Process Innovation with Analytics in Manufacturing

Exploration through Demonstrator Development

Bojer, Casper Solheim

DOI (link to publication from Publisher):
10.54337/aau679682469

Publication date:
2023

Document Version
Publisher's PDF, also known as Version of record

PROCESS INNOVATION WITH ANALYTICS IN MANUFACTURING

EXPLORATION THROUGH DEMONSTRATOR DEVELOPMENT

BY
CASPER SOLHEIM BOJER

DISSERTATION SUBMITTED 2023

AALBORG UNIVERSITY
DENMARK

PROCESS INNOVATION WITH ANALYTICS IN MANUFACTURING

EXPLORATION THROUGH DEMONSTRATOR DEVELOPMENT

by

Casper Solheim Bojer

AALBORG UNIVERSITY
DENMARK

Dissertation submitted October 31st, 2023

Dissertation submitted: October 31st, 2023

PhD supervisor:: Prof. Charles Møller,
Aalborg University / Aarhus University

PhD committee: Associate Professor Peter Nielsen (chair)
Aalborg University, Denmark

Professor Murat Kulahci
Technical University of Denmark, Denmark

Professor Jan vom Brocke
University of Münster, Germany

PhD Series: Faculty of Engineering and Science, Aalborg University

Department: Department of Materials and Production

ISSN (online): 2446-1636
ISBN (online): 978-87-7573-607-2

Published by:
Aalborg University Press
Kroghstræde 3
DK – 9220 Aalborg Ø
Phone: +45 99407140
aauf@forlag.aau.dk
forlag.aau.dk

Printed in Denmark by Stibo Complete, 2024

CV

Casper Solheim Bojer was born in Aalborg in April 1994. After a natural sciences high school education from Aalborghus Gymnasium, he studied Global Business Engineering at Aalborg University, where he graduated in 2017. This was followed by a master's in engineering in Operations and Supply Chain Management from Aalborg University in 2019, where he specialized in data and analytics in Operations, especially for demand forecasting. After graduating, he worked for one year as a Research Assistant at Center for Industrial Production at Aalborg University, where he continued his research in forecasting, culminating in two publications in the International Journal of Forecasting. In 2020, he started as a PhD Student at Center for Industrial Production at Aalborg University as part of the MADE FAST research program from Manufacturing Academy of Denmark. His research is co-sponsored by a large Danish manufacturer, where he has been closely collaborating with the Operations IT department and participated in several analytics projects.

ENGLISH SUMMARY

How can manufacturing companies improve their operational processes with analytics? Improving process performance in manufacturing has long been a concern in academia and practice. Manufacturers are increasingly expecting digital technologies such as analytics to enable the next generation of performance improvements. Improvement can be achieved by redesigning operational processes to take advantage of the ability of analytics to automate or augment tasks – what I refer to as Process Innovation with Analytics. For example, analytics can improve uptime on machines by moving from an interval-based to a predictive maintenance process or improve quality through analytics-enabled automation of quality inspection.

Adopting analytics and realizing performance improvements through process innovation has proved challenging. Analytics is a general-purpose technology, so manufacturers must conduct considerable exploration activities to discover and assess process innovation use cases. As analytics is rarely offered as a finished product or service, assessing use cases requires some degree of analytics development. Analytics development projects, however, are known for high failure rates and a slow transition from proof-of-concept to deployment and piloting.

The motivation for the dissertation is two-fold. Practically, it is motivated by the adoption challenges faced by manufacturers. Academically, it is motivated by the scarcity of empirical research on process innovation with analytics. The research featured close collaboration with industry and took outset in a concrete problem faced by a large Danish manufacturer and retailer: How can manufacturers quickly explore and assess the potential of analytics for process innovation? The dissertation thus sets out to 1) improve understanding of process innovation with analytics and 2) develop prescriptive knowledge for process innovation with analytics.

The research used Action Design Research (ADR) as the research method. Action Design Research is well suited for developing prescriptive knowledge and understanding through engaged design and evaluation activities. The object of design has been an *analytics demonstrator*, which consists of explorative development of an analytics system to gain insights into potential process innovation use cases. More specifically, the empirical activities consisted of 1) participative observation in three analytics initiatives and 2) the development of five analytics demonstrators. In an exploratory phase, the first four demonstrators were used to develop an approach for analytics demonstrator development and extract design principles. This was followed by an evaluation phase, where the fifth demonstrator instantiated, evaluated, and modified them. Finally, managerial prescriptions were developed based on the lessons learned from the engaged design activities.

The dissertation is organized into two parts. In the first part, I motivate the research, establish the relevant background, and present a summary and discussion of the research process and findings. The second part contains the four research papers that make up the main research output. The four papers address the following research questions:

- Which contributions can be developed throughout the ADR research process? (Paper 1)
- How should process innovation with analytics be conceptualized? (Paper 2)
- What drives development speed in machine learning-based process innovation demonstrators? (Paper 3)
- How should analytics demonstrators be developed? (Paper 4)

Paper 1 played a crucial role in establishing the research design through a conceptual investigation of the potential for academic contributions in ADR. The investigation highlighted the potential inherent in the rich empirical data generated by ADR, which can form the basis for case studies and theorizing. This potential for theorizing is leveraged in Paper 2, which develops a conceptualization of process innovation with analytics and proposes a research agenda. The conceptualization highlights that process innovation with analytics is a transformation process requiring coordinated development of analytics, process, and IT infrastructure. A multiple case study of the analytics demonstrators is presented in Paper 3 with the goal of identifying enablers of fast development. The case study highlighted the importance of two factors for fast development: 1) loose coupling between the demonstrator and the existing IT infrastructure and 2) the use of high-level solution building blocks to limit custom development. Based on these insights, Paper 4 formalizes an approach and design principles for fast development of analytics demonstrators, which is subsequently evaluated in a final demonstrator. The approach is based on incremental development and strategies to reduce analytical and infrastructure complexity in demonstrators.

Based on the findings from these four papers, six managerial prescriptions are developed and presented, covering both technology management and IT infrastructure concerns. Overall, the dissertation establishes that a fast assessment of the process innovation potential of analytics is possible, given the proper IT infrastructure and a manageable level of analytics complexity. A key implication for operations and innovation managers is that IT infrastructure aspects of process innovation with analytics should be considered early on and influence both prioritization and organization of initiatives. On the other hand, IT management should focus on developing an IT infrastructure that facilitates loosely coupled development and deployment of analytics systems.

DANSK RESUME

Hvordan kan produktionsvirksomheder forbedre deres operationelle processer ved hjælp af analytics? Forbedring af proces performance har været et fokusområde i forskning og praksis i mange år. I dag forventer produktionsvirksomheder i stigende grad at digitale teknologier, som analytics, vil muliggøre en ny generation af performanceforbedringer. Disse forbedringer kan opnås ved at redesigne operationelle processer således, at de udnytter analytics evne til at forbedre informationsgrundlaget for eller automatisere operationelle aktiviteter – hvad jeg kalder for *procesinnovation med analytics*.

Implementering og værdiskabelse ved brug af analytics har dog vist sig at være udfordrende. Analytics er en generisk teknologi, så produktionsvirksomheder må investere i en betragtelig udforskning af teknologien for at finde frem til og vurdere potentialet i forskellige anvendelsesområder. Da analytics stadig sjældent tilbydes som et færdigt produkt eller service, kræver det en vis grad af analytics udvikling at vurdere potentialet for et givet anvendelsesområde. Analytics udviklingsprojekter er dog kendt for høje fejlrater og en langsom modningsproces fra proof-of-concept stadiet til pilotanvendelse.

Denne afhandling er praktisk motiveret af de implementeringsvanskeligheder som produktionsvirksomheder møder, og akademisk motiveret af manglen på empirisk forskning omhandlende procesinnovation med analytics. Forskningen har været udført i tæt samarbejde med industrien og har taget udgangspunkt i en konkret praktisk problemstilling i en stor dansk produktionsvirksomhed: Hvordan kan produktionsvirksomheder hurtigt udforske og afdække potentialet i procesinnovation med analytics? Afhandlingen sigter dermed efter at 1) forbedre forståelsen af procesinnovation med analytics, og 2) udvikle præskriptiv viden for procesinnovation med analytics.

Forskningen anvendte Action Design Research (ADR) som metode. Action Design Research er velegnet til at udvikle præskriptiv viden og forståelse gennem design- og evalueringsaktiviteter, som udføres i en organisatorisk kontekst. Design objektet har været en *analytics demonstrator*, som består af eksplorativ udvikling af et analytics system med det formål at opnå indsigt i dets potentiale for procesinnovation. Mere specifikt bestod de empiriske forskningsaktiviteter af 1) deltagende observation i tre analytics initiativer og 2) udviklingen af fem analytics demonstratorer. I en eksplorativ fase blev den deltagende observation og de fire første demonstratorer brugt til at udvikle en fremgangsmåde for analytics demonstrator udvikling og dertilhørende designprincipper. Dette blev efterfulgt af en evalueringsfase, hvor den femte demonstrator blev anvendt til at instantiere, evaluere, og til sidst modificere fremgangsmåden og designprincipperne. Endeligt blev ledelsesmæssige forslag udviklet på baggrund af læringerne fra forskningsprocessen.

Afhandlingen er organiseret i to dele. I den første del motiverer jeg forskningen, etablerer relevant baggrundsviden, og præsenterer et sammendrag og diskussion af forskningsprocessen samt dens resultater. Den anden del indeholder de fire videnskabelige artikler, som udgør det primære forskningsbidrag. De fire artikler adresserer de følgende forskningsspørgsmål:

- Hvilke forskningsmæssige bidrag kan udvikles gennem hele ADR forskningsprocessen? (Artikel 1)
- Hvordan skal procesinnovation med analytics konceptualiseres? (Artikel 2)
- Hvad driver udviklingshastighed i demonstratorer fokuseret på maskinlærings-baseret procesinnovation? (Artikel 3)
- Hvordan skal analytics demonstratorer udvikles? (Artikel 4)

Artikel 1 spillede en vigtig rolle i at etablere forskningsdesignet igennem en konceptuel undersøgelse af potentialet for forskningsmæssige bidrag i ADR. Denne undersøgelse fremhævede potentialet i det rige empiriske materiale, som genereres i ADR, hvilket kan være grundlag for casestudier og teoretisering. Dette potentiale for teoretisering blev udnyttet i Artikel 2, hvor der udvikles en konceptualisering af procesinnovation med analytics samt en tilhørende forskningsagenda. Denne konceptualisering fremhæver, at procesinnovation med analytics er en transformationsproces, som kræver koordineret udvikling af analytics, processen, og IT-infrastrukturen. Et multipelt casestudie af analytics demonstratorerne præsenteres i Artikel 3 med det formål at identificere faktorer, som bidrager til høj udviklingshastighed. Casestudiet fremhævede vigtigheden af to faktorer for at opnå høj udviklingshastighed: 1) løs kobling mellem demonstratoren og den eksisterende IT-infrastruktur, og 2) brugen af *byggeklodser* i udviklingsprocessen. Med udgangspunkt i dette, udvikler og formaliserer Artikel 4 en fremgangsmåde og designprincipper for hurtig udvikling af analytics demonstratorer, som derefter evalueres i en sidste demonstrator. Fremgangsmåden er baseret på inkrementel udvikling og strategier til at reducere analytics og infrastruktur kompleksitet i demonstratorer.

Baseret på resultaterne fra disse fire artikler udvikles der seks ledelsesmæssige forslag, som omhandler teknologiledelse og IT-infrastruktur. Overordnet set etablerer afhandlingen muligheden for hurtig afdækning af potentialet for procesinnovation med analytics, givet den rigtige IT-infrastruktur og en håndterbar analytics kompleksitet. En vigtig implikation for drifts- og innovationsledere er, at IT-infrastruktur aspektet i procesinnovation med analytics bør vurderes tidligt i innovationsprocessen og have indflydelse på både prioriteringen og organiseringen af initiativer. IT-ledelsen bør derimod fokusere på at udvikle en IT-infrastruktur, som muliggør løs koblet udvikling og udrulning af analytics.

ACKNOWLEDGEMENTS

I have been very fortunate to receive substantial support throughout the sometimes arduous and lonesome journey that undertaking a PhD study can be. I want to say a big thank you to everyone who has supported me along the way.

First, I would like to thank my colleagues at Center for Industrial Production at Aalborg University for contributing to an excellent research environment where it is possible to do research that makes a difference in practice. I have met nothing but support from all my colleagues and have enjoyed our many talks about research and everything beyond. I am particularly grateful for the support from my supervisor, Charles Møller. The IS field and design science research were new to me, and I have learned a lot from our many discussions during the PhD.

Second, I want to express my gratitude to the industrial partner for supporting me and allowing me to pursue what I thought was interesting and relevant. Special thanks go to Morten Lundsgaard Degn for initiating the project, engaging in thoughtful discussions, and providing organizational support whenever needed. I would also like to thank Georg Ørnskov Rønsch for bringing great energy into our collaborative projects and paving the way for the use of data science. My work would undoubtedly have been harder without the significant work he had already undertaken. I would also like to thank all the colleagues from Operations IT I collaborated with along the way. Thanks, in particular, to Nikolai Svalebæk, Ronald Dahm Larsen, Paul Wittig, Kasper Johnsen, Afonso Taborda, and Sigurdur Magnusson.

Third, I would like to thank the hosts of my external research stay at the Department of Informatics at the University of Oslo. Thanks to Professor Bendik Bygstad and Egil Øvrelid for hosting me and offering to collaborate with me on my research. The stay was highly rewarding intellectually, and I truly enjoyed it. I am grateful for having had the opportunity to learn about the "craft" of qualitative IS research in the Scandinavian tradition from someone capable of doing both practically relevant and theoretically significant research.

Finally, I would like to express my immense gratitude to my family. They supported me throughout this journey without doubting my abilities and always encouraged me. A big thanks goes out to my parents for their relentless support. Last but not least, an enormous thanks to my girlfriend Nanna, who has stood by my side all the way. I will remain forever grateful for your unquestioning support, including listening to all my complaints about the weird beast that academia can sometimes be.

Aalborg, October 2023

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDED PAPERS

Paper 1

Bojer, Casper Solheim, and Møller, Charles. (2022). *Towards a Scheme for Contribution in Action Design Research*. In International Conference on Design Science Research in Information Systems and Technology (pp. 376-387). Cham: Springer International Publishing.

Paper 2

Bojer, Casper Solheim, and Møller, Charles. (2023a). *Conceptualizing Process Innovation with Analytics: A Pragmatic Framework and Research Agenda* [Manuscript submitted for publication].

The paper has been submitted to *Business & Information Systems Engineering* and is currently under review. An earlier version of the paper was presented at the Pre-ICIS SIGDSA Symposium in Copenhagen, December 2022.

Paper 3

Bojer, Casper Solheim, Bygstad, Bendik, and Øvrelid, Egil. (2023). *Speeding up Explorative BPM with Lightweight IT: The Case of Machine Learning* [Manuscript submitted for publication].

The paper has been submitted to *Information Systems Frontiers* and is currently undergoing revisions before a second round of review.

Paper 4

Bojer, Casper Solheim, and Møller, Charles. (2023b). *Developing Analytics Demonstrators for Process Innovation: An Infrastructural Perspective* [Manuscript submitted for publication].

The paper has been submitted to *European Journal of Information Systems* and is currently under review.

CHAPTER 1. INTRODUCTION

Improving operational performance has long been a subject of interest to manufacturing firms and academics interested in operations management. Various philosophies and approaches for improving operational performance have been developed, such as Total Quality Management, Lean, and Six Sigma. In recent years, manufacturing organizations are increasingly looking towards the implementation of digital technologies as the new means for performance improvement. Under headers such as Industry 4.0 (Lasi et al., 2014) and Smart Production (Møller et al., 2023), visions of future cyber-physical production systems relying extensively on digital technologies and resulting in considerable productivity improvements have been put forward. The digital technologies involved vary in nature from cyber security, cloud computing, industrial internet of things, big data and analytics, system integration, autonomous robots, additive manufacturing, augmented reality, and simulation (Boston Consulting Group, 2015).

Analytics, one of these technologies, is concerned with using data and quantitative methods to drive decision-making and action (Davenport & Harris, 2017). The use of analytics in operations is, however, anything but new. Quantitative methods have a long history in operations, where they have been used to forecast demand, drive production planning and scheduling, route vehicles, monitor the quality and stability of processes, and much more. The increased interest in the use of analytics in operations is partly due to technical advances that have reduced data capture, storage, and computation costs and algorithmic advances that have increased the performance of analytics on various tasks. As a result, new use cases have become both technically and economically feasible. Research has shown that analytics can impact process performance by increasing productivity, quality, and speed (Tarafdar et al., 2017; Enholm et al., 2021). These benefits of analytics can be achieved through two different mechanisms: 1) the generation of insights, leading to better process management and actions, and 2) process innovation, where analytics enables change to the process. This dissertation focuses on the second mechanism, namely using analytics to innovate processes, achieved through augmentation and automation of tasks.

Despite the potential of analytics, implementing and realizing operational performance improvements has proven challenging (Davenport & Malone, 2021). Many organizations struggle in their adoption of analytics and outside high-tech front runners such as Facebook, Amazon, and Google, few have succeeded in widespread adoption. In a Boston Consulting Group and MIT CISR survey from 2017, 50% of respondents reported not having adopted advanced analytics, while 25% reported not seeing significant benefits from their adoption (Tarafdar et al., 2019). A McKinsey survey from the same year found that less than 20% of respondents had adopted analytics at scale in their organizations (McKinsey, 2018). More recent surveys paint a slightly brighter picture regarding adoption, although widespread adoption and benefits remain missing for many organizations. McKinsey found that half of the respondents' organizations use AI in at least one function (McKinsey 2020; McKinsey 2022), while Boston Consulting Group found that only 10% of companies reported significant benefits (Ransbotham et al., 2020). For manufacturing in particular, a recent survey suggests that only around one out of six adopters of AI have succeeded in meeting their expectations (Boston Consulting Group, 2023).

One reason behind this adoption challenge seems to be the scale of change required. Research has shown that realizing value from analytics requires a coordinated change of organizational and technical systems (Dremel et al., 2017; Dremel et al., 2020; Tim et al., 2020). As if implementing analytics in general was not difficult enough, using analytics to innovate processes requires even greater change in the form of development of organizational IT infrastructures and adjustments to processes and organizational systems (Davenport & Miller, 2022). Another aspect complicating the adoption of analytics for process innovation is that analytical technologies are general-purpose technologies that were not built for a specific domain, such as manufacturing. As a result, technology adoption requires considerable exploration activities to identify, assess, and potentially implement use cases (Eley & Lyytinen, 2022; Maghazei et al., 2022). For analytics, this exploration requires a technology development process comprised of people with a deep understanding of the existing processes and IT in addition to technical expertise in analytics (Tarafdar et al., 2019).

This thesis investigates how manufacturing firms can implement analytics to innovate operational processes and improve performance. It focuses on the assessment stage of the technology adoption process, where use cases are translated from the abstract (e.g., use machine learning to predict quality), to the concrete and assessed through the development of analytics system prototypes and pilots. The research is motivated by a problem that is ultimately practical and has the dual goal of contributing to research and practice. In the following section, I elaborate on the context of the research and the concrete practical problem faced by a large global manufacturer and retailer based in Denmark that motivated the research.

1.1. RESEARCH CONTEXT AND PRACTICAL MOTIVATION

The research presented in this thesis is conducted in close collaboration with the Danish manufacturing industry. The research is part of a larger initiative by the organization Manufacturing Academy of Denmark (MADE)[1] to improve the competitiveness of the Danish manufacturing industry by bringing together researchers from Danish universities, manufacturing companies, research and technology organizations (RTO), and approved technology service organizations (GTS) in applied research projects.

---

[1] https://www.made.dk/en/

This research is part of and funded by the third MADE research platform, MADE FAST[2], which focuses on developing and testing technologies and methods that improve flexibility, agility, sustainability, and talent development in industry. MADE FAST brings together five universities, three GTS, four RTOs, and 50 industrial companies that partly co-fund the research. The platform consists of five workstreams, each addressing one specific strategic area of the platform, as illustrated in Figure 1.1. Each workstream contains multiple research projects that are organized as a collaboration between 1) one or more Danish manufacturing organizations, 2) one or more Universities, and 3) potentially one or more RTO or GTS. The research presented in this thesis is the outcome of one such research project in the workstream *Value chain execution and optimization*[3]. The name of the workstream hints at its overall objective, but more specifically, it aims to address the improvement of value chains by enabling data-driven decision-making. This is to be accomplished by 1) developing platforms that enable data integration and 2) developing methods, tools, and models for optimizing value chains based on data. The research projects are to take the shape of development projects that are tested in industry and address real problems of the participating manufacturing companies while also contributing to the academic knowledge base in one or more domains.

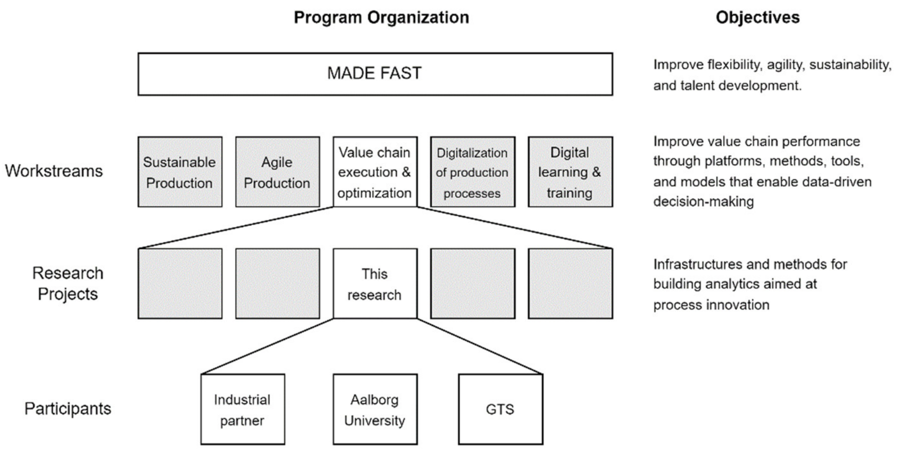


Figure 1.1 Illustration of the organization and objectives of the MADE FAST research platform

The focal research project was a collaboration between a large global manufacturer and retailer based in Denmark (henceforth referred to as *the industrial partner*) and Center for Industrial Production at Aalborg University. The industrial partner was in the process of a digital transformation journey. As part of this digital transformation,

[2] https://www.made.dk/en/made-fast/

[3] https://www.made.dk/en/made-fast/value-chain-optimization/

the industrial partner aimed to increase the extent of data-driven decision-making in operations. In the operations strategy, analytics was one of several technologies under the Industry 4.0 umbrella that would contribute to realizing productivity improvements in operations.

The main stakeholder in the research project was the Vice President (VP) of the Operations IT department at the industrial partner. In charge of IT for the operations function, his department was a key contributor to the development of the analytical technology necessary to support data-driven decision-making. At the start of the research project in the summer of 2020, the industrial partner was still at the exploration stage of adoption. Several analytics development initiatives had been carried out with varying success. While advanced analytics had been successfully adopted in a few isolated areas, it had yet to reach wide-scale adoption in operations. Foreseeing an increasing demand for the development of analytics technologies, the VP of Operations IT entered the project to obtain knowledge on 1) how to organize for fast development of analytics solutions and 2) where the use of analytics would provide value in operations. The practical problem to be solved was framed in a project charter based on an agreement between the industrial partner, stakeholders from MADE, the workstream leader (my supervisor Charles Møller), and I as follows:

- Develop an approach for fast and agile development of analytics solutions in operations
- Implement three to four analytics demonstrators in operations

Approach is used here in the broad sense and encompasses different types of prescriptive knowledge related to how to organize development, including methods, tools, infrastructure, roles, and responsibilities. Demonstrator refers here to the development of a working prototype to address a specific use-case and obtain insights into its utility as a solution. Having elaborated on the practical problem that motivated the research, I now turn to reviewing what existing research has to offer on the development of analytics solutions for improving operational processes.

1.2. BACKGROUND & STATE-OF-THE-ART

What does existing research have to offer when it comes to conducting process innovation with analytics in manufacturing? In this section, I establish the necessary background to discuss my findings and review the state-of-the-art. First, I cover different approaches to process innovation in operations, including the currently popular Industry 4.0 approach, which relies on analytics as one of the technological means for process innovation. I then move on to analytics, where I introduce the concept and present insights from both organizational and technical literature on how to develop and implement analytical systems and use them to derive organizational value. Given that development and implementation of analytics relies on IT, I then present findings from IS and information management literature on the role of IT in digital process innovation and cover different ways IT can be organized to contribute to innovation. Finally, I briefly summarize the main takeaways from the reviewed literature.

In Table 1.1, I provide definitions for key concepts used in the thesis as a reference point. As will be clear from the review below, process innovation with analytics is a complex and multi-faceted phenomenon that can be studied from many angles. The focus in my research has been the intersection of analytics, process innovation, and digital infrastructure, as visualized in Figure 1.2.

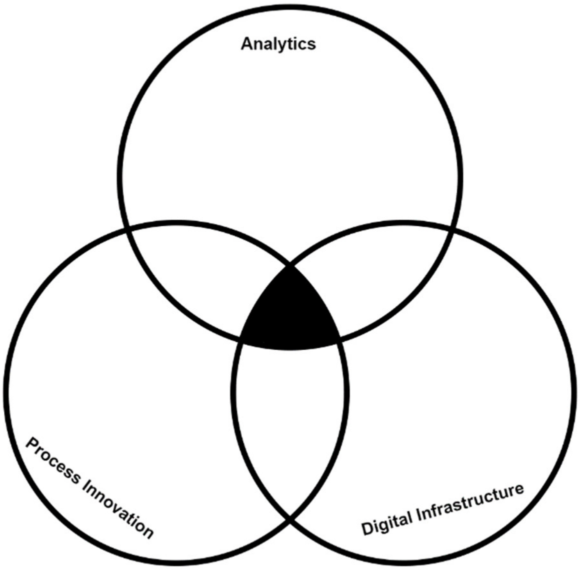


Figure 1.2 Illustration of the research focus

Table 1.1 Main concepts used and their definitions

Concept	**Definition**	**Reference(s)**
Analytics	"The extensive use of data, statistical and quantitative analysis, explanatory and predictive models, and fact-based management to drive decisions and actions. The analytics may be input for human decisions or may drive fully automated decisions."	Davenport & Harris (2017)
Process Innovation	"[A] new or significantly improved production or delivery method. This includes significant changes in techniques, equipment and/or software"	OECD/Eurostat (2005)
Lightweight IT	"a socio-technical knowledge regime, driven by competent users' need for solutions, enabled by the consumerization of digital technology, and realized through innovation processes."	Bygstad (2017)
Heavyweight IT	"a socio-technical knowledge regime, driven by IT professionals, enabled by systematic specification and proven digital technology, and realized through software engineering."	Bygstad (2017); Bygstad & Øvrelid (2020)
Digital Infrastructure	"a shared, open (and unbounded) heterogenous and evolving socio-technical system […] consisting of a set of IT capabilities and their user, operations, and design communities"	Hanseth & Lyytinen (2010)
Boundary Resource	"the software tools and regulations that serve as the interface for the […] relationship between the platform owner and the application developer"	Ghazawneh & Henfridsson (2013)

1.2.1. PROCESS INNOVATION IN OPERATIONS

Changing processes to improve performance is a key part of operations management (Slack et al., 2013). The change is often referred to as either process improvement or process innovation, based on the degree and nature of change. Many different definitions of process innovation exist both within and across fields. In the thesis, I adopt the OECD definition of process innovation as a "new or significantly improved production or delivery method. This includes significant changes in techniques, equipment and/or software" (OECD/Eurostat, 2005). In this definition, changes to the execution of processes would not be considered process innovation. Instead, the process must change or leverage new technology, such as an analytics system, to be considered a process innovation. Compared to the definition of Davenport (1993), the definition does not limit process innovation to the scope of cross-functional business processes and can include innovations at lower levels.

