Interactive Event Log Forensics: Detecting and Quantifying Timestamp Imperfections

Authors: Fischer DA, Goel K, Andrews R, van Dun C, Wynn MT, Röglinger M

Accepted in: Information Systems

Abstract: Timestamp information recorded in event logs plays a crucial role in uncovering meaningful insights into business process performance and behaviour via Process Mining techniques. Inaccurate or incomplete timestamps may cause activities in a business process to be ordered incorrectly, leading to unrepresentative process models and incorrect process performance analyses. Thus, the quality of timestamps in an event log should be evaluated thoroughly before the event log is used for any Process Mining activity. To the best of our knowledge, research on the quality assessment of event logs remains scarce. Our work presents a user-guided and semi-automated approach for detecting and quantifying timestamp-related issues in event logs. We define 15 metrics related to timestamp quality across two axes: four levels of abstraction (event, activity, trace, log) and four quality dimensions (accuracy, completeness, consistency, uniqueness). The approach has been implemented as a prototype and evaluated regarding its design specification, instantiation, and usefulness in artificial and naturalistic settings by including experts from research and practice. Overall, our approach paves the way for a systematic and interactive enhancement of event log quality during the data preprocessing phase of Process Mining projects.

Keywords: Process Mining, Event log, Data quality, Timestamps, Quality assessment

# 5 Research Article #3: ProcessGAN: Creating Process Design Options through Generative Machine Learning

Authors: van Dun C, Moder L, Kratsch W, Röglinger M

Working Paper, submitted to Decision Support Systems

Extended abstract:

Business processes are at the very heart of organizations as they reflect their day-to-day operations, encompassing several management levels within and across organizational boundaries, ultimately supporting the achievement of organizational goals. Accordingly, there is extensive research on the design, improvement, and overall management of business processes. Business process management (BPM), which is the associated management discipline, has gained momentum and evolved into a crucial enabler of organizational performance (Dumas 2018). Among activities related to BPM, business process improvement (BPI) is considered the most value-adding one (Zellner 2011). It involves introducing new process designs to address issues or capitalize on opportunities. BPI holds the potential to increase the quality, customer satisfaction, cost, and revenue of business processes, contributing to organizational success.

So far, many BPI approaches have been proposed. Virtually all of them must be carried out manually owing to the creativity-intense nature of BPI. The very act of BPI lacks computational support, making it time-consuming and labour-intensive. Thus, there is a need for BPI itself to be improved (Al-Anqoudi et al. 2021).

Given the availability of data and processing power, Artificial Intelligence (AI) has already been successfully applied to support or replace human activities in several areas. However, AI is not used for idea generation in BPI, although AI also offers significant potential for creative tasks beyond established areas in the context of computational creativity (CC) (Haefner et al 2021). As CC has become popular in recent years, there are already examples in practice mostly based on generative machine learning. Thus, our research objective is as follows: *How can generative machine learning support the creation of business process improvement ideas?*

To answer this question, we build and evaluate ProcessGAN, a novel approach to BPI based on generative adversarial networks, supporting users in the creative task of developing new process designs. In doing so, we follow the design science research (DSR) paradigm (Gregor and Hevner 2013). We iteratively develop and prototypically instantiate ProcessGAN. Our work

builds a bridge between BPI and CC, two domains that have developed in isolation. To the best of our knowledge, we are the first to provide a proof-of-concept showing that generative machine learning can be applied to BPI.

From a theoretical perspective, we aim at stimulating scientific discourse at the intersection of both domains that has high potential for research and – at the same time – substantial practical relevance. Accordingly, ProcessGAN may inspire other approaches based on CC to support BPI, serving as a foundation for systematically exploring the entire class of automated process improvement systems (APISs).

From a practical perspective, our work entails implications for multiple stakeholder groups. First, regarding prospective users such as process designers or analysts, we provide an open-source software prototype. For vendors of process mining and other BPM solutions, ProcessGAN serves as a proof-of-concept and a foundation for further development of their solutions. It shows that vendors should think about using approaches from the CC domain. By doing so, BPM vendors would not only provide tools for descriptive, diagnostic, or predictive tasks, but also incorporate a prescriptive perspective on top, a development that would ultimately benefit prospective users and boost the further uptake of BPM solutions.

Keywords: Business Process Improvement, Business Process Redesign, Generative Adversarial Networks, Generative Machine Learning, Artificial Intelligence

References:

Al-Anqoudi Y, Al-Hamdani A, Al-Badawi M, Hedjam R (2021) Using Machine Learning in Business Process Re-Engineering. BDCC 5:61. https://doi.org/10.3390/bdcc5040061

Dumas M, La Rosa M, Mendling J, Reijers HA (2018) Fundamentals of Business Process Management. Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-56509-4

Gregor S, Hevner AR (2013) Positioning and Presenting Design Science Research for Maximum Impact. MIS Quarterly 37:337–355

Haefner N, Wincent J, Parida V, Gassmann O (2021) Artificial intelligence and innovation management: A review, framework, and research agenda. Technological Forecasting and Social Change 162, 120392. https://doi.org/10.1016/j.techfore.2020.120392.