One way of distinguishing different approaches to process innovation is to consider whether they adopt an explorative or exploitative orientation (Rosemann, 2014). Exploitation is concerned with improving the efficiency of how existing work is carried out – *doing what you already do, but better*. Exploration is concerned with looking to the environment for opportunities to change how work is carried out for the better -– *doing things differently and better*. Strategic management research has argued that companies need to be ambidextrous, i.e., capable of both exploration and exploitation at the same time, to survive and prosper (Tushman & O'Reilly, 1996; O'Reilly & Tushman, 2013).

Older approaches to process innovation with roots in industrial engineering are primarily oriented towards exploitation by removing waste or variability in existing processes. In its early days, Industrial Engineering and scientific management were concerned with measurement and simplification of work to design better ways of working that improved productivity (Harmon, 2010). Inspired by the Toyota Production System (Womack et al., 1990), Lean emphasized continuous employee-driven improvement, a focus on flow rather than resource efficiency, and the removal of waste and variation from processes (Hopp & Spearman, 2021). Whereas some techniques of lean, such as value stream mapping, can include larger-scale change, Lean has generally focused on incremental improvements and eschewed the use of complex technology. A similar internal focus in found in Six Sigma, with its focus on improving quality through the removal of variation in processes (Rosemann, 2014).

Newer approaches based on harnessing the potential of new technology, particularly IT, have had a more explorative orientation. The cleanest example of an approach with an explorative orientation was Business Process Reengineering (BPR). BPR focused on leveraging innovations in IT to radically redesign processes, starting from a clean slate (Davenport, 1993; Hammer & Champy, 1993). Key to BPR implementation was considering IT capabilities in redesigning processes and using the redesigned process to drive IT implementation (Davenport, 1993; Hammer & Champy, 1993). Owing to the uncertainties involved in large-scale redesign, prototyping the redesigned process was recommended as a best practice to facilitate learning and validate its performance (Davenport & Short, 1990; Davenport, 1993). Despite the enthusiasm around BPR in the 1990s, it has since fallen out of favor due in part to difficulties in implementing and realizing the radical change envisioned. Instead, management of processes has moved towards a more balanced approach in the form of a portfolio of incremental and radical projects (Davenport, 1995). Along these lines, Business process management (BPM) has emerged as a balanced and holistic approach to managing and improving business processes. BPM is concerned with the ongoing management of business processes throughout their lifecycle of 1) discovery, 2) analysis, 3) redesign, 4) implementation, and 5) monitoring and controlling (Dumas et al., 2013). Whereas BPM covers both radical and incremental change, in practice, BPM has often focused on incremental innovation at the cost of more radical innovation (Benner & Tushman, 2003; Rosemann, 2014). Recent research is, however, working to expand the scope of BPM to exploration under the header of explorative BPM (Rosemann, 2014; Grisold et al., 2019). This stream of research is in its early stages, and there is thus still a need for knowledge on how to successfully implement exploratory process innovation (Baier et al., 2022). The research that does exist suggests use of lightweight engineering processes for process innovation rather than traditional heavier engineering and development (Schmiedel & vom Brocke, 2015; Bygstad & Øvrelid, 2020; Baier et al., 2022). Similarly, Baiyere et al. (2020) emphasize that process modeling is less important in the digital transformation of processes and argues for increased flexibility from IT infrastructure and actors. Furthermore, Baier et al. (2022) identify several new implementation success factors that suggest the importance of partner involvement, digital ambition and attitude, data analysis, and infrastructure readiness in digital process innovation.

In the manufacturing context, Industry 4.0 has emerged as a vision for process innovation that aims to take advantage of new technology (Lasi et al., 2014). To implement this vision in practice, many manufacturers have set up Industry 4.0 as a process improvement or transformation program. Industry 4.0 is similar to BPR and explorative BPM in that it takes outset in exploring the potential of new technologies to change processes for the better. There are, however, also important differences owing in part to the manufacturing context of Industry 4.0. First, compared to BPR, it is recognized that the existing processes and technology must be considered instead of starting from scratch (Kagermann et al., 2013, p. 26). Many Industry 4.0 initiatives take place in brownfield environments characterized by legacy equipment, as investment in greenfield setups is often infeasible from an economic point of view. Dealing with legacy equipment necessitates investments in connectivity and retrofitting equipment with sensors to enable digital access and data capture (Tran et al., 2022). Second, manufacturing processes are also often less changeable than business processes in general, as their structure is largely embedded not only in IT systems but also in hardware, i.e., manufacturing technology.

Whereas BPR offered several methodologies for carrying out exploratory process innovation, empirical work on Industry 4.0 is still in the early stages, with little guidance on how to organize and carry out the innovation process. Research on Industry 4.0 implementation has so far examined aspects such as tensions and resolution strategies (Dieste et al., 2022), building of dynamic capabilities (Chirumalla, 2021), implementation patterns (Frank et al., 2019), critical success factors (Hoyer et al., 2020), and assessed I4.0 on the innovation characteristics of Rogers (1995) (Tortorella et al., 2021). These studies highlight the important role of IT infrastructure in adoption and implementation (Hoyer et al., 2020; Chirumalla, 2021; Tortorella et al., 2021; Dieste et al., 2022). They further suggest the importance of knowledge and skills (Hoyer et al., 2020; Chirumalla, 2021; Dieste et al., 2022) and culture (Chirumalla et al., 2021; Dieste et al., 2022). In terms of adoption, Frank et al. (2019) find a sequential implementation pattern of increasing complexity, where organizations first adopt cloud, followed by IoT, then big data, before finally adopting analytics, resulting in use cases starting with monitoring before turning towards AI.

Zooming in on the innovation process in Industry 4.0, recent research has conceptualized the process as identifying, piloting, and scaling use cases (Maghazei et al., 2022), owing to the exploratory nature of the innovation process (Maghazei et al., 2022; Eley & Lyytinen, 2022). Rather than a linear process of technology adoption that starts with a business case, experimenting and piloting are necessary to validate use cases before a business case can be developed (Maghazei et al., 2022). Chirumalla (2021) suggests that agile cross-functional teams, a continuous digital improvement strategy, and a common data layer should be leveraged in the *seizing* stage of innovation, which roughly corresponds to the *piloting* phase of Maghazei et al. (2022).

Most existing research considers Industry 4.0 at the overall level and does not address technology-specific differences and their implications for the innovation process. There are a few notable exceptions that tackle analytics specifically. These studies have provided insights into analytics use cases in manufacturing. Lorenz et al. (2022) suggest that advanced analytics can be used for production planning and scheduling, quality inspection, quality driver identification, process control, and maintenance planning, while Gröger (2022) identifies predictive maintenance, predictive quality, and engineering in the loop, where usage data is used to improve product design, as key use-cases of analytics in manufacturing. Studies have also examined the technical implementation of infrastructure for Industry 4.0 analytics. Gröger (2018; 2021; 2022) reports on experiences from Industry 4.0 analytics implementation at Bosch. A key contribution includes an architecture for an analytics platform that facilitates reuse of analytical solutions. Furthermore, he highlights the challenges associated with democratizing data science, establishing robust data management, and holistic data governance. Bonnard et al. (2021) similarly report on the implementation of an Industry 4.0 analytics platform in two Brazilian companies and demonstrate how it enabled the development of analytics use cases. While these studies provide more concrete insights into use cases and platforms for Industry 4.0 analytics, they do not address how the platforms and their design or the specific use cases influence the development and innovation process.

1.2.2. ANALYTICS & ANALYTICAL SYSTEMS

Analytics covers an amalgam of technologies and use cases that are constantly evolving. Many different definitions are on offer in literature, emphasizing different aspects from tools and technologies, over activities and transformation processes, to capabilities and culture (Holsapple, 2014). In the dissertation, I adopt the definition by Davenport and Harris (2017) of analytics as "the extensive use of data, statistical and quantitative analysis, explanatory and predictive models, and fact-based management to drive decisions and actions. The analytics may be input for human decisions or may drive fully automated decisions". This definition emphasizes that analytics is at the intersection of technology, decision-making, and quantitative methods (Mortenson et al., 2015) and can be used both to augment and automate decisions.

Owing to the broad nature of analytics as a term, several categories and terms have been developed to define and distinguish different types of analytics. A widely influential scheme from Gartner (2014b) suggests categorizing analytics into three types based on the nature of their information output: 1) descriptive analytics focuses on what has happened, 2) predictive analytics focuses on what will happen, and 3) prescriptive analytics focuses on which action to take. Another way of classifying analytics focuses on data characteristics and distinguishes between streaming vs. batch analytics, small vs. big data analytics, or the origin domain of the data, e.g., IoT or web analytics. Davenport et al. (2010) suggest that another important factor is whether analytics is embedded into business processes or used offline for analysis. Whereas embedding analytics into business processes is more complex, they also argue that the potential benefits are higher (Davenport et al., 2010).

From a technical perspective, analytics relies on IT systems that consist of several components. Whereas analytics systems can be highly complex (Sculley et al., 2015), research has suggested simpler conceptualizations consisting of a few higher-level components in the form of data, model, and system components (Schneider et al., 2023) or data, model, and action components (Davenport & Miller, 2022). These conceptualizations highlight the general nature of analytics systems in that they consume data (*data*), leverage data or analytical models to produce an output (*model)*, which is then either presented to a user in a graphic user interface or sent to another IT system (*system/action)*. Analytical models can range from simple to complex and include various quantitative models marketed under different labels, such as statistics, machine learning, data mining, and AI. It should be noted that not all AI is analytical, as some scholars also consider robots and robotic process automation as AI (e.g., Benbya et al., 2021).

Construction of an analytics system requires technology development, as out-of-the-box standard systems are rare. Until recently, analytics development has consisted of developing custom data collection and processing components, training models by customizing existing algorithms, and developing custom graphic user interfaces or system integrations. Several development methodologies exist to structure and support the development process, such as CRISP-DM (Wirth & Hipp, 2000) and various extensions (Martínez-Plumed et al., 2019; Huber et al., 2019; Studer et al., 2021), the Team Data Science Process from Microsoft (Microsoft, 2023), and the business analytics methodology (Hindle & Vidgen, 2018). In general, these methodologies emphasize iterative development with the involvement of domain experts to ensure a thorough understanding of the business problem and the development of capable models. In practice, the use of the methodologies has been mostly limited to the development of data processing and model components. Once a data scientist arrives at a useful model that business stakeholders have validated, the implied process has been a handover to the IT department that would handle deployment like any other IT project (see, e.g., the description of CRISP-DM in Wirth & Hipp, 2000). This model focus in analytics development has recently been questioned, as organizations often struggle with deployment (Davenport & Malone, 2021; Vial et al., 2021) and underestimate the technology maturation required (Lavin et al., 2022). Deployment of analytics systems often requires considerable infrastructure (Sculley et al., 2015), and recent research suggests that these aspects should be considered throughout the project rather than left for the IT department to solve at the end (Vial et al., 2021; Davenport & Malone, 2021). The nature of analytics development also seems to be changing, owing to a technology maturation with more standard services and components in the marketplace. Cloud vendors are increasingly offering analytics-as-a-service, such as anomaly detection services and chatbots, and model components that can be customized, such as pre-trained models for time series, computer vision, or natural language processing. Emerging research is calling into question the traditional analytics development methodologies in the current era of AI and starting to draw the contours of new approaches that merge aspects of design thinking, agile, big data analytics, and IT integration (Dolata et al., 2022) or agile, data science, and project management (Vial et al., 2023).

From an organizational perspective, research has investigated how value is realized from analytics. This research has highlighted the importance of a socio-technical and holistic perspective on analytics (Dremel et al., 2017; Dremel et al., 2020; Tim et al., 2020). Investing in the foundational elements, such as data, infrastructure, software, and human skills and knowledge, is not enough – the elements must be orchestrated and organized into a coherent whole to result in an analytics capability (Mikalef et al., 2018). This orchestration encompasses strategic planning, development, use, and governance of analytical assets (Mikalef et al., 2020). The concrete realization of value from analytics happens through use, where it can generate insights, or augment or automate tasks (Enholm et al., 2021; Spring et al., 2022), resulting in either insights, improved process efficiency, or process innovation (Enholm et al., 2021). The use of analytics for either insights, augmentation, or automation places different demands on capabilities. Whereas insight requires strong data science capabilities and deep domain knowledge, augmentation further requires a mature data infrastructure, while automation additionally requires integration with existing IT systems (Shollo et al., 2022). As a result, Shollo et al. (2022) found that implementation progressed sequentially from simple (i.e., insights) to more complex use cases. A similar implementation pattern was observed by Dremel et al. (2017) in their longitudinal case study of analytics at AUDI, Tim et al. (2020) in their study of analytics at Rovio, and Zhang et. (2022) in their study of analytics for air pollution management in China, underlining the generality of the implementation pattern. While this stream of research provides important insights into realizing value from analytics, few empirical studies examine process innovation uses of analytics. One notable exception is the configurational analysis of Mikalef and Krogstie (2020), which suggests that technical skills are important for incremental process innovation with analytics, whereas managerial skills, organizational learning, and culture are important for more radical process innovation.

1.2.3. THE ROLE OF IT IN DIGITAL PROCESS INNOVATION

The organization of IT in companies plays a role in its ability to conduct digital process innovation. In most organizations, responsibility for IT is centralized in the IT department, although change to this arrangement might be underway (Legner et al., 2017). Despite harboring the key IT experts and talent in most organizations, mention of the IT department is associated with anything but enthusiasm in most organizations looking to undertake IT-enabled innovation. What gives?

Since the days of BPR, the IT department has been recognized as a key player in process innovation, contributing to the identification of opportunities and to the implementation of the IT underlying new process designs (Davenport, 1993; Hammer & Champy, 1993). Nonetheless, many organizations implementing BPR ended up in situations where their IT infrastructures ended up as bottlenecks to realizing their redesigns (Stoddard & Jarvenpaa, 1995; Broadbent et al., 1999). Research since then has confirmed that IT infrastructure flexibility, IT management capability, and IT personnel expertise are important enablers of process innovation (Kim et al., 2011).

A stream of research in IS has examined in more detail how the technical and social organization of IT either enables or constrains the ability of organizations to conduct IT-enabled innovation. Using the terms *information infrastructure* or *digital infrastructure*, this stream has investigated how the complex network of infrastructure and applications in or across organizations, along with their users, developers, and operators, impacts IT-enabled innovation (Hanseth & Monteiro, 1998; Ciborra et al., 2000). Using the lens of digital infrastructure, the challenge of IT development or innovation is not arriving at a functional new system but rather developing a system that fits into the existing complex network of systems, often referred to as *the installed* *base* (Ciborra et al., 2000; Aanestad et al., 2017). Due to its complexity, the installed base of systems can inhibit change, and inadequate attention to and respect for the installed base of systems in implementation of new IT has been brought forward as a key cause of failure in development (Ciborra et al., 2000).

Given the critical role played by the digital infrastructure in innovation, how to change it to become an enabler rather than a constraint has been the topic of considerable research. One point of discussion has been the degree to which a digital infrastructure can be centrally designed or changed, for example, by corporate IT management. Whereas IT management literature historically assumed IT to be within the control of top management (e.g., Weill & Broadbent, 1998), Ciborra et al. (2000) critiqued this assumption. They argued instead for infrastructures as being outside managerial control and subject to *drift*, owing to their complexity and the resulting unintended side effects of change attempts (Ciborra et al., 2000). Later work seems to have arrived at a (more reasonable) middle ground, where the digital infrastructure is partly controllable (Tilson et al., 2010; Rolland et al., 2015; Törmer & Henningsson, 2018; Koutsikouri et al., 2018). In addition to this tension between planned and emergent change, research has further proposed short vs. long-term and local vs. global as key dilemmas and tensions facing attempts to change infrastructures (Edwards et al., 2007). In terms of how to change infrastructure, two overall approaches to developing infrastructure are proposed in the literature, which are linked with the different views on the control issue. Taking outset in the assumption of managerial control, the enterprise architecture approach proposes centralized management of digital infrastructure based on mapping as-is and to-be states and designing incremental transition strategies realized through development projects (Ross et al., 2006). Skeptical of the assumption of central control, the digital infrastructure approach instead proposes more bottom-up decentralized development of infrastructures (Ciborra et al., 2000). This research has suggested design principles for development that respect the installed base, favoring the reuse of existing elements and small-step incremental (Hanseth & Lyytinen, 2010; Grisot et al., 2014) and modular change (Aanestad & Jensen, 2011).

Research has also examined what makes for an enabling digital infrastructure. The architecture and governance arrangements of a digital infrastructure have been identified as key aspects that influence the evolution of a digital infrastructure and its ability to enable or impede innovation (Bygstad, 2010; Henfridsson & Bygstad, 2013; Grisot et al., 2014; Bygstad & Øvrelid, 2020; Hanseth & Modol, 2021). This research has highlighted the importance of a flexible infrastructure (Bygstad, 2010; Grisot et al., 2014) and a loosely coupled architecture and decentralized control to enable innovation (Henfridsson & Bygstad, 2013; Bygstad & Øvrelid, 2020). In terms of concrete designs of digital infrastructures exhibiting these characteristics, two different and related approaches have received considerable interest in both research and practice: 1) bimodal IT and 2) platforms.

With bimodal IT, IT for operational execution and IT for customer-facing innovation are separated structurally into different units with different governance arrangements (Gartner, 2014a; Haffke et al., 2017; Ross et al., 2019). It is thus an attempt to achieve ambidexterity through structural separation. Bimodal IT has been a response to the often-missing ability of the IT organization to enable exploratory innovation. Research has similarly suggested distinguishing between *heavyweight IT* and *lightweight IT* and organizing them separately (Bygstad, 2017) (see Table 1.1 for definitions). Heavyweight IT is focused on systematic engineering and proven technologies, which, while resulting in stable and secure solutions, is not conducive to rapid exploration and innovation (Bygstad, 2017). This is instead the domain of lightweight IT, which focuses on innovation, experimentation, and fast development using relatively standard digital technologies in collaboration with users (Bygstad, 2017). Owing to the different natures of lightweight and heavyweight IT, research recommends that they be organized separately and loosely coupled to prevent heavyweight IT from slowing down innovation (Bygstad, 2017). Loose coupling can, however, be difficult to achieve in practice owing to the silo-oriented nature of many IT applications (Bygstad & Hanseth, 2018).

Another popular and related response to the demand for innovation has been organizing digital infrastructures according to the logic of platforms. Digital platforms are based on a stable core of functionality, subject to control by the platform owner, and a variable periphery of applications, often developed by third parties (de Reuver et al., 2018). The peripheral applications rely on the platform core for functionality, which it accesses through well-defined interfaces or *boundary resources* (Ghazawneh & Henfridsson, 2013). Whereas initial interest in digital platforms was on externally oriented platforms focused on drawing on the innovation capabilities of external actors, research has recently focused on platforms as an organizing model for corporate digital infrastructures to enable innovation (Bygstad & Hanseth, 2018; Törmer, 2018; Vestues & Rolland, 2021). In this arrangement, heavyweight IT systems in the digital infrastructure are turned into platforms by the development of boundary resources, such as APIs or service buses that facilitate access to data and IT functionality in the core (Bygstad & Hanseth, 2018; Bygstad & Øvrelid, 2020). These boundary resources, on the one hand, allow for governance and control of the platform core, while simultaneously providing loosely coupled access to its functionality for lightweight IT innovation initiatives (Bygstad & Øvrelid, 2020).

1.2.4. SUMMARY

To summarize, considerable research has examined 1) how to innovate operational processes, 2) how to develop and implement analytics, and 3) how to introduce new IT and change digital infrastructures, all of which are important aspects in the implementation of process innovation with analytics in manufacturing. In terms of innovation, research has highlighted the need for lightweight iterative development and experimentation and the use of prototyping and piloting to assess the affordances of analytics for process innovation. For this to succeed, however, requires an enabling digital infrastructure. More technical research points to the importance of an analytics platform, while IS research highlights the importance of loosely coupled access to the data and functionality of the IT application landscape. Implementing analytics for process innovation or changing the enabling infrastructure is, however, challenging due to the inertial effects of the installed base of systems. The challenge is further exacerbated in the manufacturing domain, which not only has to deal with legacy IT but also legacy equipment and processes that have a lower changeability due to their physical nature. To tackle this complex challenge, research recommends adopting a socio-technical perspective, starting with simpler problems and adopting an incremental change strategy.

1.3. RESEARCH OBJECTIVES

As evidenced by the review of the state-of-the-art, analytics has the potential to significantly improve process performance through the redesign of processes, but adoption is proving to be challenging. Existing research on analytics, digital process innovation, and digital infrastructure all provide partial insights into the challenges facing organizations. However, there is a scarcity of research that specifically addresses the use of analytics for process innovation. The few studies that do are mainly organizational-level studies or concerned with the impacts of analytics for process innovation rather than its process and content, i.e., how it is carried out.

The research therefore sets out to address two different objectives: 1) improving understanding of process innovation with analytics, and 2) developing prescriptive knowledge for process innovation with analytics. Table 1.2 provides a description of how the thesis will contribute more concretely to the two research objectives by achieving four sub-objectives.

1.4. THESIS STRUCTURE

The thesis is organized into six chapters, along with an appendix containing the four appended research papers. The thesis presents a more comprehensive description and discussion of the research process and outcomes as compared to the appended papers, including how they relate and contribute to the research objectives.

The first chapter introduced the overall research context and problem. It introduced the empirical phenomenon of process innovation with analytics and elaborated on the research context and the practical challenge that motivated the research. It then proceeded to present an overview of the state-of-the-art in the areas of analytics, process innovation in operations, and digital infrastructure that combine to make up the academic knowledge bases that the research builds upon. Based on the state-of-the-art and the practical problem, the research objectives were finally presented.

Table 1.2 Description of research objectives

Research Objective	Sub-objective	Motivation
Improving understanding of process innovation with analytics (Objective 1)	Developing a framework to support engagement in process innovation with analytics	Analytics research mainly covers analytics aimed at insight rather than operational process innovation, and thus a broader conceptualization of analytics is needed
	Understanding the drivers of development speed and complexity in analytics aimed at process innovation	Organizations are struggling to realize value from advanced analytics and finding that it is a long and complex journey. Understanding the drivers of development speed is necessary to engineer an approach that achieves high levels of speed.
Developing prescriptive knowledge for process innovation with analytics (Objective 2)	Developing an approach and design principles for fast development of analytics demonstrators aimed at process innovation	The practical problem of the industrial partner that initiated the research. The addition of design principles ensures the presence of an academic contribution.
	Developing managerial prescriptions for process innovation with analytics	Contribute to increasing the success rate of analytics initiatives in the manufacturing industry.

The second chapter presents the research design that was employed to meet the research objectives. First, I present a clarification of my research philosophical position. Second, I present the methodology employed. I start by elaborating on action-oriented design science research, which was the overall methodology adopted in the research. I then introduce Action Design Research as the specific genre of action-oriented design science research I adopted to achieve the dual goals of practical relevance and academic knowledge contribution. Finally, I describe in more detail how I applied and adapted ADR in my research design to deliver on the research objectives through a combination of engaged design work (the topic of Chapter 3) and theorizing (the topic of Chapter 4 and section 5.1).

The third chapter provides an overview of the empirical work conducted as part of the research. First, I briefly introduce the empirical context, focusing on the organization of IT-enabled process innovation at the industrial partner, to provide further background and insights into the context. I then present the analytics initiatives I took part in throughout the research, which is followed by details on how I used participative observation and design and action to gather empirical data and build artefacts.

The fourth chapter contains a summary of the findings of the appended papers that make up the main scientific contribution of the thesis. The overall purpose and contribution of each paper and their linkages are briefly discussed and followed by a summary of each paper.

The fifth chapter provides a discussion of the research. First, based on my research findings, I develop managerial prescriptions for how manufacturers should engage in process innovation with analytics, covering both innovation and IT infrastructure concerns. I then move on to a discussion of research quality before finally discussing research contributions and implications.

The sixth and final chapter contains a conclusion of the thesis by summarizing the main contributions of the thesis and commenting on its relevance and significance for research and the Danish manufacturing industry. This is followed by a brief highlight of the main limitations of the research and suggestions for further research.

CHAPTER 2. RESEARCH DESIGN

2.1. RESEARCH PHILOSOPHICAL POSITION

Researchers can adopt different philosophical positions in their research. Each philosophical position is a particular view of the world and how one can acquire knowledge about it. The choice of a particular philosophical position has implications for how research problems are conceptualized and which methods are deemed appropriate to create what research is ultimately about – new knowledge. The research philosophical positions on offer often differ in terms of their ontological (i.e., what the world is like), epistemological (i.e., how we can acquire knowledge about the world), and axiological assumptions (i.e., the aims or values underlying research). In terms of ontology, a key assumption is whether there is a *real* world that is independent of our conception of it. Regarding epistemology, a key assumption is whether knowledge can be acquired by objective means or whether it is ultimately a subjective process. For axiology, a key question is whether the goal of research is to adopt a value-neutral stance and describe the world as it is or to change the world for the better. In case the latter stance is taken, an additional assumption concerns what *better* means – *better* for who?

This dissertation adopts pragmatism as the overarching philosophical position. Developed initially by Charles Sanders Peirce, pragmatism is concerned primarily with creating knowledge that is *useful* in guiding action to achieve certain aims (Goldkuhl, 2012; Chang, 2022). It thus differs from both positivism, with its orientation towards generating explanations and uncovering general laws, and interpretivism, with its orientation towards increasing understanding of specific contexts (Goldkuhl, 2012). Pragmatism has been applied in many disciplines that are concerned with understanding and improving practice, including operations management (Boer et al., 2015), information systems (Ågerfalk, 2010; Goldkuhl, 2012), and organization studies (Farjoun et al., 2015). The choice fell on pragmatism as it aligns well with the research objective of creating knowledge that is useful for Danish manufacturing companies and my inclination towards research that makes a difference in practice.

Owing to its focus on utility rather than truth, pragmatism does not presuppose a particular ontology. The overarching ontological position adopted in this thesis is realist in the sense that it subscribes to the position that 1) there exists an external reality outside our direct control, and 2) any engagement with reality is ultimately framed by our minds (Chang, 2022). The position is essentially what Godfrey-Smith (2021) terms *common-sense realism.* Following Chang (2022), I adopt a pluralistic ontological position, subscribing to the idea that 1) different ontologies can be appropriate for supporting different aims and 2) that ontology building is an important part of pragmatism: “Identifying the realities that we should be dealing with is a crucial part of any process of inquiry" (Chang, 2022, p. 129). At a lower level of abstraction, the research domain sits at the intersection of engineering and social sciences. As such, an important and widely discussed ontological question (especially in IS research, see, e.g., Sarker et al., 2019) concerns the relationship between technology and the social world. On this question, the dissertation adopts a socio-technical point-of-view where the social and technical are ontologically separate but mutually influence each other.

Pragmatism relies on a process of inquiry to construct knowledge for action and change. Inquiry is described by Goldkuhl (2012) as "[…] an investigation into some part of reality with the purpose of creating knowledge for a controlled change of this part of reality." Inquiry starts with a problem situation to be resolved and consists of iterations of action, learning, and adjustment to beliefs, capabilities, methods, and potentially even aims until the problem has been satisfactorily resolved (Chang, 2022, p. 48). In this process of learning from experience, the researcher builds up what Ågerfalk (2010) terms *knowledge-through-action,* which can take the shape of both descriptive and prescriptive knowledge (Goldkuhl, 2012).

In terms of values, the research has been undertaken with the aim of contributing to the digital transformation of the Danish manufacturing industry. As such, it does not aim to be value-neutral. The research has been conducted in collaboration with and co-sponsored by management at the industrial partner and has consequently responded to the problems perceived by managers in IT and operations at the industrial partner. Ultimately, the practical aims have been the use of IT to increase productivity and efficiency. I acknowledge that this aim is not necessarily shared by all stakeholders in the manufacturing industry in general or at the industrial partner. In particular, the digital transformation of operations stands to impact employees and the characteristics of their jobs. However, these changes have not been a focus of the research, although attempts have been made to include users of the new technology, when possible, in development efforts to allow them to at least influence the process.

2.2. METHODOLOGY

This section introduces the research methodology used to satisfy the dual goals of developing both an academic contribution and solutions with practical relevance. I first introduce the overall paradigm that I leveraged before moving onto a description of the specific method adopted and its application.

2.2.1. ACTION-ORIENTED DESIGN SCIENCE RESEARCH

The research adopted design science research (DSR) as the overall research paradigm (Hevner et al., 2004). DSR is a pragmatist research approach (Hevner, 2007) focused on the development of novel and innovative artifacts to solve problems in organizations or society while simultaneously contributing to the academic knowledge base (Hevner et al., 2004). As an approach, DSR has been conceptualized as consisting of three cycles: 1) a relevance cycle, where the researcher engages with the environment to identify important problems and test solutions, 2) a rigor cycle, where the researcher draws on the academic knowledge base in design activities and contributes with new knowledge, and 3) a design cycle, where the researcher iterates between constructing and evaluating artifacts (Hevner, 2007).

While DSR offers a high-level structure and requires the development of novel artifacts, researchers otherwise have significant leeway in the goals pursued and how DSR is carried out. This has resulted in a diversity of *genres* of DSR (Rai, 2017; Peffers et al., 2018). The genres differ, amongst other things, in terms of the researcher's engagement with practice. I adopted an action-oriented DSR approach based on close practical engagement with the goal of developing a specific solution to the problems of the industrial partner, followed by a generalization of the solution to make up the main academic knowledge contribution. This is essentially what Iivari (2015) termed *DSR Strategy 2.*

Multiple genres exist even within action-oriented DSR, including Action Design Research (ADR) (Sein et al., 2011) from the IS community and Intervention-based Research (IBR) (Oliva, 2019) originating within the OM field. Both genres share the importance of intervention in practice, but they differ in their views on the role of theory and the significance of the design activity. While IBR focuses on intervention as a means for theory development and testing (Oliva, 2019), ADR is concerned with the development of artifacts through interventions and generalization of the design knowledge in the form of design principles (Sein et al., 2011). Their differences highlight an interesting aspect of action-oriented DSR, namely, that it has the possibility to create a variety of different contributions. Rather than being limited to artifact development and outputs of design theorizing (as emphasized in, e.g., Gregor & Hevner, 2013), action-oriented DSR can furthermore result in empirical and theoretical contributions (Goldkuhl & Sjöström, 2021). I discuss the potential for various contributions further in the first appended paper (*Paper 1*). Suffice it to say, in this view, action-oriented DSR projects can essentially be considered variations of a case study, generating empirical data that are potentially useful in theorizing. I adopted this view in my research along with ADR as the specific *genre* or method for my research. I elaborate on ADR and my application of it in the following sections.

2.2.2. ACTION DESIGN RESEARCH

ADR is concerned with the development of novel artifacts in an organizational context and thus emphasizes that design and evaluation should not be conceptualized as separate activities (Sein et al., 2011). Through this process of design and evaluation, design knowledge about both the artifact and its context is developed (Sein et al., 2011). As eloquently put by the authors of ADR: "to an ADR researcher, the greater concern is gaining understanding via design" (Purao et al., 2013, p. 79). ADR, in its original form, consists of four stages conducted in an iterative fashion: 1) Problem Formulation, 2) Building, Intervention, and Evaluation (BIE), 3) Reflection and Learning, and 4) Formalization of Learning (Sein et al., 2011). A later update frames the process as two nested loops with *Formalization of Learning* running in parallel: 1) *Problem Formulation* - *BIE* – *Reflection &* Learning as the outer loop, and 2) multiple BIE cycles as the inner loop. (Sein & Rossi, 2019). The updated ADR method is visualized in Figure 2.1.

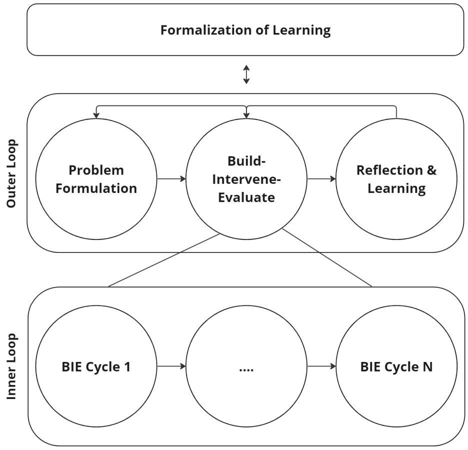


Figure 2.1 Visualization of the stages in ADR (inspired by Sein & Rossi, 2019)

In ADR, the object of design is the artifact in its context: the *ensemble artifact* (Sein et al., 2011). The scope of the ensemble artifact is thus broader than the merely technical IT artifact as it includes the organizational context of use, such as work practices and policies (Purao et al., 2013). While the possibility of design in social systems is a topic of ongoing debate (e.g., Pentland & Feldman (2008) and Beverungen (2014) for routines; Ciborra et al., (2000) and Koutsikouri et al. (2018) for infrastructures), the position taken in ADR is that the context can be designed (at least to some extent): "we argue that such ensemble artifacts can […] be designed as the whole package – tools, routines, procedures, and even policies" (Purao et al., 2013, p. 78). In addition to their broader scope, ensemble artifacts are further characterized by their emergence and the role of organizational participants in design. The ensemble artifact is the result of guided emergence: a combination of the researcher's design and interaction with the organizational context, including unintended consequences (Sein et al., 2011). Owing to this emergence, design is not solely the domain of the researcher but can include inputs from participants and existing practices (Purao et al., 2013). In the following section, I explain how I adapted ADR to deliver on the research objectives.

2.2.3. APPLICATION OF ACTION DESIGN RESEARCH

As introduced above, ADR is concerned with the development of design knowledge and understanding by building and evaluating ensemble artifacts. Application of ADR thus requires establishing:

- What the ensemble artifact is (the object of study)
- How building and evaluating the ensemble artifact allows for meeting the objectives (the research process and knowledge contribution)

I address these questions in turn below. Beforehand, it is necessary to emphasize that the research design in practice was emergent. My understanding of the ensemble artifact and the knowledge contributions made possible through the design efforts evolved throughout the research process based on reflection and learning.

The object of study

Defining the ensemble artifact requires establishing the IT artifact(s) in this research and its relevant organizational context. At a general level, the research is interested in constructing knowledge on the phenomenon of process innovation with analytics as it takes place in large manufacturing organizations by participating in the innovation process. At the concrete level, the research accomplished this by building and evaluating two sets of IT artifacts: 1) analytics systems and 2) the analytics development process. Together with a process innovation use case, these two IT artifacts make up the analytics demonstrator. The ensemble artifact is visualized in Figure 2.2 and elaborated on below.

The analytics demonstrator is an exploratory analytics development project with the goal of assessing the value and feasibility of a particular use case of analytics. It thus makes up a part of the middle activity in the modern exploratory process innovation cycle of 1) identifying use cases, 2) assessing use cases, and 3) implementing or scaling use cases (Rosemann, 2014; Maghazei et al., 2022). To assess the value of analytics, the demonstrator constructs one or more analytics system prototypes and subjects them to evaluation for feedback with users or domain experts, i.e., demonstration. The prototypes can have various levels of fidelity and range in scope from a component (e.g., an analytical model) to a working system with real and live data. The demonstrators in the research project were generally of high fidelity, consisting of a working analytics system prototype that interfaced with the real IT infrastructure at the industrial partner. The development process consisted of the set of activities, resources, and choices made to develop the analytics demonstrator.

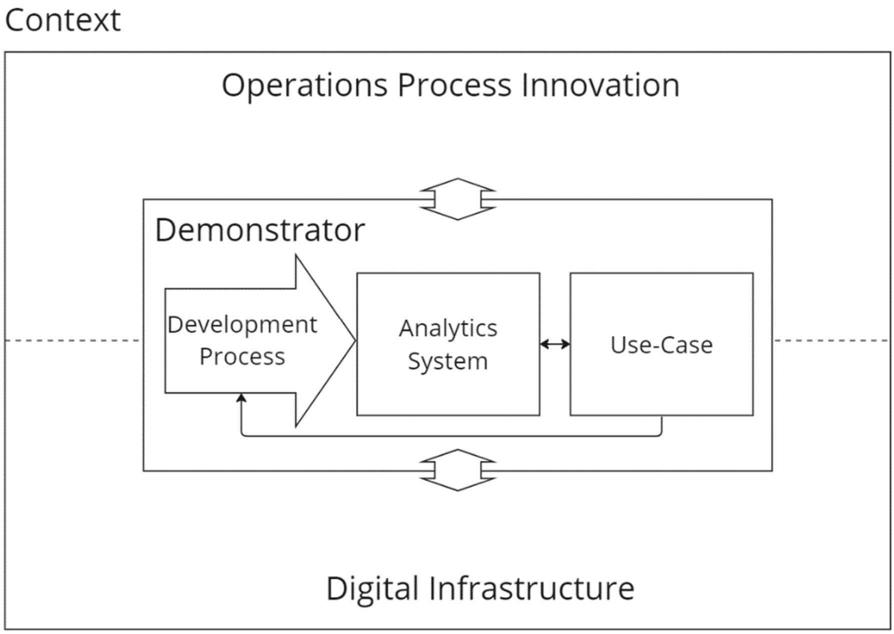


Figure 2.2 Illustration of the ensemble artifact

The relevant organizational context consisted of the organizational and technical infrastructure for innovation. Owing to the orientation and main stakeholders of the research, I was focused on the innovation infrastructure in operations and IT. The digital infrastructure included aspects such as the architecture of the IT landscape and individual systems and the governance and organization of IT. The operations process innovation infrastructure consisted of aspects such as the innovation agenda, the organization of process innovation in operations, prioritization of analytics initiatives, and the resources available for the different initiatives. The analytics demonstrators interacted with and were developed within this context, which ultimately shaped and became inscribed in their design. Developing the demonstrators allowed me to build up understanding and knowledge about the effects of the context on the development process and vice-versa. This was aided by the fact that multiple demonstrators were to be carried out, introducing some variation in contextual aspects while keeping the rest constant.

The research process & knowledge contribution

I leveraged the ADR method of Sein & Rossi (2019) (illustrated in Figure 2.1) in my research to develop the ensemble artifact. First, I sought to develop several analytics demonstrators, each producing concrete IT artifacts in the form of analytics system prototypes and their development processes. Second, at the more abstract level, the goal was to develop an approach for the development of analytics demonstrators.

Each BIE cycle consisted of the construction and evaluation of an analytics demonstrator and provided insights into the feasibility of the process innovation use case.

The research was divided into three overall phases: 1) an exploration phase, following the ADR process; 2) an evaluation phase, where the formalized design knowledge produced in the exploration phase was tested through instantiation in a new demonstrator; and 3) a proposal phase, where managerial prescriptions were developed based on learnings from the first two phases. The evaluation phase was introduced to supplement the essentially inductive logic of ADR (Sein & Rossi, 2019) with deductive theory-testing logic as employed in Action Research and IBR. A very similar design was adopted by Miah et al. (2019), who conceptualized the deductive stage as a three-stage process of *Instantiate-Evaluate-Modify*. An overview of the research process is provided in Figure 2.3. I refer the reader to the appendix in *Paper 4* for a more in-depth description of how the research applied the principles for ADR research (Sein et al., 2011).

The exploration phase consisted of participative observation in three analytics initiatives at the industrial partner and execution of four BIE cycles. The initial problem formulation was created as the result of extensive participative observation in the data platform team at the industrial partner and updated based on learnings from the following BIE cycles and further participative observation in two analytics initiatives. Reflection was carried out throughout the phase and consisted, amongst other things, of extensive scanning of academic literature for research that could support the BIE process. It was in this process that I discovered the concepts of lightweight IT and digital infrastructure that became central lenses in the research. The phase ended with formalization of learning, where the approach was conceptualized formally, design principles were extracted from the demonstrators, and the class of problems addressed was defined.

The evaluation phase consisted of a final test of the approach and design principles in a new demonstrator. First, a suitable problem was scoped for development of a demonstrator in collaboration with stakeholders from the industrial partner. The approach and the design principles were then instantiated through the construction of an analytics demonstrator. The demonstrator itself was evaluated concurrently, while the approach and design principles were evaluated after the end of the demonstrator, following reflection and learning. The design principles and approach were then modified to accommodate the new learnings.

The proposal phase consisted of a further loop of reflection and formalization of learning. In this phase, the scope changed from the demonstrator to the ensemble artifact with the goal of developing prescriptive knowledge for management on how to manage process innovation with analytics and develop a supporting IT infrastructure. This was primarily a conceptual activity, where lessons learned were extracted from the previous two phases and considered in relation to existing literature to develop managerial prescriptions.

The research design is summarized in Table 2.1.

Table 2.1 Overview of research design

Phase(s)	**Objective**	**Method**	**Presented in**
1 & 2	Developing analytics demonstrators	ADR: Problem Formulation, BIE, Reflection & Learning	Section 3
1 & 2	Developing a framework to support engagement in process innovation with analytics	Conceptual: Grounded in literature review and experience from four demonstrators and participative observation.	Section 4.2
1 & 2	Understanding the drivers of development speed and complexity in analytics aimed at process innovation	Abductive analysis of four demonstrators as *design science (Goldkuhl & Sjöström, 2021) or action cases (Braa & Vidgen, 1999)*	Section 4.3
1 & 2	Developing an approach and design principles for fast development of analytics demonstrators aimed at process innovation	ADR: Explore (one full ADR cycle with four BIE cycles) and Evaluate (one cycle of Instantiate-Evaluate-Modify (Miah et al., 2019))	Section 4.4
3	Developing managerial prescriptions for managers looking to engage in process innovation with analytics	ADR: Reflection & Formalization of Learning following two first phases of ADR.	Section 5.1

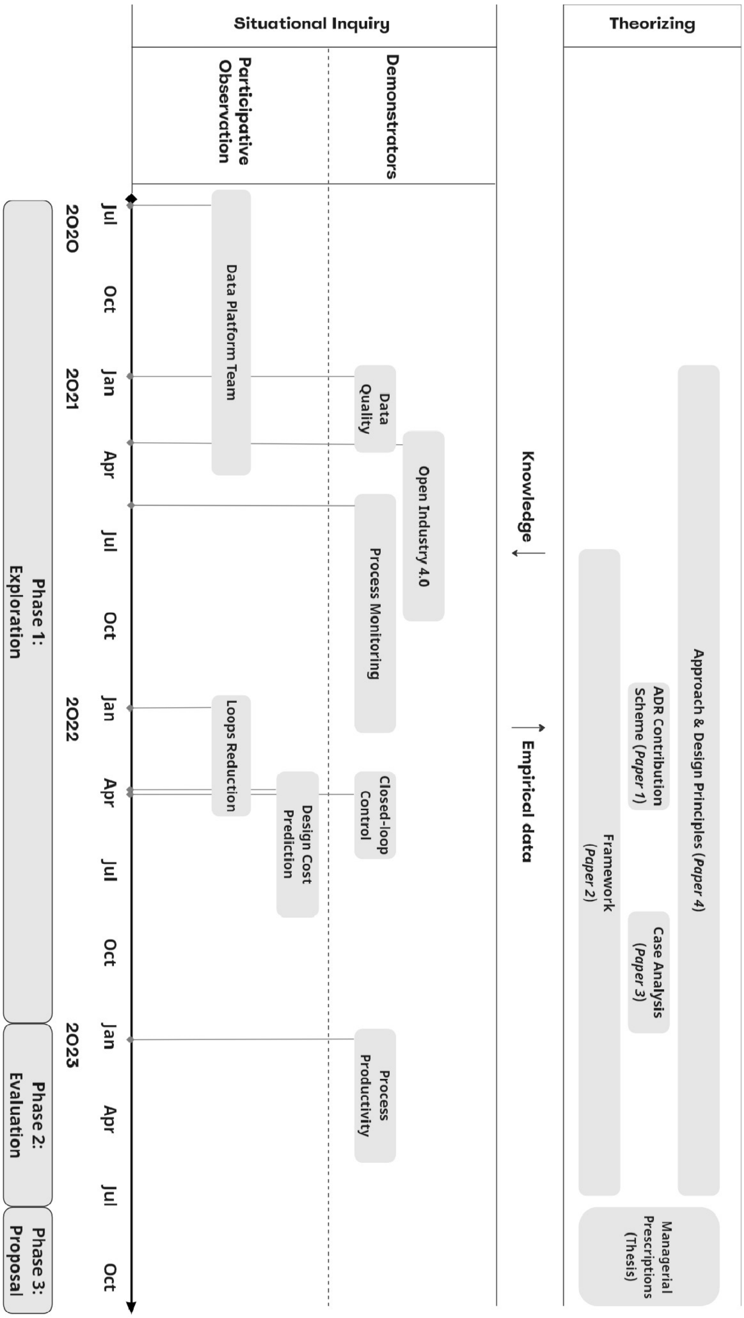


Figure 2.3 Overview of the research process

CHAPTER 3. EMPIRICAL FOUNDATION

In this chapter, I present an overview of the empirical context and work that made up a key part of the research process, namely the situational inquiry. This work consisted of engagement in analytics initiatives at the industrial partner using participative observation and design and action. To facilitate understanding, I first elaborate on empirical concerns that had influence across the individual analytics initiatives, namely how IT-enabled innovation was organized at the industrial partner. I then provide an overview of the analytics initiatives before describing in more detail how I used participative observation and design and action in the analytics initiatives.

3.1. EMPIRICAL CONTEXT

To provide insights into the organization of IT-enabled innovation in operations at the industrial partner, I first describe key organizational units and their place in the organizational structure. I then turn to elaborating on the socio-technical architecture and governance of IT-enabled process innovation. The description of architecture and governance is organized in terms of two different *configurations* (inspired by Hanseth & Modol, 2021), as a change was made to the architecture and governance arrangement while my research was underway.

3.1.1. KEY ACTORS AND ORGANIZATIONAL STRUCTURE

Owing to the size of the industrial partner, many actors are involved in some form as stakeholders in IT-enabled innovation in operations. Three units from IT and operations emerged as key actors in this work across individual areas in operations: 1) Operations Transformation, 2) Operations IT, and 3) the Data Center of Excellence.

On the operations side, the Operations Transformation unit is a key stakeholder, with the head of the unit reporting to the Chief Operations Officer. Responsible for the transformation of operations, the unit includes several subunits focused on specific transformation efforts such as long-term continuous improvement, Industry 4.0, and supply chain innovation. It is furthermore home to Product Managers, who act as key stakeholders in translating business needs into development tasks for the IT organization. I interacted mainly with the Industry 4.0 team, which relied extensively on collaboration with IT (Operations IT and the Data Center of Excellence) and open innovation to identify, assess, and demonstrate the potential value of Industry 4.0 technologies. As part of these open innovation activities, the Industry 4.0 team collaborates extensively with various universities and oversees several PhD students through the MADE research programs.

On the IT side, the Operations IT unit is a key stakeholder, with the head of the unit reporting to the Chief Digital Officer (CDO). Operations IT is responsible for all IT required to support operations, including both regular development and IT operations activities and innovation activities in Operations Transformation. In recent years, the unit has decreased its use of external consultants and vendors and hired more full-time developers and IT managers. The unit has embraced agile development principles and organizes development in IT product teams that have ownership over all IT applications within a defined area (the areas often align with a subprocess). The last key IT stakeholder is the Data Center of Excellence, where the head of the unit also reports to the CDO. The unit consolidates expertise and capabilities related to data to support efforts across the whole organization. Its responsibilities include data strategy and governance, developing and operating enterprise data infrastructure, such as the organizational data platform, and developing and operating analytical systems, such as BI and data science.

3.1.2. DIGITAL INFRASTRUCTURE CONFIGURATION 1: BIMODAL IT & IOT DATA PLATFORM

When I started my research with the industrial partner in July 2020, Operations IT was organized largely according to the logic of bimodal IT. The agile IT product teams were responsible for supporting operations within their defined product area through the operation and maintenance of existing applications, development of new features in the applications, and development or implementation of new applications as specified in their product backlogs. A Product Manager was assigned to each product team and was responsible for translating business requirements into development work and prioritizing the backlog. In addition to the product teams, a specific incubation team had been established to carry out the explorative work required to support the innovation agenda of Operations Transformation. The incubation team acted as an IT partner to Operations Transformation and was responsible for supporting the exploration and assessment of new technologies and use cases. If a use case proved promising enough to transition to the pilot stage, then the IT product teams were to take over ownership and implementation of the pilot. A Data Science unit in the Data Center of Excellence furthermore had the responsibility to support data science efforts across the organization. However, resource allocation decisions had prioritized use cases primarily in marketing and customer-focused applications. As a result, the Data Center of Excellence was only majorly involved in a single operations data science initiative focused on demand forecasting.

In terms of architecture, the IT landscape consisted of a complex heterogeneous set of applications and infrastructure acquired over many years to support operations. An ERP system made up a key IT platform supporting many processes in operations. Several initiatives were underway to modernize the application landscape, as many applications did not live up to modern IT best practices, such as exposing data and functionality through APIs. One such initiative was the introduction of a team responsible for the development and operation of an IoT data platform to support data-driven innovation projects in operations. While the industrial partner already had an enterprise data platform, the design and functionality offered by the enterprise platform had been shaped to a large extent by customer-facing use cases, resulting in a pure-cloud setup that was less suitable for IoT use cases. The IoT platform was created to address this gap and bridge factory and cloud environments while also decoupling the applications owned by product teams from innovation activities. All product teams were to supply key data from their applications to the IoT platform. This was to be realized by creating data pipelines to move data from the source systems to the IoT platform data lake. While product teams were supposed to develop and own the data pipelines, in practice, the IoT platform team ended up taking the task to sustain momentum. In the following year, an increasing number of applications delivered data to the IoT platform, which was largely deemed a success. This lasted until April 2021, following a decision to restructure the overall IT organization at the industrial partner. This marked the transition to a new configuration.

3.1.3. DIGITAL INFRASTRUCTURE CONFIGURATION 2: AMBIDEXTROUS PRODUCT TEAMS & GLOBAL DATA PLATFORM

In April 2021, a reorganization of IT took place at the industrial partner following the hiring of a new CDO. The result of the reorganization was a transition away from the bimodal setup towards an ambidextrous setup and a move to one global data platform. In the new setup, the IT product teams became owners of all development activities within their product area, including exploratory innovation activities. This rendered the incubation team superfluous and resulted in it being disbanded. Another incubation-focused team consisting of two employees was, however, created in the Data Science unit at the Data Center of Excellence, which was dedicated to incubating data science use cases in operations.

The reorganization also meant that the IoT data platform would be discontinued in favor of a modernized version of the enterprise data platform owned by the Data Center of Excellence. The updated enterprise data platform was inspired in many ways by the IoT data platform and leveraged similar technology. For a period, the two data platforms coexisted with ownership moved to the Data Center of Excellence before the team supporting the IoT platform was disbanded. All product teams were from then on to deliver data to the enterprise data platform and work towards migrating any data from the IoT platform to the enterprise data platform. In practice, the result was a scattering of the operations data across the two platform data lakes, which remained the case in the last demonstrator I developed in spring 2023. Several modernization initiatives also followed in the period after the reorganization, including an API initiative to API-enable legacy systems and applications. As a result, an increasing number of applications in operations became API-enabled throughout the following two years, although the journey is still in progress as of October 2023.

3.2. ANALYTICS INITIATIVES

In this section, I introduce the analytics initiatives that make up the empirical foundation of the research. These initiatives consisted of one infrastructure building initiative and seven analytics demonstrators. An overview of the seven demonstrators is presented in Table 3.1, including the digital infrastructure the demonstrators were embedded in, the analytical technology used, the IT capability provided, along with the process targeted and the potential process impact.

The demonstrators were all embedded in the corporate digital infrastructure at the industrial partner, more specifically, the digital infrastructure supporting manufacturing and engineering processes. Enterprise systems were a key source of data throughout the demonstrators. Six demonstrators relied on data originating from ERP, whereas others relied on business intelligence systems (BI), warehouse management systems (WMS), manufacturing execution systems (MES), and computer-aided design (CAD) systems. The enterprise systems were all (except for the MES at the Test Facility) owned by IT product teams. The enterprise systems were the sources of master data and event logs related to the production, quality, and engineering design processes. Another key data source was IoT data collected from manufacturing equipment (OT), such as sensor measurements and parameter settings. The infrastructure for the IoT data ranged from custom-built data pipelines to a combination of MES and custom-built integration components. Furthermore, the demonstrators all relied on the data platform and cloud infrastructure services.

In terms of analytics, the demonstrators varied widely in analytical complexity. The demonstrators ranged from descriptive to prescriptive analytics and from off-the-shelf to custom-developed artifacts. Key technologies were anomaly detection, open-source ML-based classifiers and regressors (e.g., models from scikit-learn, keras, and xgboost), and dashboards. The prescriptive analytics use case relied on a combination of ML-based predictions and rule-based logic rather than the traditional prescriptive approach of mathematical programming.

The demonstrators mainly aimed to improve process performance in manufacturing-related processes. Three of the demonstrators focused on improving the performance of a core manufacturing process, moulding, by reducing scrap or downtime. Two other demonstrators focused on reducing costs or rework in the engineering design process that designed and delivered moulds for the manufacturing process. The remaining two demonstrators mainly aimed to demonstrate the generic IT capabilities offered, namely anomaly detection and near-real-time visualization. The potential process impact of analytics ranged from minimal, such as better decisions or increased visibility, to redesigns of support processes in manufacturing, such as maintenance, production preparation, and inspection. In the following subsections, I elaborate further on my use of participative observation and the activities of the BIE cycles I conducted.

Table 3.1 Overview of the analytics demonstrators

	Digital Infrastructure		Analytics		Process	
Case	**Source Systems**	**Data**	**Technology**	**IT Capability**	**Type**	**Potential Impact**
Data Quality Anomaly Detection	Data Lake (origin various, e.g., WMS, ERP, IoT)	Various Relational Tables	Anomaly Detection System	Detect anomalies in data quality	IT Operations	Faster detection of issues. Redesign data pipeline monitoring process.
Process Monitoring	QIS, WMS, MES, ERP, Data Lake	Quality log, Master Data, Process (Sensor, Machine Parameters)	ML Classifier, Multivariate SPC	Detect quality issues	Moulding	Real-time quality evaluation. Redesign of inspection process.
Design Cost Prediction	ERP, BI	Master Data, Design Specifications, Accounting Data	ML Regressor	Predict costs at design stage	Engineering Design	Better decisions. Minimal process change.
Open Industry I4.0	OT, OPCUA Server	Equipment Status	Near real-time Dashboard	Visualize production status	Packing	Improved visibility. Minimal process change.
Loops Reduction	CAD, ERP	Master Data, Design Specifications, Design Log	Clustering, ML Classifier, ML Regressor	Predict issues at design stage	Engineering Design	Reduce rework. Minor redesign of the process.
Closed-loop Control	MES (Test Facility), ERP, QIS	Process (Sensor, Machine Parameters), Master Data, Dimensional Measurements	ML Classifier & Rule-based prescriptions	Prescribe better machine settings	Moulding	Reduced variation, better quality. Redesign of production preparation process.
Process Productivity	Data Lake, ERP	Process (Sensor, Machine Parameters), Production Log, Master Data	Near real-time Dashboard, Rule-based Alerting	Detect process instabilities	Moulding	Real-time stability evaluation. Redesign of maintenance process.

3.2.1. PARTICIPATIVE OBSERVATION: ORGANIZATIONAL ANALYTICS INITIATIVES

An important source of data in my research was participative observation in three organizational analytics initiatives. One of these was an infrastructure building initiative, whereas the two others were analytics demonstrators. Table 3.2 provides an overview of these three initiatives and my role in them. It is worth noting that most of the observation took place remotely due in part to the COVID-19 situation and due to the virtual organization of work caused by geographically distributed participants.

I started out the research process by being embedded in the newly formed IoT Data Platform team at the industrial partner. As the name implies, the team was responsible for building and operating an IoT data platform to support analytics activities with operations data. My role on the team was to be the data science expert. The goal of my participation was to better understand the problem and solution spaces related to analytics demonstrator development at the industrial partner. The team relied on agile development, and I participated in the team's daily standups, weekly team meetings, and sprint demonstrations. I furthermore took on smaller development tasks to get experience with the development of data pipelines and the platform itself, actively participated in architectural discussions, and provided advice on platform functionality related to data science. The participation was primarily conducted remotely via Microsoft Teams due to the COVID-19 pandemic, which forced the team to work mostly from home. I took field notes throughout the process, although the detail level and frequency of note-taking varied. I furthermore had and retained access to the internal documentation wiki, source code repositories, and the DevOps boards where development activity was tracked. Participating on the team allowed me to obtain a detailed understanding of the work that goes into developing and operating data and analytics infrastructure and the technology landscape of the organization.