Zellner G (2011) A structured evaluation of business process improvement approaches. Bus Process Mgmt Journal 17:203–237. https://doi.org/10.1108/14637151111122329

# 6 Research Article #4: Customer-Centric Vehicle Routing: Incorporating Customer Centricity into Last Mile Delivery

Authors: van Dun C, Fehrer T, Kratsch W, Röglinger M

Working Paper, submitted to Electronic Markets

Extended abstract:

In essence, last mile delivery is a business process with a high degree of customer interaction (Frank et al. 2020). This process is characterized by delivering products from one or more depots to customers in multiple locations with a limited number of couriers. The goal of related decision models and algorithms has been to minimize delivery costs and other efforts against the backdrop of operational challenges. Such approaches can be solved computationally to achieve either approximate or full optimality. It is worth noting, however, that with this focus on minimizing costs, most decision models and algorithms take a short-term, efficiency-driven view of the last mile delivery process, while neglecting a long-term customer-centric perspective. Customer satisfaction of last mile delivery is largely determined by delivery time (Barkaoui et al. 2015). Neglecting this relation may lead to repetitive routings with unintended consequences for customer satisfaction in demographics with unfavorable characteristics.

One way for a business to avoid such unintended consequences is to add a customer-centric perspective to last mile delivery. Recently, the business process management (BPM) community has attempted to account for customer centricity in the design and improvement of business processes (Frank et al. 2020; Kreuzer et al. 2020). From this, we can infer that it will be further beneficial for businesses to establish a balance between customer-centric and efficiency-driven perspectives in last mile delivery processes. Our research question follows: *How can last mile delivery be enhanced by incorporating long-term customer centricity?*

To answer this question, we adopt the DSR paradigm proposed by Gregor and Hevner (2013) to build our decision model called Customer-Centric Vehicle Routing (C2VR). Incorporating C2VR into vehicle routing enables last mile delivery businesses to strike an economically viable balance between efficiency (e.g., in terms of costs) and customer centricity. C2VR uses prescriptive knowledge, as gained from logistics and operations research (OR), to conceptualize key constructs and solution algorithms of VRPs. To provide decision support, C2VR draws on descriptive knowledge of customer centricity and customer relationship management.

The decision model assists organizations in determining how incoming orders should be prioritized based on two factors: delivery costs and customer lifetime value. The objective function we introduce to our model strikes a balance between operational efficiency and the customer-centric perspective. The decision model overcomes systematic location-based discrimination of customers by considering their historic satisfaction levels, detecting disadvantaged customers, and prioritizing their orders to prevent customer dissatisfaction. C2VR constitutes an extension of our research on customer-centric last mile delivery by generalizing our decision model to work with all VRP subtypes (van Dun et al. 2020).

The decision model is, to the best of our knowledge, the first model and instantiated application that solves a dynamic vehicle routing problem in the context of last mile delivery while navigating a cost-effective balance between short-term efficiency and long-term customer-centricity. We provide a way to approximate customer satisfaction measures based on delivery times. We also present a model for the CLV based on customers' future order probability which, in turn, is based on their present customer satisfaction.

Keywords: Vehicle Routing Problem, Last Mile Delivery, Customer Centricity, Decision Model, Routing Optimization

References:

Barkaoui M, Berger J, Boukhtouta A (2015) Customer satisfaction in dynamic vehicle routing problem with time windows. Applied Soft Computing 35:423–432. https://doi.org/10.1016/j.asoc.2015.06.035

Frank L, Poll R, Röglinger M, Rupprecht L (2020) Design heuristics for customer-centric business processes. BPMJ 26:1283–1305. https://doi.org/10.1108/BPMJ-06-2019-0257

Gregor S, Hevner AR (2013) Positioning and Presenting Design Science Research for Maximum Impact. MIS Quarterly 37:337–355

Kreuzer T, Röglinger M, Rupprecht L (2020) Customer-centric prioritization of process improvement projects. Decision Support Systems 133:113286. https://doi.org/10.1016/j.dss.2020.113286

van Dun C, Fehrer T, Kratsch W, Wolf N (2020) Customers Like It Hot and Fast – Incorporating Customer Effects into the Meal Delivery Process. 28th European Conference on Information Systems, ECIS 2020, Marrakech, Morocco.

# 7 Research Article #5: IT Availability Risks in Smart Factory Networks – Analyzing the Effects of IT Threats on Production Processes Using Petri Nets

Authors: Berger S, van Dun C, Häckel B

Published in: Information Systems Frontiers

Abstract: In manufacturing, concepts like the Internet of Things or Cyber-physical Systems accelerate the development from traditional production facilities towards smart factories. Thereby, emerging digital technologies increasingly connect information networks with production processes, forming complex smart factory networks (SFNs). Due to their reliance on information flows and the high degree of cross-linking, SFNs are, in particular, vulnerable to IT availability risks caused by attacks and errors. Against this backdrop, we present a modelling approach for analyzing the effects of IT threats on production processes. Based on Petri Nets, we provide modular SFN components for modelling SFN architectures and for simulating stochastic attack and error propagation. With this, we support the analysis and comparison of different SFN architectures regarding spreading effects, availability of information and production components, and associated effects on productivity. Our approach enables and serves as a foundation for decision support on SFN layouts from a risk perspective and the derivation of IT security mitigation measures in both research and practice. We evaluate our artefact by implementing and applying a software prototype in artificial and real-life settings.

Keywords: Smart Factory Network, Information Network, Production Network, IT Availability Risks, Attack Propagation, Petri Nets