I also participated as an observer in two analytics development activities at the industrial partner. The first of these, the Loops Reduction initiative, was a four-month project where an external vendor collaborated with the organization to identify and develop an analytics use case. The initiative was orchestrated as a collaborative, agile development project. The vendor carried out project management and analytics development, while an internal team of IT developers, managers, and business stakeholders from the industrial partner provided support and business context. At the request of the VP of Operations IT, I took part in this internal team, playing the advisory role of data science expert. I participated via Microsoft Teams in daily standups, sprint demonstrations, and weekly meetings. Remote participation was a natural choice, as the vendor organized the project virtually. All these observations were documented extensively in field notes to facilitate later analysis. Following this initiative allowed me to gain extensive insights into how the vendor, an internationally well-renowned and leading data science organization, carried out analytics demonstrator projects.

I became involved in the last analytics development initiative, Design Cost Prediction, following the Loops Reduction initiative, where I met an engineer who had developed a promising machine learning model to predict the realization costs of their engineering designs. He needed IT support to deploy his model in an ML application, and I thus teamed up with another IT engineer to participate in deploying the ML application. The deployment was carried out through multiple physical workshops that first attempted to understand the problem and model developed before moving into architecture discussions and finally realization. I contributed mainly with designing the architecture and deploying the application on the data platform. I documented the workshops extensively using field notes for further analysis. Participating in the initiative provided me with insights into the deployment phase for a custom-developed model, which I had no practical experience with at the time.

Table 3.2 Overview of initiatives in which I conducted participative observation

Initiative	**Role**	**Observations**	**Outcome**
IoT Data Platform	Team Member: Data Science Expert	Daily Standups, Weekly Team Meetings, Sprint Demos, Architecture Discussions, Pair Coding. Documented in field notes.	Insight into analytics infrastructure development and technology landscape.
Loops Reduction	Advisor: Data Science Expert	Daily Standups, Sprint Demos, Weekly Meetings. Documented extensively in field notes.	Insight into a professional and highly resourced analytics demonstrator project.
Design Cost Prediction	Advisor/Co-developer: Data Science Expert	Multiple workshops consisting of both hands-on co-development and architecture and use-case discussions. Documented extensively in field notes.	Insight into the deployment phase of a custom-developed ML model using the organizational data platform.

3.2.2. DESIGN & ACTION: BUILDING AND EVALUATING ANALYTICS DEMONSTRATORS

The main source of data in my research consisted of the five analytics demonstrators I developed at the industrial partner. Table 3.3 provides an overview of the five demonstrators and my role in them. In all the demonstrators, I acted as either the sole designer or a co-designer in collaboration with employees at the industrial partner, external consultants, or another university researcher. Key stakeholders for many of the demonstrators were employees from the Operations Transformation unit, given their responsibility for innovation related to operations. These employees had roles as either Innovation Manager or Product Manager. Other operations stakeholders included data managers and process engineers, who were involved in developing the manufacturing process.

The BIE cycles all consisted of the development of analytics artifacts using the data infrastructure of the industrial partner. By developing the artifacts in context, I was able to gain rich insight into the effects of IT and data infrastructure on analytics development, which I would otherwise not have been able to obtain. Owing to the data platform consolidation that took place during my research (explained in section 3.1), early demonstrators were built on the IoT Data Platform, whereas later demonstrators were built on the global data platform. This mainly made a difference in terms of the data sources that had been integrated with the platforms, as the underlying technology used for the platforms was similar.

Except for the *Process Monitoring* demonstrator, the artifacts developed were all functional analytics system prototypes that had been integrated with the operational infrastructure using the data platform. This integration made it possible to leverage real and near-live data in the demonstrators. The scope of the demonstrators thus differed from the smaller scope common to most analytics development projects, which rely on a static extract of historical data to develop and evaluate models.

Evaluation was naturalistic (Venable et al., 2016) and concurrent, as recommended in ADR (Sein et al., 2011). This included ongoing demonstrations within the demonstrator team and to a broader group of stakeholders, often at the end of the demonstrators. This broader group included IT engineers and IT managers who had an interest in the initiatives. From a research perspective, key outputs were the artifacts themselves, i.e., their source code and architecture, resulting documents and presentations, as well as field notes, which I used to document the BIE cycles. The first two BIE cycles were documented only sporadically in field notes, while the last three were documented more extensively.

Table 3.3 Overview of the five demonstrators developed using BIE cycles

Demonstrator	**Team & Role**	**BIE**	**Learnings**
Data Quality Anomaly Detection	Designer. Developed while embedded in IoT Data Platform Team.	Development of data pipelines and configuration of an anomaly detection system. Evaluated through use with real data.	Insight into demonstration with off-the-shelf analytics system.
Open Industry 4.0	Designer of analytics part on a team with researchers, external consultants, and Innovation Manager.	Development of streaming data pipelines and near-real-time dashboard in PowerBI. Evaluated by physical demonstration at workshop and through several follow-up presentations.	Insight into IT-OT link in analytics demonstrators.
Process Monitoring	Co-Designer on a team with Innovation Manager (also Industrial PhD)	Development of datasets and multiple ML models. Evaluated through statistical evaluation and through presentation to IT and engineering stakeholders.	Insight into the complexity and heterogeneity of the data landscape.
Closed-loop Control	Co-Designer of Analytics part. Team with three data managers and external consultants from a GTS.	Development of data pipelines and deployment of existing ML model and rule-based prescriptions. Evaluated through proof-of-concept use with data managers and presentation of findings to managers.	Insight into the development affordances provided by an enabling IT infrastructure.
Process Productivity	Designer on a team with a Product manager, two users, and an Innovation Manager.	Development and deployment of data pipelines, near real-time dashboards, and rule-based anomaly detection. Evaluated through demonstrations with users.	Effectiveness of the developed design knowledge.

CHAPTER 4. RESEARCH FINDINGS

In this chapter, I present a summary of the findings from each of the four appended papers that make up the research contribution. Although the papers differ in method and scope, they are all outcomes of the ADR process and linked to the overall objective of producing academically and practically relevant knowledge related to process innovation with analytics. Table 4.1 presents an overview of the four papers with a short summary of their purpose, method, and findings.

The first paper is focused on methodology and contributed to the research design developed as part of the research process. The second paper leverages the empirical experiences from ADR and a review of literature to develop a conceptualization of process innovation with analytics and an associated research agenda. The third paper contains an abductive case analysis of four of the demonstrators conducted as part of the ADR process to provide insights into the factors driving development speed. The fourth and final paper builds on the conceptualization in Paper 2 and the factors identified in Paper 3 to report on the ADR project as a whole and extract design principles. Figure 4.1 illustrates the linkages between the papers. In the following sections, I provide a more in-depth summary of the four papers and supplement it with further reflections.

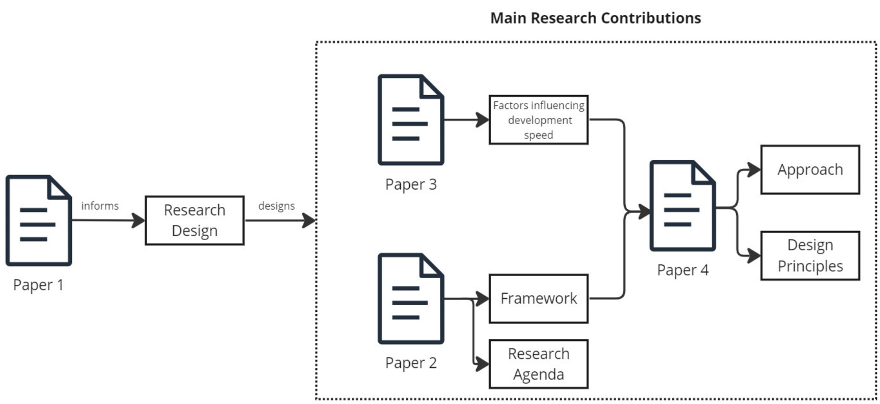


Figure 4.1 Illustration of the linkages between the four papers

Table 4.1 Overview of the research findings for each of the four papers

	Paper 1	**Paper 2**	**Paper 3**	**Paper 4**
Title	Towards a Scheme for Contribution in Action Design Research	Conceptualizing Process Innovation with Analytics - A Pragmatic Framework and Research Agenda	Speeding up Explorative BPM with Lightweight IT: The Case of Machine Learning	Developing Analytics Demonstrators for Process Innovation: An Infrastructural Perspective
Purpose	Conceptualize potential contributions in ADR to support research design.	Conceptualize process innovation with analytics and develop a practically relevant research agenda	Investigate determinants of development speed in exploratory ML-enabled process innovation	Develop prescriptive knowledge for analytics demonstrator development in large, established organizations
Method	Literature Review, Conceptual, Case Illustration	Action Design Research, Conceptual	Abductive Case Analysis	Action Design Research
Findings	ADR can contribute with different empirical, theoretical, and artefactual contributions throughout its phases.	Process innovation with analytics requires a coordinated transformation of analytics, process, and infrastructure. New governance and development approaches are needed for this class of problems.	Lightweight IT can speed up assessment and technical implementation but requires a loosely coupled digital infrastructure and the use of building blocks in development.	An approach and a set of design principles for developing analytics demonstrators. The importance of considering digital infrastructure in development.

4.1. PAPER 1: WHICH CONTRIBUTIONS CAN BE DEVELOPED THROUGHOUT THE ADR RESEARCH PROCESS?

The first paper set out to identify and conceptualize the potential for contributions in action-oriented DSR. We were motivated by a very concrete research design problem that I was facing in the early stages of my research. With the transition to dissertations consisting of a collection of articles, Danish universities expect PhD students to develop and publish multiple articles during their PhD, ideally in high-quality journals. The methodological DSR discourse in top IS journals focused mainly on the design artifacts and generalizations thereof in the form of design principles or design theory as the main contribution and "publishable unit" in DSR (e.g., Gregor & Jones, 2007; Sein et al., 2011; Gregor & Hevner, 2013; Gregor et al., 2020). A prototypical research design would consist of a series of conference articles presenting interim results, followed by one summary paper submitted to a top journal. This was somewhat at odds with the institutional expectations I was facing. Inspired in part by Intervention-based Research (Oliva, 2019), a different way of conceptualizing action-oriented DSR in OM, I set out to explore and conceptualize possibilities for contribution from action-oriented DSR throughout the research process.

To identify and conceptualize contribution possibilities, I relied mainly on a review of the literature on contributions in DSR in both IS and OM. In this process, I used various keyword searches on Google Scholar, as well as snowball sampling. I then engaged in a conceptual phase, where I synthesized the various conceptualizations of contributions in DSR and related them to the research phases of ADR. I finally provided a small evaluation of the applicability of the conceptualization by using it to support development of my research design.

The main conclusion was that action-oriented DSR can contribute with more than design artifacts, design principles, and design theory. Action-oriented DSR generates a lot of empirical material throughout its phases. This empirical material can be the basis for other types of contributions, including empirical contributions, theorizing products, and more formal theory building. Furthermore, as emphasized in IBR and Action Research, theoretical frameworks are confronted with reality through intervention, presenting an opportunity to test theories in terms of their pragmatic validity. Table 4.2 illustrates the conceptualization developed. For elaboration on the differences between the four types of theories leveraged in the conceptualization, I refer the readers to the descriptions offered in the paper. It should be noted that the conceptualization developed leverages the research phases presented in the elaborated ADR process, namely *Diagnosis*, *Design*, *Implementation*, and *Evolution* (Mullarkey & Hevner, 2019), rather than the phases in Sein et al. (2011), which I presented in section 2.2.2 (*Problem Formulation*, *Build-Intervene-Evaluate*, *Reflection and Learning*, *Formalization of Learning*). The two process models are, however, compatible. The elaborated ADR model is essentially a way of classifying the ADR iterations of Sein et al. (2011) according to the nature and purpose of the BIE cycles – whether they are focused on understanding the problem (*Diagnosis)*, conceptual design of artifacts (*Design)*, implementation of artifacts (*Implementation)*, or improving already implemented artifacts (*Evolution*).

Table 4.2 Conceptualization of potential contributions in the four ADR phases (from Bojer and Møller (2022))

Potential Contribution	**Diagnosis**	**Design**	**Implementation**	**Evolution**
Theory Building				
Design Theory	Hypothesis, Propositions		Inductive & Abductive Theory Building	
Substantive Technological Theory				
Type I-IV Theory	Theorizing Products		Theorizing Products, Inductive & Abductive Theory Building	
Practical Theory	Theorizing Products, Theory Modifications		Theorizing Products, Theory Modifications, Inductive & Abductive Theory Building	
Theory Testing				
Design Theory			X	X
Substantive Technological Theory			X	X
Type I-IV Theory	X		X	X
Practical Theory	X	X	X	X
Non-theory				
Rich Empirical Descriptions	X	X	X	X
Artifacts	X	X	X	X

Reflections

Looking back at the paper, it is hard not to notice that the research design I arrived at in the paper was different from the research design I ended up with. For one, I was optimistic about the number of publications, which is not uncommon for early-stage PhD students. The change in research design is, however, to be expected due to the emergent nature of ADR research. As noted in the paper: "the research design and publication strategy will […] have to be revisited as the research process unfolds" (Bojer & Møller, 2022, p. 385).

After presenting at the DESRIST 2022 conference, I discussed publication and research design challenges with one of the creators of ADR, who sympathized greatly and opined that ADR-based dissertations ought to be monographs. As is evident from this thesis, I decided not to follow that suggestion. Instead, I relied on the potential for empirical and theorizing products as contributions to develop the two other main contributions (*Paper 2* and *Paper 3*) that make up the research. Whether this was a wise decision remains to be seen.

Reflecting on the scheme, I no longer believe that making the distinction between substantive technological theories and design theories and type I-IV and practical theories adds enough value to justify the added complexity. One thing is, however, certain to me: further research on how to best scope out various contributions action-oriented DSR is warranted.

4.2. PAPER 2: HOW SHOULD PROCESS INNOVATION WITH ANALYTICS BE CONCEPTUALIZED?

The goal of the second paper was to reconceptualize analytics in the broader context of process innovation and propose a research agenda reflecting this updated conceptualization. It became clear throughout my engagements with the industrial partner that existing socio-technical analytics research was mainly focused on using analytics to obtain insights to guide tactical and strategic decision-making. A similar conclusion was made by Badakshan et al. (2022). The existing conceptualizations of analytics, however, seemed to ignore a lot of the complexities I had faced in my engagement with analytics demonstrators aimed at operational process innovation. As has since been noted by Davenport & Miller (2022), this use of analytics seemed to have much in common with business process reengineering (BPR) (Davenport, 1993; Hammer & Champy, 1993), but at the time, no one else seemed to have published on the connection. At the same time, it was also clear that analytics was different from the traditional software that had been used in the BPR era. The paper thus sought to establish an updated conceptualization and clarify exactly how process innovation with analytics was different from traditional IT-enabled process innovation.

The conceptualization was developed as a result of theorizing during the ADR project. It is essentially a generalized conceptualization of the problem space of the research, which was developed and refined over a more than two-year-long process. The research agenda was developed by engagement with relevant areas of literature where existing knowledge fell short in providing practical insights.

The paper proposes that process innovation with analytics consists of coordinated development of analytics, the digital infrastructure, and the process. The development is influenced by and takes place within the context of the existing processes, digital infrastructure, and organizational governance. The outcome is a changed process, enabled by an analytics system and changes to the digital infrastructure. Figure 4.2 depicts the conceptualization. The paper further establishes some important differences between process innovation with analytics and traditional IT-enabled process innovation. Specifically, the scope of process change is smaller, and technological exploration and development are required prior to process design to discover process innovation affordances.

We further proposed the need for research into 1) digital infrastructures for process innovation with analytics, 2) the relationship between process change and analytics development, and 3) governance of process innovation with analytics. Table 4.3 illustrates the detailed research agenda proposed.

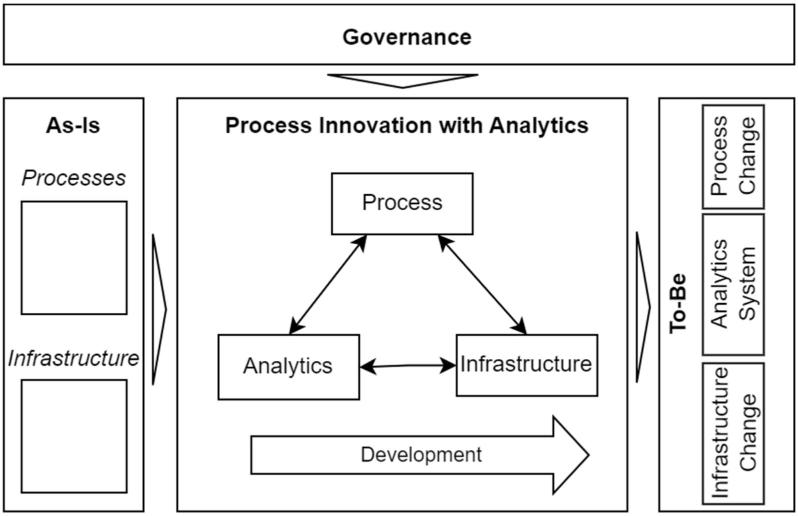


Figure 4.2 Conceptualization of Process Innovation with Analytics (from Bojer and Møller (2023a))

Reflections

While the focus of the paper was to support further research into process innovation with analytics, the updated conceptualization has several implications that were not possible to touch upon in the paper due to space limitations. Most significantly, the conceptualization implies that the complexity and scope of process innovation with analytics initiatives differ along three dimensions: 1) IT infrastructure, 2) analytics, and 3) process. A further implication is that initiatives can vary widely in their complexity and scope, making them qualitatively different in significant ways. I find it helpful to imagine each initiative as being located somewhere in the three-dimensional space visualized in Figure 4.3.

Table 4.3 Research agenda proposed for process innovation with analytics (from Bojer and Møller (2023a))

Research Direction	Research Challenge	Promising Angles
Digital infrastructures for process innovation with analytics	Understanding the influence of the existing digital infrastructure	Adoption of an information infrastructure lens in qualitative research (e.g., Ciborra et al., 2000; Hanseth & Lyytinen et al., 2010)
	Transition strategies: From constraining to enabling infrastructure	Contextualization of insights from enterprise architecture (Ross et al., 2006) and classic infrastructure literature (Broadbent et al., 1999). Exploring the potential applicability of *data ecosystems* (Gröger, 2021) and *platformization* (Bygstad & Hanseth, 2018).
Exploring the relationship between process change and analytics development	Understanding the impact of the existing process	Retrospective and longitudinal case studies of process innovation with analytics focusing on the *process* of process change and the benefits realized.
	Constructing integrated methodologies for developing analytics-enabled processes	Contextualizing and integrating analytics development and process innovation methodologies.
Advancing our understanding of governance of process innovation with analytics	Identifying and understanding successful governance configurations	Contextualizing existing governance models, e.g., Lightweight IT (Bygstad & Iden, 2017) Case studies and configurational analysis to examine the interplay of context, governance configurations, and their impacts.

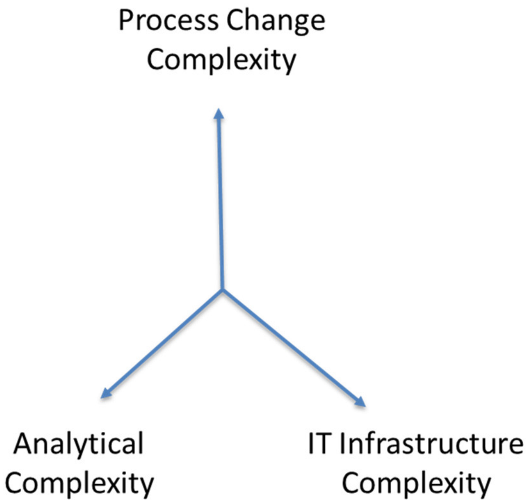


Figure 4.3 Complexity in Process Innovation with Analytics

As an illustration of one end of the complexity spectrum, one could imagine an analytics initiative developing a dashboard using existing BI data to augment a task, resulting in only minor process changes. On the other end, an example would be a project to redesign quality processes by leveraging custom-developed reinforcement learning, requiring a new data collection infrastructure to capture machine sensor data and digital means of controlling production machines through APIs. The different nature of such projects should obviously be factored in when selecting and resourcing initiatives. If the goal is to demonstrate value in the short term to gain momentum in organizational transformation, it is necessary to limit the overall complexity of prioritized initiatives by managing trade-offs along the three dimensions.

In terms of the conceptualization or framework, reaching the right level of detail was a continuous struggle through the theorizing process. Earlier versions of the framework were more complicated and sought to include details on the components of the three key development areas (process, infrastructure, analytics) and their relationships. In the end, the simplified version presented in the paper emerged following attempts to simplify and distill the lower-level representations. Nonetheless, these lower-level details remain important for supporting action within each development area. It is furthermore worth acknowledging that other organizational factors that have been found relevant for IT-enabled process innovation and analytics, such as culture and capabilities, remain relevant for process innovation with analytics. These factors are likely also to require transformation or development to realize process innovation with analytics but have been outside the scope of this research.

4.3. PAPER 3: WHAT DRIVES DEVELOPMENT SPEED IN ML-BASED DEMONSTRATORS?

The third paper aimed to identify key factors that drive development speed in analytics demonstrators using ML. Achieving a deep understanding of the drivers of development speed would be necessary to engineer an approach capable of fast demonstration and assessment. The research was motivated by the difference in development speed I had observed in the demonstrators in which I had been involved. These demonstrators had aimed to quickly explore the process innovation affordances of ML for a particular use case, but two of the demonstrators had been less successful in achieving this goal. How to quickly assess the process innovation affordances of new technology was one of the challenges identified in the discourse on explorative BPM (Rosemann, 2014). However, little empirical research existed to answer the question. Examining the literature, the distinction between lightweight and heavyweight IT (Bygstad, 2017; Bygstad & Øvrelid, 2020) emerged as a seemingly useful lens to explain the failure or success of the demonstrators in terms of speed. It was, however, unclear whether and when the concept of lightweight IT could apply to ML as it did not fit the ideal type of lightweight IT. We thus set out to identify explanations for the differences in development speed in the analytics demonstrators and investigate the applicability of the lightweight IT concept.

To identify the factors driving development speed, we relied on an abductive analysis (Dubois & Gadde, 2002; Sætre & van de Ven, 2021) of multiple analytics demonstrators. We selected the five demonstrators I had been involved in, which leveraged ML. The *Process Monitoring* demonstrator was not presented in the paper for reasons of brevity, as the analysis provided results similar to the *Loops Reduction* demonstrator. The analysis relied on the creation and iterative adjustment of an analytical framework and explanation through confrontation with literature and the empirical material.

The conclusion of the analysis was that two factors influence development speed in particular: 1) the nature of *coupling* between the development process and *heavyweight IT,* that is, the digital infrastructure operated by the IT department, and 2) the extent and nature of *building blocks* used in the development process. In terms of coupling, development of ML-based demonstrators typically requires interaction with *heavyweight IT* systems to 1) acquire data for model development and 2) integrate with systems for data and functionality access in the deployment of the prototypes. Our analysis showed that when coupling is tight, the development process becomes reliant on extensive support from the IT department to access the necessary data and functionality, which slows development speed significantly. This slowdown is caused by the need to coordinate across teams with different priorities and development cultures. Conversely, we found that loose coupling positively influences development speed. Under loose coupling, the resources necessary can be accessed largely independently from the IT department through, e.g., APIs or existing datasets published on a data platform. Our analysis further showed that the extensive use of higher-level building blocks contributes positively to development speed by reducing the scope and complexity of the development effort. The building blocks in the cases ranged from boundary resources (Ghazawneh & Henfridsson, 2013), such as APIs and BI views, to infrastructure offered in developer platforms, and solution components, such as software components and packages. Both identified factors are related to the digital infrastructure: coupling is influenced largely by its architecture and use of building blocks is influenced by the IT capabilities offered, i.e., whether relevant infrastructure and boundary resources are available.

Based on these findings, we proposed that ML can be considered as *lightweight IT* under the conditions of loose coupling and extensive use of building blocks. It was also under these circumstances that our case analysis showed that development speed was highest. From an explorative BPM perspective, the findings suggest that *lightweight IT* enables fast exploration and assessment of process innovation affordances. Table 4.4 provides an overview of the case analysis.

Table 4.4 Summary of the case analysis of the demonstrators (from Bojer et al. (2023))

Cases	**Coupling**	**Use of Building Blocks**	**Light vs. Heavy**	**Speed**
Loops Reduction	*Tight Coupling*	*Partly*	Heavyweight	Slow
Design Cost Prediction	*Loose/No Coupling*	*Partly*	Mediumweight	Medium
Closed-loop Control	*Loose Coupling*	*Extensively*	Lightweight	Fast
Data Quality Anomaly Detection	*Loose Coupling*	*Extensively*	Lightweight	Fast

Reflections

The paper presents a relatively simple yet powerful high-level explanation for the different outcomes I observed in the demonstrators. Most importantly, the findings have clear implications for action when it comes to the development of analytics demonstrators. Analytics projects are often prioritized by assessing potential business value and ease of implementation and selecting projects that score high on both dimensions (see, e.g., Hindle & Vidgen, 2018). The findings suggest that decision-makers pay close attention to coupling and building block availability when assessing ease of implementation. Ideally, demonstrators that can leverage mature building blocks and be executed with loose coupling to IT should be prioritized. However, even when prioritization decisions have been made without consideration of these factors, it is often possible within the scoping of the development project to make decisions that reduce the extent of coupling and further the use of higher-level building blocks. Examples include which data sources to include in the project scope and how to access them, as well as ensuring a thorough assessment of available technologies before deciding to build custom ML models.

A valid question concerns whether it makes sense to prioritize ML projects that can be run as lightweight IT. Should organizations not select the opportunities that have the highest potential business value and then ensure the resources are available to make it happen? To the extent that the business case for the analytics initiative is strong enough or considered a strategic priority, the answer is yes. In practice, this was often not the case for the demonstrators I participated in. The potential business value of the analytics use cases was often highly uncertain. Often, it was not even known whether solving the problem with analytics was feasible, and part of the aim of the demonstrator was to discover that. In settings such as these, heavyweight IT resources often end up being allocated towards incremental development tasks with a more certain payoff, preventing the demonstrators from progressing.

The use of abductive reasoning means that the explanation should be viewed as a hypothesis or proposition. Although it aligns with both the empirical material and draws on arguments from existing literature on lightweight IT and digital process innovation (Bygstad, 2017; Bygstad & Øvrelid, 2020) and innovation as recombination (Henfridsson et al., 2018), it has not been subjected to the degree of testing that is characteristic of deductive research. The mechanisms are admittedly relatively high-level abstractions of more detailed mechanisms, such as the effects of specific building blocks. While the explanation is tested through use in the development of the final demonstrator in Paper 4, further testing is required to assess the quality of the explanation in different contexts.

4.4. PAPER 4: HOW SHOULD ANALYTICS DEMONSTRATORS BE DEVELOPED?

The fourth paper sought to deliver on the main objective of the research project, namely, to construct prescriptive knowledge for analytics demonstrators. Specifically, the goal was to develop an approach for fast development of analytics demonstrators at the industrial partner and to derive generalized design principles as a contribution to the literature on analytics development. While the existing literature provides several analytics development methodologies (e.g., Wirth & Hipp, 2000; Microsoft, 2023; Martínez-Plumed et al., 2019), they are light on advice when it comes to the socio-technical aspects of development (Vial et al., 2023). Furthermore, they do not deal substantially with the deployment aspect of analytics, which takes on greater importance in analytics demonstrators, where the goal is to develop an operational analytics system prototype and not just an analytical model.

To develop the approach and design principles, I relied on the first two phases of the research design presented in section 2.2.3. The first phase was the exploratory phase, which leveraged participative observation and development of four analytics demonstrators at the industrial partner to construct and formalize an approach and design principles. The second phase was the evaluation phase, which consisted of instantiating and evaluating the formalized knowledge in a new demonstrator and updating the approach and design principles based on learnings.

The result of the research was the approach for analytics demonstrator development illustrated in Figure 4.4. The approach relies on an infrastructure building phase to put into place the necessary infrastructure for subsequent iterative and user-focused analytics development. In the infrastructure phase, reusable data pipelines are constructed that make the necessary data available in a data platform. In the user-focused demonstrator phase, the data platform is leveraged to rapidly develop and deploy the data, model, and system components that make up an operational prototype that can be shown to users. The presence of the reusable data pipelines allows the demonstrators to be built with real “live” data and deployed as part of the digital infrastructure without the data pipelines being specific to the demonstrator initiative.

We furthermore extracted six design principles (see Table 4.5 for a detailed overview) for demonstrators that address the socio-technical aspects of analytics development, including technology choices, scoping, and organizing:

- *Technology choices*: selecting standard and flexible technologies whenever possible.
- *Scoping*: preferring self-service data sources and existing infrastructure and starting simple with a reduced scope.
- *Organizing*: ensuring access to infrastructure resources during demonstrator development.

Reflections

The development approach advocated for in the paper is largely shaped by software development with learning and adoption in mind. One of the key requirements by the main stakeholder at the industrial partner was that the approach would be able to quickly put systems in the hands of users to facilitate learning. The developed approach thus has much in common with agile development. Although agile development is not uncommon in analytics development, it is generally limited to the data preparation and model development stages, where data and models are iteratively improved through user feedback. The move from model to system and subsequent deployment, however, often happens following a stage-gate style decision. In contrast, the approach developed shows how it is possible, with the right infrastructure in place, to make deployment a part of the agile development process, where an operational system prototype is improved based on feedback from use. This has certain similarities to what Hertzum et al. (2012) refer to as pilot implementation, where the focus is on the deployment of systems in actual use settings with the aim of obtaining feedback for development rather than learning about implementation issues. While I succeeded in obtaining feedback from use in four of the demonstrators, the scope of use and the extent of mutual adjustment of process and technology achieved could have been greater. I do not believe this was due to shortcomings of the approach, but rather by a lack of management support and resources on the user side.

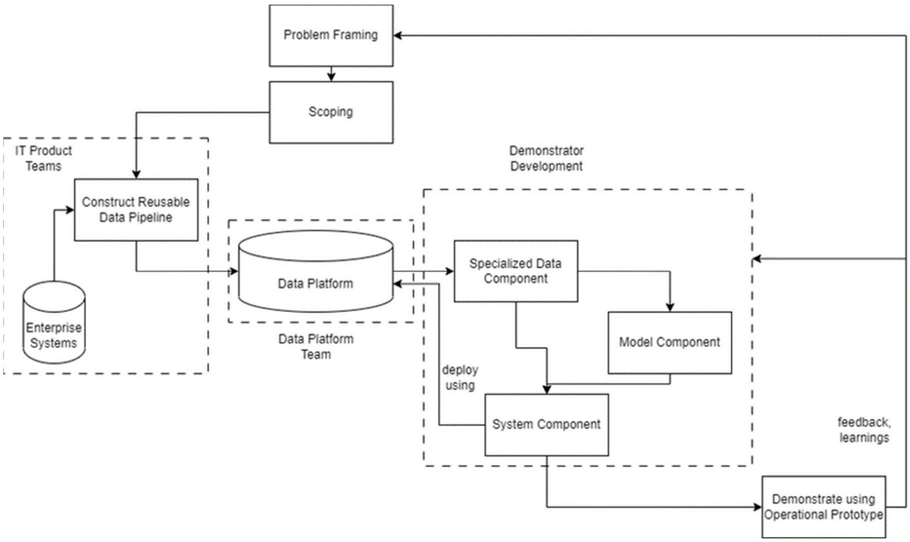


Figure 4.4 Depiction of the demonstrator development approach (from Bojer and Møller (2023b))

The approach additionally pays greater attention to adoption compared to traditional analytics development methods. In her influential theory of the diffusion of innovation, Rogers (1995) suggests that the adoption of an innovation is influenced by the following characteristics of the innovation: 1) relative advantage, 2) complexity, 3) compatibility, 4) trialability, and 5) observability. Adoption in the case of analytics demonstrators concerns both adoption from the point-of-view of the users and the IT department. Except for relative advantage, which is specific to the use case and not influenced directly by the approach, the approach addresses each of the four other factors:

- Trialability and Observability are increased by focusing on putting working systems in the hands of users.
- Complexity and Compatibility from the point-of-view of IT is reduced by relying on standard technologies and existing infrastructure to the extent possible

Table 4.5 Developed design principles (from Bojer & Møller (2023b))

Design Principle	Description
DP1: Use standard components. - **DP1.1**: Leverage existing ML services, Automated ML, pre-trained models, and models in that order. - **DP1.2**: Leverage standards for packaging and deploying models. - **DP1.3**: Leverage existing lightweight solutions for UI.	To speed up initial prototype development, use standard components for data (e.g., generic data pipelines and APIs), model (e.g., AutoML, open-source software), and system (e.g., ML-as-a-Service, SaaS, model deployment standards) components when possible.
DP2: Prefer self-service data sources.	To speed up initial prototype development, prioritize self-service data sources in problem framing.
DP3: Start simple, demonstrate quick wins. - **DP3.1**: Augmentation before Automation – Visualization before Predictive, and Prescriptive Analytics - **DP3.2**: Small scale projects with few data sources and stakeholders.	To enable fast development and user involvement, start simple in framing the problem and incrementally demonstrate value before increasing complexity.
DP4: Select flexible technologies. - **DP4.1**: Select technologies with configurable rules and thresholds. - **DP4.2**: Select technologies with configurable user interfaces.	To facilitate fast iterations in prototype development, select flexible technologies.
DP5: Leverage the installed base. - **DP5.1**: Prioritize and leverage data that is already in use. - **DP5.2**: Leverage existing infrastructure and integrations over building new infrastructure	To increase success rate and speed in analytics demonstrators, leverage and build upon the existing IT systems and infrastructure.
DP6: Ensure access to infrastructure developers during analytics initiatives	To prevent infrastructure adjustments from slowing down demonstrators, ensure access to infrastructure developers during analytics initiatives.

CHAPTER 5. DISCUSSION

In this chapter, I discuss the process and outcome of the research. First, I discuss how the findings of my research contribute to understanding how manufacturers should conduct process innovation with analytics and develop managerial prescriptions along the way to satisfy the second research objective. I then move to a discussion of the methodology, where I comment on the quality of the research. Lastly, I discuss the contributions and implications of the research.

5.1. HOW SHOULD MANUFACTURERS CONDUCT PROCESS INNOVATION WITH ANALYTICS?

I now turn to addressing the remainder of the second research objective by drawing on my findings to develop managerial prescriptions. I first discuss what the research findings suggest regarding how to manage process innovation with analytics activities, covering prioritization, organization, and execution. I then turn to how IT in manufacturing organizations should support process innovation with analytics.

5.1.1. MANAGING PROCESS INNOVATION WITH ANALYTICS

How should managers approach process innovation with analytics? Does it differ from the management of analytics more generally or process innovation in operations with other technologies? Recent research highlights that a key characteristic of analytics is that it, along with many other digital technologies, is a general-purpose technology (May et al., 2020; Eley & Lyytinen, 2022). To innovate with general-purpose technologies, organizations must first identify use cases and then explore their potential through experiments and pilots before deciding whether and where to invest in further scaled adoption (Rosemann, 2014; Maghazei et al., 2022). This thesis took outset in this need for exploration of the potentials of analytics and, through participation in eight analytics initiatives at the industrial partner, sought to answer how it should be carried out. Two of the key findings that emerged are that 1) analytics for process innovation is more complex than traditional analytics, and 2) analytics initiatives can differ widely in their content and complexity. Given these differences, I argue that managers should adopt a differentiated approach towards process innovation with analytics.

First, I propose that managers of process innovation with analytics need to understand and assess the infrastructure complexity of use cases before prioritizing larger investments in their development. This forces managers to do an initial pre-screening of initiatives and ensure enough clarity on how the analytics system will interact with other IT systems to deliver its output. This pre-screening allows for early detection of use cases where significant IT infrastructure development would be required to enable piloting. As noted by Vial et al. (2021) and Davenport & Malone (2021), many organizations do not consider such data accessibility or deployment concerns before much later in the project, resulting in projects that stall out and take a long time to reach the pilot deployment stage. They both recommend that deployment aspects be considered throughout the project (Vial et al., 2021; Davenport & Malone, 2021). My findings showed that data and functionality accessibility of the IT infrastructure is a key concern with implications for both the feasibility and organization of initiatives. In particular, my research suggests that it is not just whether data is machine-accessible that matters, as suggested by Vial et al. (2021), but also whether it is accessible in a loosely coupled manner. Understanding the infrastructural complexity at the initial stage will enable better prioritization through a better assessment of the *feasibility* (Vial et al., 2023) or *ease of implementation* (Hindle & Vidgen, 2018) of the initiative, which can then be traded off vs. potential business value. Furthermore, it can contribute to ensuring that initiatives are organized appropriately, which brings me to the second proposal.

Second, I propose that managers adopt a two-pronged approach to the organization of process innovation with analytics that distinguishes between heavyweight and lightweight analytics. An overview of the characteristics of heavyweight and lightweight analytics is presented in Table 5.1. In *Paper 3*, I established that even complex analytics initiatives using ML can be run successfully as *lightweight IT* (Bygstad, 2017), given the right circumstances and development strategy. In *lightweight* demonstrators, the focus is on the *technology-in-use* (Orlikowski, 1995) through iterative development and deployment of the analytics system to gather user feedback. Modeling and data science thus take a backseat and become the means to an end rather than the end itself. The lightweight approach is optimal from a process innovation point-of-view (Bygstad, 2017; Bygstad & Øvrelid, 2020), as piloting the technology in use allows for quick learning about the feasibility and value of the use case. It furthermore optimizes for the mutual adaptation of technology and process that is often necessary to adopt technology successfully (Leonard-Barton, 1988). The *lightweight demonstrator* thus presents a way of realizing the iterative design of analytics-enabled work that has been proposed in work on AI implementation (Tarafdar et al., 2017; Davenport, 2018b, p. 57). The viability of the lightweight demonstrator, however, hinges on a fast development and deployment loop. As such, it should be adopted when initiatives can be carried out with loose coupling to the IT infrastructure and the analytics complexity is low or medium.

When infrastructure or analytics complexity is high, then the heavyweight analytics R&D approach should instead be adopted. This is essentially the traditional *data science* or analytics approach, where the focus is on experimentation to define and solve the analytical problem, i.e., developing and validating a useful analytics artifact. The development is shielded from infrastructure complexity by working with batch extracts of data from the IT infrastructure at the cost of reducing the scope to models or system prototypes with static data. As recommended by Vial et al. (2021), IT can work on developing reusable infrastructure that enables piloting and deployment of the use case while analytics R&D is carried out. Ultimately, IT infrastructure might still become the bottleneck to deployment of a pilot, but at least the infrastructure work is started earlier this way.

Table 5.1 Comparison of Analytics R&D and Analytics Demonstrators

	Heavyweight Analytics R&D	**Lightweight Demonstrators**
Focus	Defining the analytical problem and developing a model with sufficient analytical capability.	Exploring the potential of analytics through development of the analytics system with feedback from actual use
Logic	Experimentation	Developmental Piloting
Optimize for	Modeling iterations	Feedback from use
Key participants	Data scientist (Model Developers), Domain Experts	Data scientist (Generalists), Software Engineers, Users
Suitable when:	Analytical Complexity: Medium – High Coupling to IT Infrastructure: Tight	Analytical Complexity: Low - Medium Coupling to IT Infrastructure: Loose

Third, I propose that managers tackle complexity incrementally by adopting a bias towards simplicity and small-scale initiatives in their exploration initiatives. Whereas managers should dream big when it comes to process innovation with analytics, lessons from both IT infrastructure (Ciborra et al., 2000; Hanseth & Lyytinen, 2010; Aanestad & Jensen, 2011), BPR (Davenport, 1995; Stoddard & Jarvenpaa, 1995), Industry 4.0 (Frank et al., 2019), analytics (Dremel et al., 2017; Shollo et al., 2022), and project management more generally (Flyvbjerg, 2021) suggests that implementation is better approached incrementally and using modularity. My findings suggest that this remains true for process innovation with analytics, even when organizations are at the stage of exploring the technology. In process innovation with analytics, even getting to the piloting stage can be a long journey and require significant work. By prioritizing analytics use cases with large complexity on both analytical and infrastructural dimensions, organizations risk becoming swamped with complexity and losing momentum and support before reaching the pilot stage. If the initiative ultimately loses support or fails to deliver on its ambitions, the organization is often left with only learning to show for its efforts. Managers should thus be careful to balance analytical and infrastructure complexity in the use cases they prioritize. Recent research on successful ML adopters has shown an incremental adoption pattern, starting with analytics for insights, moving onto augmentation use cases, and only sometimes towards automation (Shollo et al., 2022). This pattern aligns with increasing complexity of both the analytical and infrastructural complexity. The conceptualization of analytics development as the development and recombination of building blocks in *Paper 3* allows for a simple argument in favor of the incremental approach: the building blocks developed in simpler initiatives can be reused as modules in more complex initiatives. This is well-illustrated in *Paper 4,* where the final analytics demonstrator initially focused on the development of data pipelines and visualization (insights) before progressing to anomaly detection (augmentation), which leveraged the same data pipelines and visualization modules developed in the first iteration. In this way, value and working solutions can be delivered incrementally while allowing for the necessary building of capabilities in the organization. Table 5.2 summarizes the managerial prescriptions.

Table 5.2 Managerial prescriptions for management of process innovation with analytics

Aim	**Actions**	**Rationale**
To increase the success rate of adoption in process innovation with analytics	Understand and assess infrastructure complexity of use cases before prioritization.	Infrastructure complexity has large implications for the success rate of initiatives and how they should be organized
	Establish a two-pronged approach to initiatives: heavyweight and lightweight. Prioritize loosely coupled initiatives within the lightweight approach.	Take advantage of the loosely coupled infrastructure in the lightweight approach to quickly obtain feedback from use. Tackle complex initiatives by splitting infrastructure and R&D work in the heavyweight approach.
	Tackle complexity incrementally: Dream big, pilot incrementally	Smaller and simpler initiatives have a higher success rate and can build the capabilities, technical components, and momentum that enables tackling more complex initiatives later.

5.1.2. INFRASTRUCTURING FOR INNOVATION

As established in the preceding discussion, IT infrastructure plays an important role in process innovation with analytics and has significant implications for how innovation should be organized. Given the above prescriptions, how should IT managers respond and provide an IT infrastructure that supports rather than constrains the innovation process? In existing research, the socio-technical architecture and governance configuration of the infrastructure has been identified as playing a key role in the ability to support innovation (Bygstad & Øvrelid, 2020; Hanseth & Modol, 2021). In particular, loose coupling and decentralized (Henfridsson & Bygstad, 2013) or platform-based control (Bygstad & Hanseth, 2018) have been proposed as facilitating innovation. Similar findings emerged in my research, where the IT infrastructure, to a large extent, determined the ability to conduct fast iterations of analytics development and deployment. I thus propose that manufacturing organizations looking to adopt process innovation with analytics at a larger scale should develop an appropriate infrastructure to support this effort. Having the right infrastructure in place will enable more initiatives to be run as *lightweight demonstrators*, thus allowing organizations to assess and pilot use cases faster.

First, I propose that manufacturing organizations should leverage structural ambidexterity or bimodal IT in the IT organization to provide a dedicated unit to run *lightweight demonstrators* for operations. Bimodal IT as a concept is far from new (Gartner, 2014a; Haffke et al., 2017). However, so far, it has mainly been used to suggest that customer-focused or "new business"-generating IT be separated structurally from IT focused on carrying out work – *the operational backbone* (Ross et al., 2019). I suggest, on the other hand, that bimodal IT is also applied within operations to ensure that work focused on incremental development of the operational backbone (exploitative) and work focused on exploration of the potential of new technologies to change the way operational activities are conducted (exploration) remain separated. At the industrial partner, this was the case for the first part of the research process (see Section 3.1), where IT product teams handled exploitative development, whereas a dedicated unit supported longer-term discovery or exploration projects. After the removal of the dedicated unit, both IT management and Operations Transformation felt that innovation decreased, as exploration activities ended up not being prioritized within IT product teams. The findings are in line with the suggestion of Bygstad (2017) to ensure that heavyweight IT (i.e., process execution and infrastructure) and lightweight IT (innovation, process-support user-focused) remain loosely coupled organizationally. It is important to emphasize *loose coupling*, as the exploratory and exploitative units should still interact to facilitate the transfer of 1) knowledge between the units and 2) successful exploration projects to the exploitative units. Relying on outsourcing or external partners for the exploratory part alone makes achieving appropriate levels of interaction harder in practice.

Second, I propose that IT in manufacturing organizations should provide boundary resources to open up the core operational IT infrastructure for innovation initiatives. Boundary resources that provide access to data and functionality of the operational IT infrastructure are necessary to achieve organizational loose coupling of exploration and exploitation (Bygstad & Øvrelid, 2020). My findings showed that many challenges related to conducting fast, loosely coupled exploration were ultimately due to a lack of boundary resources that facilitated technical loose coupling. In the initiatives where appropriate boundary resources such as data pipelines, message queues, and APIs were present, innovation could occur freely with little interaction with exploitative teams. Different boundary resources enabled loose coupling in different development activities: ETL pipelines or on-demand batch extraction provided the ability to do model development, read-APIs, message buses, and message queues provided the ability to deploy augmentation systems, whereas write-APIs were necessary for automation use-cases. The presence of boundary resources provided benefits not just for the innovation initiatives that could occur faster (Bygstad, 2017) but also for the exploitative teams, which did not have to spend time doing ad-hoc extraction of data or developing project-specific integration points. It should be emphasized that merely providing a technical resource is often insufficient to make it a true boundary resource. Unless appropriate documentation and procedures for using the boundary resources are in place, potential users of the boundary resources will need to interact considerably with the owners. Providing boundary resources for key data and IT functionality will ultimately require a transformation of the infrastructure in most manufacturing organizations, as legacy systems rarely provide the necessary boundary resources. Rather than requiring a replacement of legacy systems, a more pragmatic strategy is to develop boundary resources for existing legacy systems (Bygstad & Hanseth, 2018; Weill et al., 2020). As mentioned above, an ideal time to undertake such projects is when the pre-screening of analytics initiatives identifies a major infrastructural gap. At the same time, IT managers should ensure that new systems provide boundary resources, e.g., by making them part of the enterprise architecture principles (Haki et al., 2021).

Third, I propose that IT in manufacturing organizations should develop a data platform that supports analytics development and deployment. Modern analytics systems can be highly complex, with many moving parts (Sculley et al., 2015). Other industrial research has also suggested the need for platforms to support Industry 4.0 analytics (Bonnard et al., 2021; Gröger, 2018; Gröger, 2021), arguing that it leads to a cleaner architecture, reuse of analytical assets, better governance, and enables democratization of data science (Gröger, 2018; Gröger, 2021). My findings corroborate that providing a data platform for development can allow analytics initiatives to focus on the components that create value for a particular use case rather than infrastructure. Leveraging a data platform for development and deployment, allowed for on-demand access to infrastructure for training models and deploying data pipelines, and even standardized deployment of analytical models as an API. This makes deployment of models considerably more feasible for data scientists without major involvement from IT engineers. As was the case with the operational IT infrastructure, it is necessary that the data platform offers boundary resources that enable developers to develop and deploy data, models, and system components loosely coupled from the data platform infrastructure team. Without these boundary resources, the data platform team will instead become the bottleneck for innovation. One key challenge for manufacturing, as compared to other functions, is that the data platform needs to provide capabilities for deployment in both the cloud and at the edge to support the full range of Industry 4.0 use cases.

Finally, it should be emphasized that the above infrastructure does not only enable the organization to carry out process innovation with analytics but will also support IT-enabled process innovation in general and analytics for insights. As such, it makes up part of the foundation necessary to support a broader digital transformation of manufacturing. A summary of the managerial prescriptions is provided in Table 5.3.

Table 5.3 Managerial prescriptions for IT infrastructure

Aim	**Action**	**Rationale**
To develop an infrastructure that enables process innovation with analytics	Leverage structural ambidexterity in IT through a dedicated operations exploration unit.	Exploitation often ends up being prioritized over exploration if not separated, leading to a lack of explorative resources.
	Provide boundary resources to open up the core operational IT infrastructure for innovation	Boundary resources allow exploration initiatives to be run loosely coupled to ordinary development, speeding up innovation.
	Provide a data platform for analytics development and deployment.	The data platform provides standards and IT capabilities for developing and deploying models, thus reducing project-level efforts to develop and deploy models.

5.2. METHODOLOGY & RESEARCH QUALITY

No research is perfect, and it thus seems appropriate to comment on the quality of the research. Applied research such as DSR is subjected to the dual requirements of research rigor and relevance (Hevner et al., 2004; van Aken et al., 2016; Baskerville et al., 2018). The close collaboration with the industrial partner and the orientation towards solving the problems they were facing provide strong evidence for practical relevance. I will thus limit the following discussion to the question of research rigor. van Aken et al. (2016) suggest that the quality of DSR research be assessed in terms of 1) pragmatic validity and 2) the quality of the explanatory component. Pragmatic validity is established through testing and concerns whether the artifact works and produces the intended effects (van Aken et al., 2016). The quality of the explanatory component concerns the explanation offered of the mechanisms producing the observed effects, and it is judged according to traditional criteria for explanatory research (van Aken et al., 2016).

In terms of pragmatic validity, the key question is whether 1) the design principles and approach developed improve development speed of analytics demonstrators and 2) whether the managerial prescriptions developed facilitate adoption of analytics for process innovation. For the design principles, the fact that they were successfully instantiated in a separate demonstrator in their natural context provides evidence that they are usable and work as intended. However, owing to the use of natural evaluation, varying resource availability in the demonstrators, and other project-specific differences, it has not been meaningful to establish quantitative estimates of the development speed improvement. Qualitatively, the development speed was fast and facilitated multiple iterations of working prototypes, with iterations taking weeks rather than months. This evidence, in combination with the grounding of the design principles in existing literature, provides support for the pragmatic validity of the design principles and approach. Regarding the managerial prescriptions, these remain unevaluated and should thus be viewed as propositions. While grounded in both experience and existing literature, evidence supporting their pragmatic validity remains somewhat limited.

When it comes to the quality of the explanatory component, the main concern is the quality of the explanations for development speed differences in *Paper 3*. In the paper, 1) loose coupling between the operational digital infrastructure and the demonstrator and 2) the extent of use of high-level building blocks are offered as key explanations for differences in development speed in analytics demonstrators. As suggested by van Aken (2016), my co-authors and I relied on cross-case analysis to establish the explanation. We relied on extensive empirical material in the form of documentation and field notes and subjected the case analysis to best practices within abductive or retroductive case analysis, primarily from the tradition of critical realist case study research (Bygstad et al., 2016). This included, amongst others, a transparent analysis process and assessing the potential for other explanatory mechanisms. Nonetheless, the identified mechanisms are necessarily partial, owing to the complexity of development projects. The reliance on participative observation and action as main data sources introduces the potential for researcher bias. This was counterbalanced by the detachedness of my co-authors, who were not involved in the overall research project and thus were able to challenge my interpretations. The credibility of the explanations is furthermore corroborated by the success achieved in the final demonstrator, which relied on the mechanisms to guide design and action.

5.3. CONTRIBUTIONS TO RESEARCH

The thesis and the appended papers collectively contribute to research on analytics, digital process innovation, lightweight IT, and action-oriented DSR. Table 5.4 provides an overview of the contributions to each of these three areas of research.

Table 5.4 Highlights of the main contributions to research

Area of Research	**Contribution**
Analytics	New development approach & design principles Conceptualization of analytics within a broader scope Research agenda for process innovation with analytics
Digital Process Innovation (OM & IS)	Insights into organization & management of analytics-based Process Innovation Research agenda for process innovation with analytics
Lightweight IT	Insights into the boundary of the concept
Action-oriented DSR	Illustration of how ADR can be adapted to balance academic and practical contribution

The first contribution of the research is the development approach and design principles developed in *Paper 4*. These contribute to the literature on analytics development, which has received some attention recently following years without major developments (Hindle & Vidgen, 2018). In particular, the design principles and approach consider socio-technical aspects of development and highlight the need to contextualize development according to characteristics of the infrastructure, thus moving beyond the task-focus and infrastructure-agnosticism present in the widely adopted CRISP-DM methodology (Wirth & Hipp, 2000).

The second and third contributions are the framework that reconceptualizes analytics in the context of process innovation and the associated research agenda developed in *Paper 2*. Analytics is a highly dynamic phenomenon due to innovations in both technologies and use contexts (Davenport, 2018a). The framework and research agenda contribute to the organizational literature on analytics (e.g., Dremel et al., 2017; Dremel et al., 2020; Tim et al., 2020; Mikalef & Krogstie, 2020; May et al., 2020; Kunz et al., 2022; Shollo et al., 2022) by providing an updated foundation for research on the increasingly prevalent process innovation use of analytics. The framework, in particular, has implications for empirical studies of AI, ML, and analytics adoption by highlighting the need to account for and examine differences in infrastructural and process change complexity.

The fourth contribution lies in the insights developed across the papers into the management and organization of analytics-based process innovation. These insights have implications for research on exploratory digital process innovation in both technology management discourses in OM and IS research. In IS, the research contributes to the multiple calls for research into the management of the latest installment of analytics, namely AI (Benbya et al., 2021; Berente et al., 2021). In particular, I provide insights into the architecture and governance configurations that support augmentation and automation (Benbya et al., 2021). My findings furthermore highlight the role of design and development in realizing the process innovation affordances afforded by analytics. This has implications for IS research leveraging the affordance perspective to study value realization with analytics and AI (e.g., Dremel et al., 2020; Tim et al., 2020), which often tends to downplay the role of design owing to the perspective's origin in the study of packaged software. The insights furthermore contribute to the explorative BPM discourse (Rosemann, 2014; Grisold et al., 2019; Baier et al., 2022) by demonstrating how lightweight IT can speed up the assessment and technical implementation of new non-standard technologies. In OM, the findings contribute to the discourse on adoption and management of Industry 4.0 (Maghazei et al., 2022; Eley & Lyytinen, 2022). As noted by Eley & Lyytinen (2022), there is currently a lack of adoption research that examines the *process* aspect of adoption of Industry 4.0 technologies. The longitudinal study of drone adoption by Maghazei et al. (2022) is a notable exception, which highlights the role of use cases and piloting in innovation. My findings similarly confirm the pivotal role of pilots and use cases, but as compared to drones, suggest that adoption and piloting of process innovation with analytics places greater demands on IT infrastructure and IT and operations collaboration.

The fifth contribution, I argue, is to further the work on lightweight IT. The research achieves this by contributing with insights into the boundaries of the concept. As noted in *Paper 3*, one of the motivations for the study was to investigate the applicability of lightweight IT to machine learning as a technology. Bygstad (2017) conceptualized lightweight IT and heavyweight IT, but ML did not fit neatly into either of the two categories. On the one hand, the development culture around ML is associated with experimentation and focused on improving processes or delivering insights to users, as is characteristic of lightweight IT. On the other hand, it is a complex technology requiring significant development expertise and integration with existing IT systems to function, as is characteristic of heavyweight IT. My findings suggest that it is not an either-or question but rather that ML can belong to both classes depending on the specifics of the technology and its application. To the extent that ML is used to automate processes and relies on extensive custom development, it is reminiscent of heavyweight IT. If, on the other hand, it is used to augment work and relies on more or less standard technologies, it should be viewed and treated as lightweight IT. These findings help establish the boundaries of the concept and contribute to the body of research applying the lightweight IT concept, which has so far been focused on apps and digital whiteboards for process support (e.g., Bygstad, 2017; Bygstad & Øvrelid, 2020) and robotic process automation (e.g., Penttinen et al., 2018; Osmundsen et al., 2019; Herm et al., 2023).

The sixth and final contribution of the thesis is as an example of how ADR can be adapted to achieve a balance between contributions to practice and academia. As concluded in *Paper 1* and discussed in *Chapter 2*, there has yet to be a consensus or "epistemic script" (Grover & Lyytinen, 2015) in the IS and OM communities regarding how to design contributions from action-oriented DSR. Whereas the initial focus in terms of academic contribution has been on the novelty and utility of the artifact and generalized design knowledge (e.g., Gregor & Hevner, 2013), others are pointing to the rich empirical data gathered in design and action, in combination with theory, as holding the potential for contribution (e.g., Oliva, 2019; Goldkuhl & Sjöström, 2021). This thesis has tried to strike a balance between a focus on the artifact and associated design principles and other contributions taking the form of an embedded case study, a conceptual framework, and a research agenda. These contributions would not have been possible without the wider framing of the demonstrators, or BIE cycles (Sein et al., 2011), as vehicles for data collection on the technology in its organizational context. This use of action-oriented DSR is in many ways close to what Braa & Vidgen (1999) termed *action cases*, which are case studies that attempt to both change and understand the problem situation, with the main difference being the additional emphasis on design in this research. In terms of practical utility, the thesis provides another illustration of how DSR projects can play a role in digital innovation initiatives at organizations (Chen et al., 2022). The research design used allowed for contributing concretely to the innovation agenda at the industrial partner through the assessment of use cases while simultaneously providing insights for managers on how to organize and manage the digital innovation initiatives.

CHAPTER 6. CONCLUSION

Many manufacturing organizations are looking towards analytics as a means for process improvement and innovation. While many promising applications have been reported on in the media and in research, adoption and implementation are proving to be a challenge. This thesis set out to improve our understanding of process innovation with analytics and develop prescriptive knowledge that supports organizations in adoption. The overall methodology adopted to achieve this was Action Design Research, which was used to develop understanding and prescriptive knowledge through the design of analytics demonstrators at a large Danish manufacturer and retailer. This engaged work at the industrial partner consisted of participative observation in three analytics initiatives and the development of five analytics demonstrators.

The first objective of improving understanding of process innovation with analytics was met by developing a framework and research agenda in *Paper 2* and through a case study of analytics demonstrators in *Paper 3*. A key finding was that process innovation with analytics differs from traditional analytics for insights in that it requires coordinated development of digital infrastructure, analytics, and processes, and its scope is thus significantly greater. Furthermore, the use and maturity of solution building blocks and a loosely coupled digital infrastructure were identified as key factors influencing the speed of exploration and piloting of analytics for process innovation.

The second objective of developing prescriptive knowledge for process innovation with analytics was achieved by extracting and evaluating design principles from the analytics demonstrators in *Paper 4* and developing managerial prescriptions based on the lessons learned in section 5.1. The findings highlight the importance of assessing and understanding the infrastructure needs of initiatives early on as an input to both prioritization and organization of initiatives. Prioritization should favor exploration of initiatives that can be run within the existing infrastructure in a loosely coupled fashion and broken down to deal with complexity in an incremental and modular fashion. These initiatives should be run as *lightweight analytics demonstrators* focusing on quickly developing and piloting the analytics systems. When infrastructure development is required, it should be separated from but coordinated with the analytics work and focus on developing reusable boundary resources.

Overall, the findings of the thesis reveal the importance of having the right IT infrastructure in place to enable manufacturing organizations to achieve wider-scale adoption of analytics. However, while IT infrastructure does play a crucial role in enabling development, deployment, and piloting of analytics technologies to improve processes, it is not enough on its own. Leveraging the infrastructure to speed up innovation requires a development approach that takes advantage of the infrastructure:

the *analytics demonstrator*. It furthermore requires a presence of vision, skills, willingness, and capacity in operations to explore and pilot analytics systems. Without these factors in place, the faster development afforded will lead to the IT and innovation organization piling up analytical models and systems that lack the feedback-from-use required to successfully adopt the technology for process innovation.

The findings presented in the dissertation have implications for large manufacturers looking to adopt analytics for process innovation. Several large Danish manufacturers are on this journey, and it is my hope that the findings prove useful for them, as success in digitalization is likely to be important in ensuring the future competitiveness of the Danish manufacturing industry.

6.1. LIMITATIONS

As is always the case, the research presented in this thesis is subject to limitations. The most obvious limitation is that the research was conducted with a single organization, and thus, the question of generalizability remains a concern. A key challenge that emerged in the research was dealing with the installed base of technology. This was the case even though the industrial partner is considered to be amongst the leaders in IT in the Danish manufacturing industry. It thus seems likely that the strategies developed to deal with a less-than-ideal installed base will remain useful in other companies where the impact of the installed base is even greater. While grounding the knowledge claims in existing theory does strengthen the potential for generalizability, the knowledge must ultimately be put to the test in other empirical contexts. As a start, this should include contexts that fall within the class of problems the research addresses, namely large established manufacturers adopting analytics for process innovation, but it might also be relevant to test the findings in other contexts characterized by complex technology infrastructures.

Another limitation concerns the practical (re-)usability of the design knowledge generated. While the design knowledge generated has been discussed with various practitioners at the industrial partner, the formalized version has so far only been instantiated by me. A valid concern is thus whether practitioners find the formalized knowledge usable and valuable (Iivari et al., 2021). Owing to practical and time constraints, this has not yet been addressed and future work thus remains to assess and evaluate the formalized design knowledge with practitioners.

6.2. FUTURE RESEARCH

As is evident from the research agenda proposed in the second appended paper, the research presented in this thesis has only started to scratch the surface of process innovation with analytics. Ultimately, the research has generated more questions than it has answered. I end by highlighting a few of these questions which I find particularly interesting. First, I suggest the need for research that furthers our understanding of how to consider process concerns in analytics development and analytics concerns in process development. This could be accomplished both by descriptive case studies of successful innovators and prescriptive approaches such as DSR.

Second, I suggest further research into innovation infrastructures for process innovation with analytics. Which infrastructure configurations are generative of process innovation with analytics? I proposed one promising configuration based on my experiences at the industrial partner, focusing in particular on the technical and organizational aspects of the configuration. Further research could fruitfully investigate successful configurations in a variety of contexts using multiple case studies. This research should investigate the effects of the actors involved in the innovation process, including aspects such as open innovation, as this was an important dimension that I did not address substantially in my research. Additionally, it should explore how the infrastructure impacts the later stages of the innovation process, including wider-scale implementation.

Furthermore, it would be of great use to senior managers to understand whether and when process innovation with analytics provides a positive return on investment. Is it worth pursuing as compared to other process improvement programs? Investments in process innovation with analytics are currently largely based on anecdotal evidence and beliefs rather than evidence. Much work remains to be done in this direction, including qualitative, configurational, and survey-based research.

LITERATURE LIST

Aanestad, M., Grisot, M., Hanseth, O., & Vassilakopoulou, P. (Eds.). (2017). Information Infrastructures and the Challenge of the Installed Base. In Information Infrastructures within European Health Care (pp. 25–33). Springer International Publishing. https://doi.org/10.1007/978-3-319-51020-0

Aanestad, M., & Jensen, T. B. (2011). Building nation-wide information infrastructures in healthcare through modular implementation strategies. The Journal of Strategic Information Systems, 20(2), 161–176. https://doi.org/10.1016/j.jsis.2011.03.006

Ågerfalk, P. J. (2010). Getting pragmatic. European Journal of Information Systems, 19(3), 251–256. https://doi.org/10.1057/ejis.2010.22

Badakhshan, P., Wurm, B., Grisold, T., Geyer-Klingeberg, J., Mendling, J., & vom Brocke, J. (2022). Creating business value with process mining. The Journal of Strategic Information Systems, 31(4), 101745. https://doi.org/10.1016/j.jsis.2022.101745

Baier, M.-S., Lockl, J., Röglinger, M., & Weidlich, R. (2022). Success factors of process digitalization projects – insights from an exploratory study. Business Process Management Journal, 28(2), 325–347. https://doi.org/10.1108/BPMJ-07-2021-0484

Baiyere, A., Salmela, H., & Tapanainen, T. (2020). Digital transformation and the new logics of business process management. European Journal of Information Systems, 29(3), 238–259. https://doi.org/10.1080/0960085X.2020.1718007

Baskerville, R., Baiyere, A., Gregor, S., Hevner, A., & Rossi, M. (2018). Design science research contributions: Finding a balance between artifact and theory. Journal of the Association for Information Systems, 19(5).

Benbya, H., Pachidi, S., & Jarvenpaa, S. L. (2021). Special Issue Editorial: Artificial Intelligence in Organizations: Implications for Information Systems Research. Journal of the Association for Information Systems, 22(2), 281–303. https://doi.org/10.17705/1jais.00662

Benner, M. J., & Tushman, M. L. (2003). Exploitation, Exploration, and Process Management: The Productivity Dilemma Revisited. The Academy of Management Review, 28(2), 238–256. https://doi.org/10.2307/30040711

Berente, N., Gu, B., Recker, J., & Santhanam, R. (2021). Managing artificial intelligence. MIS Quarterly, 45(3), 1433–1450.

Beverungen, D. (2014). Exploring the Interplay of the Design and Emergence of Business Processes as Organizational Routines. Business & Information Systems Engineering, 6(4), 191–202. https://doi.org/10.1007/s12599-014-0335-3

Boer, H., Holweg, M., Kilduff, M., Pagell, M., Schmenner, R., & Voss, C. (2015). Making a meaningful contribution to theory. International Journal of Operations & Production Management, 35(9), 1231–1252. https://doi.org/10.1108/IJOPM-03-2015-0119

Bojer, C. S., Bygstad, B., & Øvrelid, E. (2023). Speeding up Explorative BPM with Lightweight IT: The Case of Machine Learning. [Manuscript submitted for publication].

Bojer, C. S., & Møller, C. (2022). Towards a Scheme for Contribution in Action Design Research. In A. Drechsler, A. Gerber, & A. Hevner (Eds.), The Transdisciplinary Reach of Design Science Research (pp. 376–387). Springer International Publishing.

Bojer, C. S., & Møller, C. (2023a). Conceptualizing Process Innovation with Analytics: A Pragmatic Framework and Research Agenda. [Manuscript submitted for publication].

Bojer, C. S., & Møller, C. (2023b). Developing Analytics Demonstrators for Process Innovation: An Infrastructural Perspective. [Manuscript submitted for publication].

Bonnard, R., Arantes, M. D. S., Lorbieski, R., Vieira, K. M. M., & Nunes, M. C. (2021). Big data/analytics platform for Industry 4.0 implementation in advanced manufacturing context. The International Journal of Advanced Manufacturing Technology, 117(5), 1959–1973. https://doi.org/10.1007/s00170-021-07834-5

Boston Consulting Group. (2015). Industry 4.0: The Future of Productivity and Growth in Manufacturing Industries. BCG Global. https://www.bcg.com/publications/2015/engineered_products_project_business_industry_4_future_productivity_growth_manufacturing_industries

Boston Consulting Group. (2023). BCG-WEF Project: AI-Powered Industrial Operations. BCG Global. https://www.bcg.com/about/partner-ecosystem/world-economic-forum/ai-project-survey

Braa, K., & Vidgen, R. (1999). Interpretation, intervention, and reduction in the organizational laboratory: A framework for in-context information system research. Accounting, Management and Information Technologies, 9(1), 25–47. https://doi.org/10.1016/S0959-8022(98)00018-6

Broadbent, M., Weill, P., & St. Clair, D. (1999). The Implications of Information Technology Infrastructure for Business Process Redesign. MIS Quarterly, 23(2), 159–182. https://doi.org/10.2307/249750

Bygstad, B. (2010). Generative mechanisms for innovation in information infrastructures. Information and Organization, 20(3), 156–168. https://doi.org/10.1016/j.infoandorg.2010.07.001

Bygstad, B. (2017). Generative Innovation: A Comparison of Lightweight and Heavyweight IT. Journal of Information Technology, 32(2), 180–193. https://doi.org/10.1057/jit.2016.15

Bygstad, B., & Hanseth, O. (2018). Transforming Digital Infrastructures through Platformization. ECIS 2018 Proceedings. Twenty-Sixth European Conference on Information Systems.

Bygstad, B., & Iden, J. (2017). A Governance Model for Managing Lightweight IT. In Á. Rocha, A. M. Correia, H. Adeli, L. P. Reis, & S. Costanzo (Eds.), Recent Advances in Information Systems and Technologies (pp. 384–393). Springer International Publishing. https://doi.org/10.1007/978-3-319-56535-4_39

Bygstad, B., Munkvold, B. E., & Volkoff, O. (2016). Identifying Generative Mechanisms through Affordances: A Framework for Critical Realist Data Analysis. Journal of Information Technology, 31(1), 83–96. https://doi.org/10.1057/jit.2015.13

Bygstad, B., & Øvrelid, E. (2020). Architectural alignment of process innovation and digital infrastructure in a high-tech hospital. European Journal of Information Systems, 29(3), 220–237. https://doi.org/10.1080/0960085X.2020.1728201

Chang, H. (2022). Realism for Realistic People. Cambridge University Press.

Chen, K.-L., Lassen, A., Li, C., & Møller, C. (2022). Exploring the value of IoT data as an enabler of the transformation towards servitization: An action design research approach. European Journal of Information Systems, 32(4), 735–761.

Chirumalla, K. (2021). Building digitally-enabled process innovation in the process industries: A dynamic capabilities approach. Technovation, 105, 102256. https://doi.org/10.1016/j.technovation.2021.102256

Ciborra, C., Braa, K., Cordella, A., Dahlbom, B., Hepsø, V., Failla, A., Hanseth, O., & others. (2000). From control to drift: The dynamics of corporate information infastructures. Oxford University Press.

Davenport, T. H. (1993). Process Innovation. Mcgraw-Hill.

Davenport, T. H. (1995). Business process reengineering: Where it's been, where it's going. In V. Grover & W. J. Kettinger (Eds.), Business Process Change: Reengineering concepts, methods and technologies (pp. 1–13). Idea Group Publishing.

Davenport, T. H. (2018a). From analytics to artificial intelligence. Journal of Business Analytics, 1(2), 73–80. https://doi.org/10.1080/2573234X.2018.1543535

Davenport, T. H. (2018b). The AI Advantage: How to Put the Artificial Intelligence Revolution to Work. MIT Press.

Davenport, T. H., Harris, J. G., & Morison, R. (2010). Analytics at work: Smarter decisions, better results. Harvard Business Press.

Davenport, T. H., & Miller, S. M. (2022). Working with AI: Real Stories of Human-machine Collaboration. MIT Press.

Davenport, T. H., & Short, J. E. (1990). The new industrial engineering: Information technology and business process redesign. Sloan Management Review, 31(4), 11–27.

Davenport, T., & Harris, J. (2017). Competing on Analytics: Updated, with a New Introduction: The New Science of Winning. Harvard Business Press.

Davenport, T., & Malone, K. (2021). Deployment as a Critical Business Data Science Discipline. Harvard Data Science Review, 3(1).

de Reuver, M., Sørensen, C., & Basole, R. C. (2018). The Digital Platform: A Research Agenda. Journal of Information Technology, 33(2), 124–135. https://doi.org/10.1057/s41265-016-0033-3

Dieste, M., Sauer, P. C., & Orzes, G. (2022). Organizational tensions in industry 4.0 implementation: A paradox theory approach. International Journal of Production Economics, 251, 108532. https://doi.org/10.1016/j.ijpe.2022.108532

Dolata, M., Crowston, K., & Schwabe, G. (2022). Project Archetypes: A Blessing and a Curse for AI Development. ICIS 2022 Proceedings. Forty-Third International Conference on Information Systems.

Dremel, C., Herterich, M. M., Wulf, J., & vom Brocke, J. (2020). Actualizing big data analytics affordances: A revelatory case study. Information & Management, 57(1), 103121. https://doi.org/10.1016/j.im.2018.10.007

Dremel, C., Herterich, M., Wulf, J., Waizmann, J.-C., & Brenner, W. (2017). How AUDI AG Established Big Data Analytics in its Digital Transformation. MIS Quarterly Executive, 16(2), 81–100.

Dubois, A., & Gadde, L.-E. (2002). Systematic combining: An abductive approach to case research. Journal of Business Research, 55(7), 553–560. https://doi.org/10.1016/S0148-2963(00)00195-8

Dumas, M., La Rosa, M., Mendling, J., & Reijers, H. A. (2013). Fundamentals of business process management (Vol. 1). Springer.

Edwards, P. N., Jackson, S. J., Bowker, G. C., & Knobel, C. P. (2007). Understanding infrastructure: Dynamics, tensions, and design. University of Michigan. http://hdl.handle.net/2027.42/49353

Eley, T., & Lyytinen, K. (2022). Industry 4.0 Implementation: Novel Issues and Directions. Proceedings of the 55th Hawaii International Conference on System Sciences. Hawaii International Conference on System Sciences. https://doi.org/10.24251/HICSS.2022.622

Enholm, I. M., Papagiannidis, E., Mikalef, P., & Krogstie, J. (2021). Artificial Intelligence and Business Value: A Literature Review. Information Systems Frontiers, 24, 1709–1734. https://doi.org/10.1007/s10796-021-10186-w

Farjoun, M., Ansell, C., & Boin, A. (2015). PERSPECTIVE—Pragmatism in Organization Studies: Meeting the Challenges of a Dynamic and Complex World. Organization Science, 26(6), 1787–1804. https://doi.org/10.1287/orsc.2015.1016

Flyvbjerg, B. (2021). Make megaprojects more modular. Harvard Business Review, November-December 2021, 58–63.

Frank, A. G., Dalenogare, L. S., & Ayala, N. F. (2019). Industry 4.0 technologies: Implementation patterns in manufacturing companies. International Journal of Production Economics, 210, 15–26. https://doi.org/10.1016/j.ijpe.2019.01.004

Gartner. (2014a). Bimodal IT: How to Be Digitally Agile Without Making a Mess. Gartner. https://www.gartner.com/en/documents/2798217

Gartner. (2014b). Gartner Says Advanced Analytics Is a Top Business Priority. Gartner. https://www.gartner.com/en/newsroom/press-releases/2014-10-21-gartner-says-advanced-analytics-is-a-top-business-priority

Ghazawneh, A., & Henfridsson, O. (2013). Balancing platform control and external contribution in third-party development: The boundary resources model. Information Systems Journal, 23(2), 173–192. https://doi.org/10.1111/j.1365-2575.2012.00406.x

Godfrey-Smith, P. (2021). Theory and Reality: An Introduction to the Philosophy of Science (Second Edition). University of Chicago Press.

Goldkuhl, G. (2012). Pragmatism vs interpretivism in qualitative information systems research. European Journal of Information Systems, 21(2), 135–146. https://doi.org/10.1057/ejis.2011.54

Goldkuhl, G., & Sjöström, J. (2021). Design science theorizing: The contribution of practical theory. In N. R. Hassan & L. P. Willcocks (Eds.), Advancing Information Systems Theories: Rationale and Processes (pp. 239–273). Springer.

Gregor, S., & Hevner, A. R. (2013). Positioning and presenting design science research for maximum impact. MIS Quarterly, 37(2), 337–355.

Gregor, S., & Jones, D. (2007). The Anatomy of a Design Theory. Journal of the Association for Information Systems, 8(5). https://doi.org/10.17705/1jais.00129

Gregor, S., Kruse, L. C., & Seidel, S. (2020). Research Perspectives: The Anatomy of a Design Principle. Journal of the Association for Information Systems, 21(6). https://doi.org/10.17705/1jais.00649

Grisold, T., Gross, S., Röglinger, M., Stelzl, K., & vom Brocke, J. (2019). Exploring Explorative BPM - Setting the Ground for Future Research. In T. Hildebrandt, B. F. van Dongen, M. Röglinger, & J. Mendling (Eds.), Business Process Management (Vol. 11675, pp. 23–31). Springer International Publishing. https://doi.org/10.1007/978-3-030-26619-6_4

Grisot, M., Hanseth, O., & Thorseng, A. (2014). Innovation Of, In, On Infrastructures: Articulating the Role of Architecture in Information Infrastructure Evolution. Journal of the Association for Information Systems, 15(4), 197–219. https://doi.org/10.17705/1jais.00357

Gröger, C. (2018). Building an Industry 4.0 Analytics Platform. Datenbank-Spektrum, 18(1), 5–14. https://doi.org/10.1007/s13222-018-0273-1

Gröger, C. (2021). There is no AI without data. Communications of the ACM, 64(11), 98–108. https://doi.org/10.1145/3448247

Gröger, C. (2022). Industrial analytics – An overview. It - Information Technology, 64(1–2), 55–65. https://doi.org/10.1515/itit-2021-0066

Grover, V., & Lyytinen, K. (2015). New State of Play in Information Systems Research: The Push to the Edges. MIS Quarterly, 39(2), 271–296.

Haffke, I., Kalgovas, B., & Benlian, A. (2017). Options for Transforming the IT Function Using Bimodal IT. MIS Quarterly Executive, 16(2), 101–120.

Haki, K., & Legner, C. (2021). The Mechanics of Enterprise Architecture Principles. Journal of the Association for Information Systems, 22(5), 1334–1375. https://doi.org/10.17705/1jais.00696

Hammer, M., & Champy, J. (1993). Reengineering the Corporation: A Manifesto for Business Revolution. Harper Business.

Hanseth, O., & Lyytinen, K. (2010). Design Theory for Dynamic Complexity in Information Infrastructures: The Case of Building Internet. Journal of Information Technology, 25(1), 1–19. https://doi.org/10.1057/jit.2009.19

Hanseth, O., & Modol, J. R. (2021). The dynamics of architecture-governance configurations: An assemblage theory approach. Journal of the Association for Information Systems, 22(1), 5.

Hanseth, O., & Monteiro, E. (1998). Understanding information infrastructure. Unpublished manuscript. https://www.researchgate.net/profile/Eric-Monteiro-2/publication/265066841_Understanding_Information_Infrastructure/links/546dee010cf2d5ae36708bd0/Understanding-Information-Infrastructure.pdf

Harmon, P. (2010). The Scope and Evolution of Business Process Management. In J. vom Brocke & M. Rosemann (Eds.), Handbook on Business Process Management 1: Introduction, Methods, and Information Systems (pp. 37–81). Springer. https://doi.org/10.1007/978-3-642-00416-2_3

Henfridsson, O., & Bygstad, B. (2013). The Generative Mechanisms of Digital Infrastructure Evolution. MIS Quarterly, 37(3), 907–931.

Henfridsson, O., Nandhakumar, J., Scarbrough, H., & Panourgias, N. (2018). Recombination in the open-ended value landscape of digital innovation. Information and Organization, 28(2), 89–100. https://doi.org/10.1016/j.infoandorg.2018.03.001

Herm, L.-V., Janiesch, C., Helm, A., Imgrund, F., Hofmann, A., & Winkelmann, A. (2023). A framework for implementing robotic process automation projects. Information Systems and E-Business Management, 21(1), 1–35. https://doi.org/10.1007/s10257-022-00553-8

Hertzum, M., Bansler, J. P., Havn, E. C., & Simonsen, J. (2012). Pilot Implementation: Learning from Field Tests in IS Development. Communications of the Association for Information Systems, 30(1). https://doi.org/10.17705/1CAIS.03020

Hevner, A. R. (2007). A three cycle view of design science research. Scandinavian Journal of Information Systems, 19(2).

Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design science in information systems research. MIS Quarterly, 28(1), 75–105.

Hindle, G. A., & Vidgen, R. (2018). Developing a business analytics methodology: A case study in the foodbank sector. European Journal of Operational Research, 268(3), 836–851. https://doi.org/10.1016/j.ejor.2017.06.031

Holsapple, C., Lee-Post, A., & Pakath, R. (2014). A unified foundation for business analytics. Decision Support Systems, 64, 130–141. https://doi.org/10.1016/j.dss.2014.05.013

Hopp, W. J., & Spearman, M. S. (2021). The lenses of lean: Visioning the science and practice of efficiency. Journal of Operations Management, 67(5), 610–626.

Hoyer, C., Gunawan, I., & Reaiche, C. H. (2020). The Implementation of Industry 4.0 – A Systematic Literature Review of the Key Factors. Systems Research and Behavioral Science, 37(4), 557–578. https://doi.org/10.1002/sres.2701

Huber, S., Wiemer, H., Schneider, D., & Ihlenfeldt, S. (2019). DMME: Data mining methodology for engineering applications – a holistic extension to the CRISP-DM model. Procedia CIRP, 79, 403–408. https://doi.org/10.1016/j.procir.2019.02.106

Iivari, J. (2015). Distinguishing and contrasting two strategies for design science research. European Journal of Information Systems, 24(1), 107–115.

Iivari, J., Rotvit Perlt Hansen, M., & Haj-Bolouri, A. (2021). A proposal for minimum reusability evaluation of design principles. European Journal of Information Systems, 30(3), 286–303.

Kagermann, H., Wahlster, W., & Helbig, J. (2013). Recommendations for implementing the strategic initiative INDUSTRIE 4.0: Securing the future of German manufacturing industry; final report of the Industrie 4.0 Working Group. acatech. https://en.acatech.de/publication/recommendations-for-implementing-the-strategic-initiative-industrie-4-0-final-report-of-the-industrie-4-0-working-group/

Kim, G., Shin, B., Kim, K., & Lee, H. (2011). IT Capabilities, Process-Oriented Dynamic Capabilities, and Firm Financial Performance. Journal of the Association for Information Systems, 12(7), 487–517. https://doi.org/10.17705/1jais.00270

Koutsikouri, D., Lindgren, R., Henfridsson, O., & Rudmark, D. (2018). Extending Digital Infrastructures: A Typology of Growth Tactics. Journal of the Association for Information Systems, 19(10), 1001–1019. https://doi.org/10.17705/1jais.00517

Kunz, P. C., Jussupow, E., Spohrer, K., & Heinzl, A. (2022). How the Application of Machine Learning Systems Changes Business Processes: A Multiple Case Study. ECIS 2022 Proceedings. Thirtieth European Conference on Information Systems.

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

Lavin, A., Gilligan-Lee, C. M., Visnjic, A., Ganju, S., Newman, D., Ganguly, S., Lange, D., Baydin, A. G., Sharma, A., Gibson, A., Zheng, S., Xing, E. P., Mattmann, C., Parr, J., & Gal, Y. (2022). Technology readiness levels for machine learning systems. Nature Communications, 13(1), 6039. https://doi.org/10.1038/s41467-022-33128-9

Legner, C., Eymann, T., Hess, T., Matt, C., Böhmann, T., Drews, P., Mädche, A., Urbach, N., & Ahlemann, F. (2017). Digitalization: Opportunity and Challenge for the Business and Information Systems Engineering Community. Business & Information Systems Engineering, 59(4), 301–308. https://doi.org/10.1007/s12599-017-0484-2

Leonard-Barton, D. (1988). Implementation as mutual adaptation of technology and organization. Research Policy, 17(5), 251–267. https://doi.org/10.1016/0048-7333(88)90006-6

Lorenz, R., Kraus, M., Wolf, H., Feuerriegel, S., & Netland, T. H. (2022). Selecting advanced analytics in manufacturing: A decision support model. Production Planning & Control. https://doi.org/10.1080/09537287.2022.2126951

Maghazei, O., Lewis, M. A., & Netland, T. H. (2022). Emerging technologies and the use case: A multi-year study of drone adoption. Journal of Operations Management, 68(6–7), 560–591. https://doi.org/10.1002/joom.1196

Martinez, I., Viles, E., & Olaizola, I. G. (2021). A survey study of success factors in data science projects. 2021 IEEE International Conference on Big Data (Big Data), 2313–2318. https://doi.org/10.1109/BigData52589.2021.9671588

Martínez-Plumed, F., Contreras-Ochando, L., Ferri, C., Hernández-Orallo, J., Kull, M., Lachiche, N., Ramírez-Quintana, M. J., & Flach, P. (2019). CRISP-DM Twenty Years Later: From Data Mining Processes to Data Science Trajectories. IEEE Transactions on Knowledge and Data Engineering, 33(8), 3048–3061. https://doi.org/10.1109/TKDE.2019.2962680

May, A., Sagodi, A., Dremel, C., & van Giffen, B. (2020). Realizing Digital Innovation from Artificial Intelligence. ICIS 2020 Proceedings. Forty-First International Conference on Information Systems.

McKinsey. (2018). Building an effective analytics organization. McKinsey. https://www.mckinsey.com/industries/financial-services/our-insights/building-an-effective-analytics-organization

McKinsey. (2020). The state of AI in 2020. McKinsey. http://ceros.mckinsey.com/global-ai-survey-2020-a-desktop

McKinsey. (2022). The state of AI in 2022—And a half decade in review. McKinsey. https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai-in-2022-and-a-half-decade-in-review

Miah, S. J., Gammack, J. G., & McKay, J. (2019). A metadesign theory for tailorable decision support. Journal of the Association for Information Systems, 20(5).

Microsoft. (2023). What is the Team Data Science Process? Microsoft. https://learn.microsoft.com/en-us/azure/architecture/data-science-process/overview

Mikalef, P., & Krogstie, J. (2020). Examining the interplay between big data analytics and contextual factors in driving process innovation capabilities. European Journal of Information Systems, 29(3), 260–287. https://doi.org/10.1080/0960085X.2020.1740618

Mikalef, P., Pappas, I., Krogstie, J., & Pavlou, P. (2020). Big data and business analytics: A research agenda for realizing business value. Information & Management, 57(1). https://doi.org/10.1016/j.im.2019.103237

Mikalef, P., Pappas, I. O., Krogstie, J., & Giannakos, M. (2018). Big data analytics capabilities: A systematic literature review and research agenda. Information Systems and E-Business Management, 16(3), 547–578. https://doi.org/10.1007/s10257-017-0362-y

Møller, C., Madsen, O., Berger, U., Schou, C., Lassen, A. H., & Waehrens, B. V. (2023). The Smart Production Vision. In O. Madsen, U. Berger, C. Møller, A. Heidemann Lassen, B. Vejrum Waehrens, & C. Schou (Eds.), The Future of Smart Production for SMEs: A Methodological and Practical Approach Towards Digitalization in SMEs (pp. 13–28). Springer International Publishing. https://doi.org/10.1007/978-3-031-15428-7_2

Mortenson, M. J., Doherty, N. F., & Robinson, S. (2015). Operational research from Taylorism to Terabytes: A research agenda for the analytics age. European Journal of Operational Research, 241(3), 583–595. https://doi.org/10.1016/j.ejor.2014.08.029

Mullarkey, M. T., & Hevner, A. R. (2019). An elaborated action design research process model. European Journal of Information Systems, 28(1), 6–20. https://doi.org/10.1080/0960085X.2018.1451811

OECD/Eurostat. (2005). Oslo Manual: Guidelines for Collecting and Interpreting Innovation Data (3rd Edition). OECD Publishing. https://doi.org/10.1787/9789264013100-en

Oliva, R. (2019). Intervention as a research strategy. Journal of Operations Management, 65(7), 710–724.

O’Reilly, C. A., & Tushman, M. L. (2013). Organizational Ambidexterity: Past, Present, and Future. Academy of Management Perspectives, 27(4), 324–338. https://doi.org/10.5465/amp.2013.0025

Orlikowski, W. J. (1995). Action and artifact: The structuring of technologies-in-use. Sloan School Working Paper. https://dspace.mit.edu/bitstream/handle/1721.1/2600/SWP-386734131415.pdf?sequence=1

Osmundsen, K., Iden, J., & Bygstad, B. (2019). Organizing robotic process automation: Balancing loose and tight coupling. Proceedings of the 52nd Hawaii International Conference on System Sciences. Hawaii International Conference on System Sciences.

Peffers, K., Tuunanen, T., & Niehaves, B. (2018). Design science research genres: Introduction to the special issue on exemplars and criteria for applicable design science research. European Journal of Information Systems, 27(2), 129–139.

Pentland, B. T., & Feldman, M. S. (2008). Designing routines: On the folly of designing artifacts, while hoping for patterns of action. Information and Organization, 18(4), 235–250. https://doi.org/10.1016/j.infoandorg.2008.08.001

Penttinen, E., Kasslin, H., & Asatiani, A. (2018). How to Choose between Robotic Process Automation and Back-end System Automation? ECIS 2018 Proceedings. Twenty-Sixth European Conference on Information Systems.

Purao, S., Henfridsson, O., Rossi, M., & Sein, M. (2013). Ensemble artifacts: From viewing to designing in action design research. Systems, Signs & Actions, 7(1), 73–81.

Rai, A. (2017). Editor's comments: Diversity of design science research. MIS Quarterly, 41(1), iii–xviii.

Ransbotham, S., Khodabandeh, S., Kiron, D., Candelon, F., Chu, M., & LaFountain, B. (2020). Are you making the most of your relationship with AI. Boston Consulting Group. https://www.bcg.com/publications/2020/is-your-company-embracing-full-potential-of-artificial-intelligence

Rogers, E. M. (1995). Diffusion of Innovations (Fourth Edition). The Free Press.

Rolland, K. H., Ghinea, G., & Gronli, T.-M. (2015). Ambidextrous Enterprise Architecting: Betting on the Future and Hacking Path-dependencies. ECIS 2015 Proceedings. Twenty-Third European Conference on Information Systems.

Rosemann, M. (2014). Proposals for Future BPM Research Directions. In C. Ouyang & J.-Y. Jung (Eds.), Asia Pacific Business Process Management (pp. 1–15). Springer International Publishing.

Ross, J. W., Beath, C. M., & Mocker, M. (2019). Designed for digital: How to architect your business for sustained success. MIT Press.

Ross, J. W., Weill, P., & Robertson, D. (2006). Enterprise architecture as strategy: Creating a foundation for business execution. Harvard business press.

Sætre, A. S., & Van De Ven, A. (2021). Generating Theory by Abduction. Academy of Management Review, 46(4), 684–701. https://doi.org/10.5465/amr.2019.0233

Sarker, S., Chatterjee, S., Xiao Xiao, & Elbanna, A. (2019). The Sociotechnical Axis of Cohesion for the Is Discipline: Its Historical Legacy and Its Continued Relevance. MIS Quarterly, 43(3), 695–720. https://doi.org/10.25300/MISQ/2019/13747

Schmiedel, T., & vom Brocke, J. (2015). Business Process Management: Potentials and Challenges of Driving Innovation. In J. vom Brocke & T. Schmiedel (Eds.), BPM - Driving Innovation in a Digital World (pp. 3–15). Springer International Publishing. https://doi.org/10.1007/978-3-319-14430-6_1

Schneider, J., Abraham, R., Meske, C., & Vom Brocke, J. (2023). Artificial Intelligence Governance For Businesses. Information Systems Management, 40(3), 229–249. https://doi.org/10.1080/10580530.2022.2085825

Sculley, D., Holt, G., Golovin, D., Davydov, E., Phillips, T., Ebner, D., Chaudhary, V., Young, M., Crespo, J.-F., & Dennison, D. (2015). Hidden Technical Debt in Machine Learning Systems. Advances in Neural Information Processing Systems, 28. https://proceedings.neurips.cc/paper/2015/hash/86df7dcfd896fcaf2674f757a2463eba-Abstract.html

Sein, M. K., Henfridsson, O., Purao, S., Rossi, M., & Lindgren, R. (2011). Action Design Research. MIS Quarterly, 35(1), 37–56. https://doi.org/10.2307/23043488

Sein, M. K., & Rossi, M. (2019). Elaborating ADR while drifting away from its essence: A commentary on Mullarkey and Hevner. European Journal of Information Systems, 28(1), 21–25.

Shollo, A., Hopf, K., Thiess, T., & Müller, O. (2022). Shifting ML value creation mechanisms: A process model of ML value creation. The Journal of Strategic Information Systems, 31(3), 101734. https://doi.org/10.1016/j.jsis.2022.101734

Slack, N., Brandon-Jones, A., & Johnston, R. (2013). Operations management (Seventh Edition). Pearson Education.

Spring, M., Faulconbridge, J., & Sarwar, A. (2022). How information technology automates and augments processes: Insights from Artificial-Intelligence-based systems in professional service operations. Journal of Operations Management, 68(6–7), 592–618.

Stoddard, D. B., & Jarvenpaa, S. L. (1995). Business process redesign: Tactics for managing radical change. Journal of Management Information Systems, 12(1), 81–107.

Studer, S., Bui, T. B., Drescher, C., Hanuschkin, A., Winkler, L., Peters, S., & Müller, K.-R. (2021). Towards CRISP-ML(Q): A Machine Learning Process Model with Quality Assurance Methodology. Machine Learning and Knowledge Extraction, 3(2), 392–413. https://doi.org/10.3390/make3020020

Tarafdar, M., Beath, C. M., & Ross, J. W. (2017). Enterprise cognitive computing applications: Opportunities and challenges. It Professional, 19(4), 21–27.

Tarafdar, M., Beath, C. M., & Ross, J. W. (2019). Using AI to enhance business operations. MIT Sloan Management Review, 60(4).

Tilson, D., Lyytinen, K., & Sørensen, C. (2010). Research Commentary: Digital Infrastructures: The Missing IS Research Agenda. Information Systems Research, 21(4), 748–759.

Tim, Y., Hallikainen, P., Pan, S. L., & Tamm, T. (2020). Actualizing business analytics for organizational transformation: A case study of Rovio Entertainment. European Journal of Operational Research, 281(3), 642–655. https://doi.org/10.1016/j.ejor.2018.11.074

Törmer, R. L. (2018). Internal Digital Platforms and Generative Mechanisms of Digital Innovation. ICIS Proceedings 2018. Thirty ninth International Conference on Information Systems.

Törmer, R. L., & Henningsson, S. (2018). From Drift To Central Guidance: A Path Constitution Perspective On The Platformization Of An Information Infrastructure. ECIS 2018 Proceedings. Twenty-Sixth European Conference on Information Systems.

Tortorella, G. L., Fogliatto, F. S., Cauchick-Miguel, P. A., Kurnia, S., & Jurburg, D. (2021). Integration of Industry 4.0 technologies into Total Productive Maintenance practices. International Journal of Production Economics, 240, 108224. https://doi.org/10.1016/j.ijpe.2021.108224

Tran, T.-A., Ruppert, T., Eigner, G., & Abonyi, J. (2022). Retrofitting-Based Development of Brownfield Industry 4.0 and Industry 5.0 Solutions. IEEE Access, 10, 64348–64374. https://doi.org/10.1109/ACCESS.2022.3182491

Tushman, M. L., & O'Reilly, C. A. (1996). Ambidextrous Organizations: Managing Evolutionary and Revolutionary Change. California Management Review, 38(4). https://journals.sagepub.com/doi/abs/10.2307/41165852

Van Aken, J., Chandrasekaran, A., & Halman, J. (2016). Conducting and publishing design science research: Inaugural essay of the design science department of the Journal of Operations Management. Journal of Operations Management, 47, 1–8.

Venable, J., Pries-Heje, J., & Baskerville, R. (2016). FEDS: A Framework for Evaluation in Design Science Research. European Journal of Information Systems, 25(1), 77–89. https://doi.org/10.1057/ejis.2014.36

Vestues, K., & Rolland, K. (2021). Platformizing the organization through decoupling and recoupling: A longitudinal case study of a government agency. Scandinavian Journal of Information Systems, 33(1).

Vial, G., Cameron, A.-F., Giannelia, T., & Jiang, J. (2023). Managing artificial intelligence projects: Key insights from an AI consulting firm. Information Systems Journal, 33(3), 669–691.

Vial, G., Jiang, J., Giannelia, T., & Cameron, A.-F. (2021). The Data Problem Stalling AI. MIT Sloan Management Review, 62(2), 47–53.

Weill, P., & Broadbent, M. (1998). Leveraging the new infrastructure: How market leaders capitalize on information technology. Harvard Business School Press.

Weill, P., Woerner, S. L., & Harte, M. (2020). Replatforming the Enterprise. MIT Sloan CISR Research Briefing, XX(7). https://cisr.mit.edu/publication/2020_0701_Replatforming_WeillWoernerHarte

Wirth, R., & Hipp, J. (2000). CRISP-DM: Towards a Standard Process Model for Data Mining. Proceedings of the 4th International Conference on the Practical Applications of Knowledge Discovery and Data Mining, 29–40.

Womack, J. P., Jones, D. T., & Roos, D. (1990). The Machine that Changed the World. Rawson Association.

Zhang, D., Pan, S. L., Yu, J., & Liu, W. (2022). Orchestrating big data analytics capability for sustainability: A study of air pollution management in China. Information & Management, 59(5), 103231. https://doi.org/10.1016/j.im.2019.103231

APPENDED PAPERS

Paper 1: Bojer, Casper Solheim, and Møller, Charles. (2022). *Towards a Scheme for Contribution in Action Design Research*. In International Conference on Design Science Research in Information Systems and Technology (pp. 376-387). Cham: Springer International Publishing.

Paper 2: Bojer, Casper Solheim, and Møller, Charles. (2023a). *Conceptualizing Process Innovation with Analytics: A Pragmatic Framework and Research Agenda* [Manuscript submitted for publication].

Paper 3: Bojer, Casper Solheim, Bygstad, Bendik, and Øvrelid, Egil. (2023). *Speeding up Explorative BPM with Lightweight IT: The Case of Machine Learning* [Manuscript submitted for publication].

Paper 4: Bojer, Casper Solheim, and Møller, Charles. (2023b). *Developing Analytics Demonstrators for Process Innovation: An Infrastructural Perspective* [Manuscript submitted for publication].

Paper 1: Towards a Scheme for Contribution in Action Design Research

Bojer, Casper Solheim, and Møller, Charles. (2022). *Towards a Scheme for Contribution in Action Design Research*. In International Conference on Design Science Research in Information Systems and Technology (pp. 376-387). Cham: Springer International Publishing.

Towards a Scheme for Contribution in Action Design Research

Casper Solheim Bojer[1][0000-0001-8344-5416] and Charles Møller[1][0000-0003-0251-3419]

[1] Department of Materials and Production, Aalborg University, 9220 Aalborg, Denmark
{csb,charles}@mp.aau.dk

Abstract. Researchers are increasingly asked to engage with industry in research projects and contribute to both practice and academia. Action Design Research (ADR) is gaining traction in IS due to its potential to achieve this dual goal. While the practical utility of ADR projects is obvious, the role of design science research (DSR) in knowledge abstraction and accumulation is still unclear and the subject of much discussion. Some scholars suggest DSR should build theory, some that it should test theory, while others suggest that its contributions lie elsewhere. While the elaborated ADR model of Mullarkey & Hevner (2019) clarified the potential for artefactual contributions at different abstraction levels throughout the research process, other types of contribution were left for further research. Drawing on reflections from an ongoing research project using ADR, as well as research on theorizing and DSR contributions, we present a tentative conceptual scheme that considers both empirical, artefactual, theory building, and theory testing opportunities in ADR. We discuss the benefits of the scheme in identifying contribution opportunities and reflect on its utility in research design for industrially engaged DSR.

Keywords: Design Science Research, Action Design Research, Theorizing, Theoretical Contribution, Contribution.

1 Introduction

Design Science Research has become an important research approach in IS due to its future-orientation and its potential to make knowledge contributions that are both rigorous and relevant (Hevner et al., 2004). It thus meets a need in a time where researchers are increasingly asked to engage in research projects with industrial partners that deliver both practical and academic contributions. As a result of this situation, several methods for conducting industrially engaged DSR research have been introduced in recent years. Action Design Research is one such method that focuses on real-world problem-solving at a client by the introduction of an ensemble artefact and subsequent abstraction of the knowledge obtained in the process (Sein et al., 2011). Intervention-based Research (IBR) (Oliva, 2019), an Action Research-inspired method originating in Operations Management, also features real-world problem solving at a client, but focuses on the use of traditional theory building and testing for academic contributions. Industrially engaged DSR projects frequently span multiple years and it can therefore be necessary for institutional reasons to publish multiple contributions during the project, such as for doctoral students or early-career researchers. However, existing methodological work provides little guidance on how to transform an industrially engaged DSR project into multiple sequential knowledge contributions as the project unfolds. First steps towards providing such guidance are provided in the elaborated ADR model of Mullarkey & Hevner (2019) which lists potential artefactual contributions in each of their four ADR cycles (Diagnosis, Design, Implementation, and Evolution), however, they leave how ADR can make other contributions for further research. Recent research on contributions in DSR suggests that it has the potential to deliver many contributions in addition to artefacts and design principles, such as design theories (Gregor & Jones, 2007; Iivari, 2020), practical theories (Goldkuhl & Sjöström, 2021), substantive technological theory (Iivari, 2020), and empirical contributions (Goldkuhl & Sjöström, 2021).

In this paper we attempt to advance the work initiated by Mullarkey & Hevner (2019) by synthesizing existing research on contributions in DSR with an emphasis on theorizing and relating it to the elaborated ADR process model. Drawing on reflections from applying ADR and attempting to plan a series of research contributions, we expand on the potential for contribution in the four ADR cycles. As our main contributions we: 1) provide a review of perspectives on contributions in DSR; and 2) develop a conceptual scheme for contributions in the four ADR cycles, which includes empirical, artefactual, theory building, and theory testing opportunities. We thus add to the discussion of knowledge contributions and accumulation in DSR. The overview provided by our conceptual scheme supports future ADR researchers in research design by providing guidance in terms of how to identify and publish valuable knowledge contributions, thereby making it easier to achieve the dual aims of contributions to practice and research.

In the next section, we cover extant literature on industrially engaged action-oriented DSR, theoretical contributions, and DSR contributions. We then present reflections from ongoing ADR research, before presenting our conceptual scheme for ADR research contributions and applying it to our research. Finally, we discuss implications for ADR research, compare our scheme to related work, and conclude on our contribution.

2 Background

2.1 Action-oriented Design Science Research

As applied research fields are increasingly being asked to conduct industrially engaged research, they have developed action-oriented DSR methods that enable making both rigorous academic and practically relevant contributions. In IS, ADR is such a method aimed at inductively developing generalizable design knowledge by solving a specific problem through building and evaluating ensemble artifacts in an organizational setting (Sein et al., 2011). The main academic knowledge contributions in ADR are design principles that describe how to produce a (general) solution that addresses a class of problems. While theory-inspired design principles are formulated and refined throughout the process, the publishing of the design principles is presented as taking place at the end of the project in the Formalization stage, potentially with an additional contribution in the form of a theoretical refinement to the theories used (Sein et al., 2011, p. 44). Mullarkey & Hevner (2019) proposed the elaborated ADR process model, which consists of four iterative cycles each with a different purpose: 1) *Diagnosis*, 2) (conceptual) *Design*, 3) *Implementation*, and 4) *Evolution*. Each of these cycles consists of steps inspired by the original ADR model and produces different artefacts that has the potential to be formalized and published as an academic knowledge contribution. The potential for and importance of publishing the interim products of ADR is also acknowledged by Sein & Rossi (2019) in their response to the elaborated ADR model. However, Mullarkey & Hevner (2019) leave it for further research to integrate design theory development and do not address other forms of contributions.

In Operations Management (OM), IBR is gaining traction as an action-oriented DSR method. In IBR, theoretical frameworks are used to build interventions and make predictions of their results. Anomalies are considered as potentials for modifications to theory, while the organizational dynamics observed after intervening are framed as data that can be used for process theorizing (Oliva, 2019). Top journals in OM have come to place less emphasis on the artefactual contribution, focusing instead on theory testing and theory building (Chandrasekan et al., 2020). This view on theory is thus much closer to that of Canonical Action Research (CAR) (Davison et al., 2012), with the addition of the potential for in/abductively generating process theory.

2.2 Theoretical Contribution

Theory is concerned with improving our ability to understand phenomena and is one of the main communication devices used to transfer knowledge in scientific discourse. Traditionally, theory has been conceptualized as being limited to conceptual abstractions consisting of constructs, relationships, and boundary conditions, with the aim of explanation and prediction of phenomena (Bacharach, 1989). More recent discourse has broadened the scope to include theories with different purposes and formats. Gregor (2006) expanded the scope to include theories for analysis, theories for either predicting or explaining, and theories for design and action. It has likewise been recognized that theory can take different forms depending on the underlying meta-theoretical approach selected, which will in turn focus the inquiry on particular aspects of the phenomena (Burton-Jones et al., 2015).

While theories are generally highly regarded as a prime research outcome in IS, the focus on theory has recently come under critique. Avison and Malaurent (2014) suggests that we are facing a theory-fetish in IS, which prevents our field from making progress, while Alter (2017) states that the focus on theory limits the publication of a variety of other useful conceptual artefacts. In this paper we adopt an inclusive view of theory that contains all five theory types by Gregor (2006).

Individual publications rarely produce a complete theory. Most theoretical contributions either advance an existing theory slightly, or take the form of interim products of theorizing which can have an important role to play in advancing the academic discourse (Weick, 1995). Examples of interim theorizing products include conceptual frameworks, models, and diagrams. Colquitt and Zapata-Phelan (2007) argues for distinguishing between building theory and testing theory and present a taxonomy for categorizing contributions based on the degree of theory building and testing present.

2.3 DSR Contributions

While the distinguishing feature of DSR lies in artefactual contributions (Hevner, 2004), multiple authors have argued that empirical, and theoretical contributions are also possible (Ågerfalk, 2021; Goldkuhl & Sjöström, 2021). After artefactual contributions, theoretical contributions in the form of Type V (design) theories (Gregor & Jones, 2007) have arguably received the most attention in the DSR community. While full design theories are not a necessary outcome of DSR, design theorizing and knowledge abstraction, makes up an important part of a DSR contribution (Baskerville et al., 2018). Gregor & Hevner (2013) argue for distinguishing DSR contributions based on the level of abstraction and maturity, which ranges from instantiations over nascent design theory to well-developed design theories. A popular (nascent) design theoretical contribution is design principles, which are prescriptive means-end statements. Synthesizing various formulations of prescriptive statements, Gregor et al. (2020) arrives at seven building blocks of design principles: implementers, aim, user, context, mechanisms, enactors, and rationale. Attempting to add clarity to the debate on design theories, Iivari (2020) propose to distinguish between three types of design theory: theory used to derive meta-requirements (Design Theory 1); theory used to explain why meta-requirements are satisfied by the meta-design (Design Theory 2); and theory used to explain the effects of the IT artefact (Design Theory 3).

In addition to design theory as contribution from DSR, several authors propose other types of theoretical contribution. Iivari (2020) propose that DSR can contribute by testing, refining, or proposing substantial technological theories (STT) from Bunge (1966). STT's are essentially applied versions of kernel theories that are close enough to the problem context to guide design and ground design theories. In his view, DSR can thus in addition to the artefact contribute with either 1) one or more types of design theory, or 2) STT. Goldkuhl & Sjöström (2021) argue that in addition to generating design theory, DSR has the potential to contribute with both building and testing of practical theory. Practical theory is theory that offers practical utility in the design inquiry process and can include traditional theories, e.g., for description and explanation, as well as other tools that are useful in problem diagnosis, planning & design, and evaluation (Goldkuhl & Sjöström, 2021). Additionally, they suggest the potential for empirical contributions by reporting on the rich data collected and knowledge obtained as part of the design inquiry. The varied nature of DSR contributions is also acknowledged by Drechsler & Hevner (2018) that suggest the potential for theoretical contributions to both descriptive (type I-IV) and prescriptive knowledge (type V) of varying maturity, in addition to concrete instantiations. Table 1 summarizes the various viewpoints related to contributions from DSR.

Table 1. Viewpoints on Potential Contributions from DSR

Viewpoint	References	Description
Contribution as Artefact	Hevner et al., 2004	DSR can contribute with artefacts in the form of constructs, models, methods, and instantiations
Contribution as Conceptual Abstractions	Mullarkey & Hevner, 2019	DSR can contribute with a variety of conceptual abstractions aimed at diagnosis, design, implementation, and evolution.
Contribution as Empirical	Goldkuhl & Sjöström, 2021; Ågerfalk, 2021	DSR can contribute with rich empirical descriptions based on close engagement with the problem & solution.
Contribution as Theory Testing	Oliva, 2019; Chandrasekan et al., 2020	DSR can contribute with practical testing of type I-IV theories.
Contribution as Inductive Process Theory-Building	Oliva, 2019; Chandrasekan et al., 2020	DSR can contribute with inductive building of process theories explaining the observed organizational transition from pre- to post-intervention.
Contribution as Practical Theory	Goldkuhl & Sjöström (2021)	DSR can contribute with testing, refinement, and building of practical theories for diagnosis, design, and evaluation.
Contribution as Substantive Technological Theory	Iivari (2020)	DSR can contribute with development of substantive technological theories inspired by the artefacts.
Contribution as Design Theory	Gregor & Hevner (2013), Iivari (2020)	DSR can contribute with design theories that 1) theoretically ground meta-requirements, 2) explain why meta-design satisfies the requirements, 3) explain the effects of the artefact, or 4) all the above.

3 Empirical Grounding: Reflections on ongoing ADR Research

We reflect on ADR contributions by means of an ongoing three-year research project following the elaborated ADR model (Mullarkey & Hevner, 2019), where the project is currently halfway. The goal of the project is to develop an approach that is both fast and scalable for development and real-life evaluation of machine learning (ML) based IS aimed at internal process innovation. The problem setting is a large Danish manufacturing company that is in the process of building big data and analytical capabilities, but currently finds the process of development and evaluation of ML-based systems too slow to enable rapid exploration. The research project thus sits at the intersection of IS and Operations Management. In addition to the approach and associated design knowledge, it was expected that the research project would deliver more traditional contributions to existing relevant academic knowledge bases. Three knowledge bases were identified in an initial research design phase and used to theoretically ground and frame the project: business process management, enterprise architecture, and dynamic capabilities. However, it was at this stage unclear what the nature of these contributions would be.

The main artefactual outcome of the research project is the approach, which includes a high-level design process, as well as design principles, suggested architectures for the different layers of the IS, and one or more instantiations of the approach presenting proof-by-construction. Following the conceptual artefacts presented by Mullarkey & Hevner (2019), this would amount to one or more systems (the ML-based IS), and one process (the approach), and several design and diagnosis artefacts. In addition, the combination of architectures, constructs, and design principles is a form of nascent design theory (Gregor & Hevner, 2013). Table 2 showcases the expected outputs from the research project (in bold) mapped to the range of potential artefacts listed by Mullarkey & Hevner (2019).

Table 2. Potential contributions by ADR cycle according to Mullarkey & Hevner (2019) with the expected artefactual contributions of our research project in bold.

Stage	Diagnosis	Design	Implementation	Evolution
Contribution	**Conceptualization of Problem and/or Solution**, **Requirements Definition,** Technical Specification, **Assessment of Existing Tools**,	**Design Features** **Design Principles** **Models** **Architectures** **Implementation Methods** **Constructs** **Nascent Design Theory**	**Systems** Algorithms Programs Databases **Processes** **Nascent Design Theory**	Modification to any of the previous artefacts Nascent Design Theory

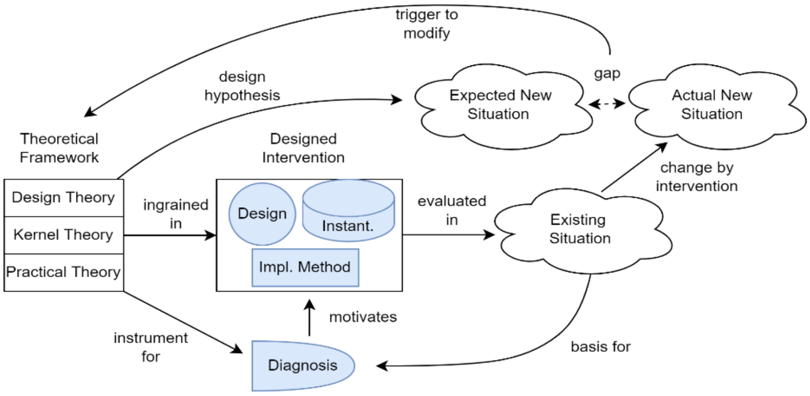
	Critical Success Factors **Nascent Design Theory**			

Engaging with the literature in one of the iterations of the *Diagnosis* phase led to the research area of big data analytics (BDA) capabilities. While some progress was made on conceptualization of BDA capabilities, empirical research was scarce, and theory related to their evolution and value creation mechanisms was nascent. This led us to the realization that our close collaboration with our industrial partner and the embeddedness of the main author presented us with the opportunity to contribute to this debate through rich empirical descriptions or early theoretical contributions. Where and how these contribution opportunities fit with the ADR process was not evident, and thus following Sein et al. (2011), our early research designs envisaged them as taking place at the end of the project, after design, evaluation, and further data collection was finished.

4 A Conceptual Scheme for ADR Contributions

Our research design evolved as the *Diagnosis* and *Design* phases unfolded based on engagement with the problem context and DSR literature. Recent contributions on the role of theory in DSR made it evident that empirical and theoretical contributions do not have to be add-ons at the end of the research project. Instead, they can take place during the project, after one or more iterations of one of the four cycles. The iterative process of rethinking our research design led us to reflect on the learnings we achieved in the process. Figure 1 shows our updated conceptualization of the role of theory in ADR, where each iteration of an ADR cycle provides the opportunity to engage in both theory testing, as emphasized in IBR and CAR, and theory building, for a variety of theory types. In the *Problem Formulation/Planning* stage, the problem and solution are grounded in existing knowledge. Practical theories can aid understanding and assessment of the situation, and design theories can be deductively developed using kernel or substantive technological theories. These theories present the theoretical framework used in the design of artefacts, as per the *Theory-ingrained artefact* principle, and the design will thus contain theory-driven hypotheses. These hypotheses are tested through *Evaluation* where the designed artefacts are introduced into the existing situation, bringing about change, and a new situation. Depending on whether the new situation matches our expected situation, the hypotheses are corroborated or falsified. During both stages, practical theories can be used to inform or shape artefact creation and evaluation. In terms of inductive theorizing, rich data is collected on the existing situation in the *Problem Formulation/Planning stage*, on the development of the artefact in *Artifact Creation*, and on the performance of the artefact in *Evaluation*, using methods such as participative observation, interviews, process performance measurements, etc. As a result, once reaching the *Reflection* stage, the researchers(s) have the results of their hypotheses tests as well as a rich empirical database that can serve as the foundation for a theoretical contribution.

Fig. 1. Role of theoretical framework and artefacts (in blue) in ADR. Inspired by (Oliva, 2019).



While there is potential for both empirical contributions, and inductive and deductive theoretical contributions in all the four cycles, the nature of the contributions will differ for each cycle due to differences in content, aim, methods, and the existing theories employed. Table 3 shows our conceptual scheme, which provides an overview of the roles of theory in each of the four stages and the potentials for contribution.

In the *Diagnosis* phase, rich empirical data on the problem situation and context is collected using, e.g., interviews, participative observation, document analysis, etc. This rich empirical data, if the context or problem is novel and interesting, can with good narrative be turned into an empirical contribution, which might inspire future type IV and type V theorizing. To structure the data collection and obtain understanding of the often-complex situation, existing practical theories and type I-IV theories can be used. By combining one or more theories, a theoretical or conceptual framework can be constructed, which can serve to produce artefacts in the form of conceptualizations of the problem and the solution space. Confronting the theoretical framework with the organizational situation allows for testing the practical validity of these theories in terms of their ability to understand and explain the problem and solution space, e.g., as assessed by practitioners. Adjustments to practical theories based on *Reflection* or development of novel ones can serve as potential contributions if formalized. Even if the practical theory proves useful without modification, reporting the test result can still make a theoretical contribution if the context of use extends the current boundaries of the theory. Kernel theory can also be used to generate theory-based *Requirements*, through the conversion to STT, thus constituting theorizing for a Design Theory 1 (Iivari, 2020).

In the *Design* phase, rich empirical data on the conceptual design process, such as actions, events, and the evolving design is collected. While this data can serve as an important foundation for theorizing about the design process, it is perhaps less useful as an empirical contribution on its own, unless some aspect of the design process followed was particularly novel or surprising. In this phase, kernel theory can be used in theorizing for a Design Theory 2 (Iivari, 2020), which explains why the design satisfies the requirements, and a Design Theory 3, which explains the effects that the introduction of the designed artefact into the problem context will produce. In both cases, the kernel theory will likely need to be translated to a STT to be concrete enough for design theory derivation. The design in this case can consists of all the six artefacts listed by Mullarkey & Hevner (2019). Practical theories might be used as a source of inspiration or for the generation of constructs, as well as serving as guidance in assessing the value of different design options (Goldkuhl & Sjöström, 2021).

Table 3. Scheme for potential contributions in each of the four ADR cycles. Potential contributions are shown by either 'X' or elaborating text.

Potential Contribution	Diagnosis	Design	Implementation	Evolution
Theory Building				
Design Theory	Hypothesis, Propositions		In-/Abductive Building	
STT	Hypothesis, Propositions		In-/Abductive Building	
Type I-IV	Theorizing Products		Theorizing Products, In-/Abductive Building	
Practical Theory	Theorizing Products, Theory Modifications		Theorizing Products, Theory Modifications, In-/Abductive Building	
Theory Testing				
Design Theory			X	X
STT			X	X
Type I-IV	X		X	X
Practical Theory	X	X	X	X
Non-theory				
Rich Empirical Descriptions	X	X	X	X
Artefacts (see Mullarkey & Hevner, 2019)	X	X	X	X

In the *Implementation* phase, an instantiation of the ensemble artifact is tested out in the organizational context thus providing the first in-situ evaluation of the instantiation, but also the problem framing, theoretical framework, and the conceptual design. Due to the emergent nature of the artefacts and the often-complex nature of the problem situation, it is likely that modifications are needed to one or more of the above elements. If detailed data is collected on the outcome of the implementation, this can serve as the foundation for Type I-IV theorizing. As an example, focusing data collection on the dynamics of the environment after introduction of the artefact can enable inductive process theorizing, as emphasized in IBR.

In the *Evolution* phase, rich empirical data can be collected on the evolution of the artefact and its environment as the ensemble artefact emerges from continual interaction and redesign (Sein et al., 2011). This data can be used for inductively theorizing about the evolution of this class of solutions and its effects on the environment. Each evaluation in the *Evolution* phase is thus a repeated test of any unmodified ingrained theories and new tests of any changes to the theoretical framework and provides the opportunity for revision to any of the previous artefacts developed or theories used.

4.1 Application for Research Design

To demonstrate the utility of our conceptual scheme, we present the results of applying it in our project for research design. The theoretical framework we arrived at for our case through multiple iterations in *Diagnosis* and *Design* can be seen in Table 4.

Table 4. Overview of the theoretical framework in our project, categorized by role.

Role of Theory	Theoretical Framework Components
Practical Theories	Business Process Management, Work Systems Theory, Enterprise Architecture
Kernel Theories	Dynamic capabilities, Digital Infrastructure, Process Innovation, Technology Innovation, Big Data Analytics capabilities (BDAC)
Design Knowledge	Architecture & Development Processes (Software Engineering & Machine Learning), Explorative Process Prototyping.

Engaging with the conceptual scheme in our current iteration of research design, we identified five potential publications, with one of them optional (#4) pending results of testing the practical theories in the first three cycles. Our identified contributions range from practical theorizing, through conceptual design theorizing, to testing of practical and design theory, and finally a case study featuring a rich empirical description and nascent inductive theorizing for BDAC, see Table 5.

Table 5. Application of the scheme for research & publication design in our case.

Cycle	Publication Number & Nature of Contribution
Diagnosis	#1: Type I Practical Theorizing + Design Theory 1 Theorizing
Design	#2: Models, Architectures + Design Theory 2 & 3 Theorizing
Implementa-tion	#3: Instantiated Approach + Tested Design Principles *#4: BPM/EA Testing & Modification – if justified*
Evolution	#5: Case Study + Nascent Type IV Theorizing for BDAC

5 Discussion

Our conceptual scheme provides a synthesis of different opportunities for contribution in ADR and relate them to the elaborated ADR model and IBR. We see the conceptual scheme as a useful tool for research design, where it can be used as a basis for exploring potential publication strategies. This is particularly relevant for early-stage researchers, who often need to publish several contributions during a multi-year research project. From our conceptualization of contribution in Figure 1 and as exemplified in our application of the scheme, the feasibility of making certain research contributions in ADR depend on 1) the results obtained by interaction with the context, and 2) the theoretical framework and methods employed. The research design and publication strategy will thus have to be revisited as the research process unfolds, but when this should happen is not addressed in the elaborated ADR model. We found that doctoral practicalities required us to make an initial design before starting ADR and revisiting it periodically.

Compared to previous work on conceptualization of contributions in DSR, we focus on the temporal aspect of the potential for contribution. Compared to the conceptualization of Dreschler & Hevner (2018), we expand on the potential for contribution to descriptive knowledge by distinguishing between practical theory, kernel theory, and substantive technological theories and emphasize the potential for empirical contributions. Compared to Maedche et al. (2021) our conceptualization suggests that it is possible for the research to be classified in different quadrants at different points of the research project, e.g., making observation-based descriptive statements in the *Diagnosis* stage, and in later stages contributing with creation-based prescriptive statements.

In line with Iivari (2020), we found that it was difficult to distinguish between artefactual and theoretical contributions, particularly when considering the abstraction principle of Mullarkey & Hevner (2019). This was the case for both practical theories and design theories. As examples, take the problem conceptualization artefact of Mullarkey & Hevner (2019) and a practical diagnostic theory as introduced in Goldkuhl and Sjöström (2021), or design principles vs. design theory. We thus believe that the DSR community stands to gain from further rigor in the discussion of contributions.

6 Conclusion

We present a conceptual scheme for potential research contributions in ADR based on a synthesis of extant literature on theorizing and contributions in DSR. We show that ADR projects have the potential to make empirical, theoretical, and artefactual contributions in each of the cycles of *Diagnosis*, *Design*, *Implementation*, and *Evolution*. We thus highlight the potential for mixed configurations of contributions throughout a DSR project. Our conceptual scheme supports industrially engaged DSR researchers in research design and publication planning, by providing an overview of the space of potential contributions. This should prove especially useful for early-stage researchers, who must deliver multiple publications during their industrially engaged research projects. The conceptual scheme we propose is only a first step towards a thorough understanding of the theorizing potential in ADR. Further research should identify exemplars of the contribution opportunities, although a challenge here is that not all ADR-based contributions are likely to be advertised as such. In addition, how to best integrate research design activities with the elaborated ADR model remains an open question.

References

1. Mullarkey, M. T., & Hevner, A. R. (2019). An elaborated action design research process model. European Journal of Information Systems, 28(1), 6-20.
2. Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design science in information systems research. MIS quarterly, 75-105.
3. Sein, M. K., Henfridsson, O., Purao, S., Rossi, M., & Lindgren, R. (2011). Action design research. MIS quarterly, 37-56.
4. Oliva, R. (2019). Intervention as a research strategy. Journal of Operations Management, 65(7), 710-724.

5. Gregor, S., & Jones, D. (2007). The anatomy of a design theory. Association for Information Systems.
6. Iivari, J. (2020). A critical look at theories in design science research. Journal of the Association for Information Systems, 21(3), 10.
7. Goldkuhl, G., & Sjöström, J. (2021). Design Science Theorizing: The Contribution of Practical Theory. In Advancing Information Systems Theories (pp. 239-273). Palgrave Macmillan, Cham.
8. Sein, M. K., & Rossi, M. (2019). Elaborating ADR while drifting away from its essence: A commentary on Mullarkey and Hevner. European Journal of Information Systems, 28(1), 21-25.
9. Chandrasekaran, A., de Treville, S., & Browning, T. (2020). Intervention-based research (IBR)—What, where, and how to use it in operations management. Journal of Operations Management, 66(4), 370-378.
10. Davison, R. M., Martinsons, M. G., & Ou, C. X. (2012). The roles of theory in canonical action research. MIS Quarterly, 763-786.
11. Bacharach, S. B. (1989). Organizational theories: Some criteria for evaluation. Academy of management review, 14(4), 496-515.
12. Gregor, S. (2006). The nature of theory in information systems. MIS quarterly, 611-642.
13. Burton-Jones, A., McLean, E. R., & Monod, E. (2015). Theoretical perspectives in IS research: from variance and process to conceptual latitude and conceptual fit. European journal of information systems, 24(6), 664-679.
14. Avison, D., & Malaurent, J. (2014). Is theory king?: questioning the theory fetish in information systems. Journal of Information Technology, 29(4), 327-336.
15. Alter, S. (2017). Nothing is more practical than a good conceptual artifact… which may be a theory, framework, model, metaphor, paradigm or perhaps some other abstraction. Information Systems Journal, 27(5), 671-693.
16. Weick, K. E. (1995). What theory is not, theorizing is. Administrative science quarterly, 40(3), 385-390.
17. Colquitt, J. A., & Zapata-Phelan, C. P. (2007). Trends in theory building and theory testing: A five-decade study of the Academy of Management Journal. Academy of management journal, 50(6), 1281-1303.
18. Ågerfalk, P. J., & Karlsson, F. (2021). Theoretical, Empirical, and Artefactual Contributions in Information Systems Research: Implications Implied. In Advancing Information Systems Theories (pp. 53-73). Palgrave Macmillan, Cham.
19. Baskerville, R., Baiyere, A., Gregor, S., Hevner, A., & Rossi, M. (2018). Design science research contributions: Finding a balance between artifact and theory. Journal of the Association for Information Systems, 19(5), 358-376. https://doi.org/10.17705/1jais.00495
20. Gregor, S., & Hevner, A. R. (2013). Positioning and presenting design science research for maximum impact. MIS quarterly, 337-355.
21. Gregor, S., Chandra Kruse, L., & Seidel, S. (2020). Research perspectives: the anatomy of a design principle. Journal of the Association for Information Systems, 21(6), 2.
22. Bunge, M. (1966). Technology as applied science. In Contributions to a Philosophy of Technology (pp. 19-39). Springer, Dordrecht.
23. Drechsler, A., & Hevner, A. R. (2018). Utilizing, producing, and contributing design knowledge in DSR projects. In International Conference on Design Science Research in Information Systems and Technology (pp. 82-97). Springer, Cham.
24. Maedche, A., Gregor, S., & Parsons, J. (2021). Mapping design contributions in information systems research: the design research activity framework. Communications of the Association for Information Systems, 49(1), 12.

Paper 2: Conceptualizing Process Innovation with Analytics: A Pragmatic Framework and Research Agenda

Bojer, Casper Solheim, and Møller, Charles. (2023a). *Conceptualizing Process Innovation with Analytics: A Pragmatic Framework and Research Agenda* [Manuscript submitted for publication].

The paper has been submitted to *Business & Information Systems Engineering* and is currently under review.

An earlier version of the paper was presented at the Pre-ICIS SIGDSA Symposium in Copenhagen, December 2022.

Conceptualizing Process Innovation with Analytics – A Pragmatic Framework and Research Agenda

Casper Solheim Bojer (corresponding author)
Department of Materials & Production
Aalborg University
Fibigerstræde 16
9220 Aalborg Øst
Denmark
E-mail: casperbojer@gmail.com

Charles Møller
Department of Mechanical and Production Engineering
Aarhus University
Katrinebjergvej 89
8200 Aarhus N
Denmark
E-Mail: charles@mpe.au.dk

Abstract

Organizations are increasingly investing in *Process Innovation with Analytics*, i.e., the usage of analytics to innovate and improve operational processes. Process innovation with analytics is a highly challenging and complex endeavor encompassing 1) redesign of processes, 2) development of digital infrastructure, and 3) analytics development. As a result, organizations need guidance on how to approach this complex challenge. While research on IT-enabled process innovation and analytics offer valuable insights, process innovation with analytics necessitates contextualization of these knowledge bases due to its distinct characteristics. This research note aims to inspire further research into process innovation with analytics by 1) reconceptualizing analytics in the context of process innovation, and 2) proposing a research agenda, consisting of three research directions and five research challenges. The reconceptualization and research agenda are based on the authors' experience from an Action Design Research study at a large global manufacturer and retailer focused on process innovation with analytics. Bridging analytical, process innovation, and infrastructure perspectives, the research note offers a foundation for future scholarly endeavors and calls for further research into 1) digital infrastructures for process innovation with analytics, 2) the relationship between process change and analytics development, and 3) governance of process innovation with analytics.

Keywords

Analytics; Process Innovation; Digital Infrastructure; Machine Learning; Research Agenda; Framework

1 Introduction

Large organizations are investing heavily in analytics. Whereas analytics has historically been used primarily to derive insights in support of tactical and strategic decision-making, advantages in AI and ML have made it possible to use analytics to improve operational processes (Tarafdar et al., 2019; Benbya et al., 2021; Davenport, 2018). We refer to this use of analytics to innovate operational processes as *Process Innovation with Analytics.* In contrast to tactical and strategic uses of analytics, *Process Innovation with Analytics* requires changing processes and embedding analytics systems as part of the redesigned process. Example applications includes predictive maintenance and predictive quality in manufacturing, or automated fraud detection and loan application handling in the financial sector.

The IS community has a rich tradition for research into IT-enabled process innovation. The seminal works by Davenport & Short (1990), Davenport (1993) and Hammer & Champy (1993) introduced methodologies for bringing IT to the forefront of process innovation and further research has demonstrated the important enabling role of IT infrastructure in radically innovating processes (Broadbent et al., 1999; Kim et al., 2011; Bygstad & Øvrelid, 2020). Similarly, a rich body of literature has emerged on analytics from an organizational perspective (e.g., Dremel et al., 2017; Dremel et al., 2020; Tim et al., 2020; Zhang et al., 2022) and technical research has developed prescriptive knowledge in the form of methodologies (e.g., Wirth & Hipp, 2000; Martínez-Plumed et al., 2022; Nalchigar & Yu, 2020) and architectures (Phillips-Wren et al., 2021; Gröger, 2021) for analytics development.

"Process innovation with analytics" is, however, different from both traditional "IT-enabled process innovation", and "analytics aimed at insights" and there is thus a need for contextualization of both knowledge bases. As an instance of "IT-enabled process innovation", "process innovation with analytics" inherits its larger scope and thus differs from "analytics for insights" in requiring both process redesign and development of digital infrastructure in addition to the development of an analytical artifact or system. Characteristics of analytics as a technology and its application in processes, however, introduce some important differences between "process innovation with analytics" and traditional "IT-enabled process innovation". First, the scope of process change is smaller, often consisting of changes to subprocesses or tasks (Sedera et al., 2016), as compared to cross-functional processes. Second, analytics is a general-purpose (May et al., 2020) and weakly structured (Eley & Lyytinen, 2022) technology requiring organizations to engage in exploration and prototyping with context-specific data to assess its affordances (Dolata et al., 2022). As a result of these differences, starting with process redesign as opposed to IT development as suggested in most redesign approaches (e.g., Davenport, 1993; Gross et al., 2019) becomes less applicable, as does taking outset in any process embedded in the IT system, as was the case with highly structured enterprise systems (Volkoff & Strong, 2013).

The phenomenon of process innovation with analytics thus calls for further research on analytics that adopts the socio-technical focus present in the research on IT-enabled process innovation and similarly foregrounds the changes required to both processes and digital infrastructure. There is both a need for explanatory research to understand how companies have successfully innovated processes using analytics and prescriptive research to update and contextualize existing methods within analytics development and IT-enabled process innovation.

In this research note, our objective is to inspire further research into process innovation with analytics along these lines by 1) reconceptualizing analytics in the context of process innovation and 2) developing a research agenda consisting of several research challenges along with promising research directions. The reconceptualization and the research agenda presented are based on the generalization and abstraction of our experiences from a three-year Action Design Research (ADR) project conducted with a large global manufacturer and retailer based in Denmark. In the ADR project, the goal was to develop prescriptive knowledge for analytics development aimed at process innovation through participation in the development of analytics infrastructure and several analytics projects (Authors, under review). It was during attempts to apply existing knowledge in this project that we became aware of the limitations in existing conceptualizations of analytics and were inspired to develop a reconceptualization that applied more generally, while still aligning with our empirical experiences.

The rest of the paper is structured as follows. First, we introduce our conceptualization of process innovation with analytics by drawing on and synthesizing three perspectives: 1) the analytics perspective, 2) the process innovation perspective, and 3) the digital infrastructure perspective. We then present three research directions connected with this conceptualization along with five research challenges. We conclude with a summary of our work and a call to the BISE community to join in addressing these challenges.

2 Conceptualizing Process Innovation with Analytics

Process innovation with analytics concerns the innovation of operational work processes where embedded analytical systems play a key part in enabling the redesigned work process. Exemplary applications of process innovation with analytics includes computer-vision or sensor-based automated quality inspections in manufacturing, predictive maintenance of industrial equipment, AI-assisted screening in radiology, and automated fraud detection in credit card transactions. In all these examples, analytical systems play a key enabling role in the redesign of processes through the use of data to either augment or automate tasks. Process innovation with analytics has some similarities to what Davenport et al. (2010) labeled *embedded analytics.* A key difference is that embedding analytics into work processes does not necessitate any change in the process. As an example, analytics can be used to augment a single task while leaving the surrounding tasks unchanged. Confusingly, *embedded analytics* as a term has also been used in the computer science field to refer to analytics on embedded devices.

In our conceptualization, process innovation with analytics consists of three major areas of development. First, it involves the development of an analytics system that delivers the analytical capabilities or affordances that enables a process redesign. Second, it requires process redesign and development of the new process. Third, it requires development of the digital infrastructure of the organization, including the integration of the analytics system with this digital infrastructure. The outcome of process innovation with analytics is a newly redesigned process where an embedded analytical system plays a key role. This new process is in turn made possible by changes to the digital infrastructure. Our conceptualization is visualized in Figure 1.

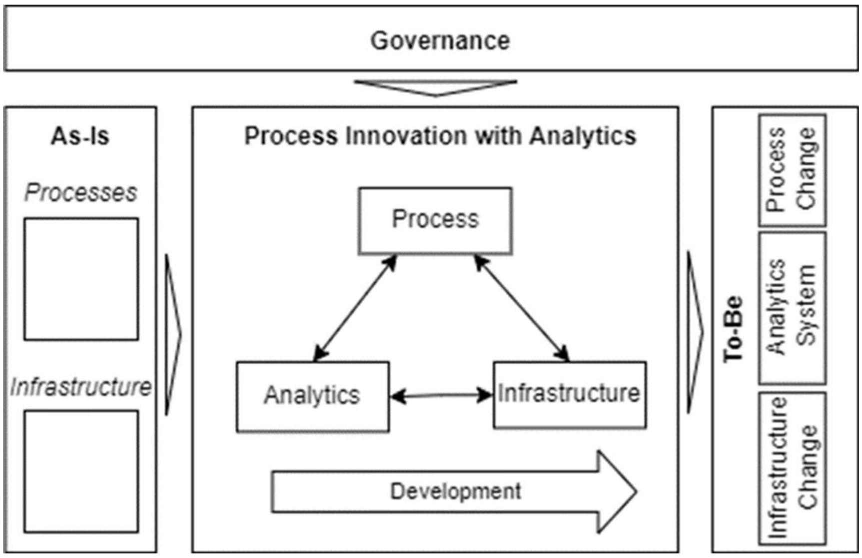


Figure 1: Conceptualization of process innovation with analytics.

The need to consider and develop both the process and the digital infrastructure means that the scope and complexity of process innovation with analytics is much greater than when analytics is used for insights. When analytics is used for insights, it mainly supports one-time or infrequent decisions, and as a result the analytical systems are at most weakly integrated with the digital infrastructure. This stands in contrast to the ongoing usage and stronger integration needed for process innovation with analytics.

Process innovation with analytics is essentially IT-enabled process innovation. It therefore shares similarities with BPR. This was also noted by Davenport & Miller (2022) in their study of AI implementations. As was the case with BPR, they noted that: "The AI system had to be deeply integrated with the organization's existing technology infrastructure and embedded in […] processes". Our experience suggests that this is not only true for advanced analytics such as AI but holds whenever any type of analytics is embedded in operational processes and becomes an integral part of their design.

Process innovation with analytics is, however, also different from BPR. For one, the scope of process innovation with analytics is most commonly at the subprocess or task-level (Sedera et al., 2016), while BPR is cross-functional in scope (Davenport, 1993; Hammer and Champy, 1993). In process innovation with analytics, the use of analytics will change the tasks themselves as they become automated or augmented but is also commonly accompanied by changes to upstream or downstream tasks (Raisch & Krakowski, 2021, p. 197), and the addition of new tasks such as (re-)training models and monitoring data and model quality (Grønsund & Aanestad, 2020). Another difference is the uncertainty associated with the capabilities of analytical systems as compared to traditional IT capabilities. With model-based analytics systems experimentation and prototyping using context-specific organizational data is required to estimate their performance (Dolata et al., 2022). Development thus becomes necessary prior to process redesign. In the following subsections, we expand on each of the three perspectives that form our conceptualization.

2.1 The Analytics Perspective

Analytics as traditionally practiced remains a core part of process innovation with analytics. Analytics is concerned with the use of data processing technology and quantitative methods to enable improved decision-making and action-taking. The process of obtaining value from analytics has been conceptualized as the transformation process of data to information, information to insights, insights to decisions, and decisions to action. Key activities in this transformation process are understanding the business context and the business problem to be addressed, understanding the data available to address the problem, preparing and processing the data, modelling of the data, evaluation of the analytical outputs, and finally deployment of the analytical outputs, if deemed desirable. Historically, deployment has often taken the shape of standalone applications or dashboards that are only weakly integrated with the technical infrastructure, such as through manual or automated batch extracts of data. This is, however, changing with the operational usage of analytics.

2.2 The Process Innovation Perspective

Adopting a process innovation perspective entails viewing analytics as the technological lever or means for improvement of processes through its informational or automation capabilities. In IT-enabled process innovation, the capabilities of IT become inputs to the design of a new and improved process (Davenport & Short, 1990). The new process design then becomes the blueprint and specification for development of the IT applications necessary to support the redesigned process. In addition to process design and IT development, realizing the redesigned process can require organizational changes such as changes to the organizational structure, new job roles, and compensation schemes (Davenport, 1993). Owing to the complexity of such projects, prototyping the new processes was recommended as a way of confirming whether the new designs worked as planned. All these aspects remain relevant for process innovation with analytics with a few caveats. First, extensive exploration and experimentation with analytics is required before new processes can be properly designed due to the uncertainty surrounding their capabilities. Second, a clean-slate approach as preached in the BPR-era is no longer applicable, as the data from the existing process will often be the key input to analytics development.

One of the key obstacles to realizing IT-enabled process innovation turned out to be IT infrastructure. Rather than being limited to the redesign of processes and supporting IT-applications, organizations discovered that IT-enabled process innovation furthermore could require considerable development of their IT infrastructure. The existing level of IT infrastructure in organizations thus act as either an enabler or a constraint in the realization of IT-enabled process innovation. This is also the case for process innovation with analytics, which brings us to the next perspective.

2.3 The Digital Infrastructure Perspective

The digital infrastructure is the landscape of IT applications and infrastructure that enables the organization to carry out work, as well as its' users and developers. The emphasis on digital infrastructure in research emerged as a result of the increased ubiquity of IT in organizations. Whereas successful introduction of a new IT application to support work once consisted mainly of developing and implementing an application that met user requirements, the pervasiveness of IT meant that new applications increasingly had to meet additional requirements that ensured a fit between the new system and the existing digital infrastructure. In the context of analytics, new analytical applications thus need to consider both how they enable new process designs, and how they extend and fit with the existing digital infrastructure. Managing the growth and development of the often-complex digital infrastructure to support organizational and process change has proven to be challenging and conflicting philosophies and approaches exist. The two main approaches in research are that of 1) the top-down and modelling-based enterprise architecture approach (e.g., Ross et al., 2006), and 2) the bottom-up and evolutionary information infrastructure approach (e.g., Ciborra et al., 2000).

All process innovation with analytics requires the development of infrastructure, but the nature and extent of development required differs across use-cases. In general, three types of infrastructure changes can be necessary to realize process innovation with analytics. First, analytics infrastructure is required to provide storage and computation resources for analytics development and deployment. Examples include data lakes, data warehouses, and clusters. Second, changes to the existing digital infrastructure can be required to collect data and to provide the necessary access to data and functionality required by the analytics systems. Third, linkages and integrations between the analytics infrastructure and digital infrastructure can be required. This includes both integrations that move data from the digital infrastructure to the analytics infrastructure and integrations that provide analytical outputs such as predictions, recommendations, and prescriptions to the digital infrastructure. In cases where data is already collected, made easily available by the digital infrastructure, and the analytical infrastructure is mature, then the infrastructure development required to realize process innovation with analytics can be relatively modest. In other cases, the infrastructure development required can be considerable.

To summarize, analytics for process innovation requires a broader conceptualization than present in existing analytics research and it differs from traditional IT-enabled process innovation in that it deals with weakly structured technologies resulting in more local process change. Specifically, process innovation with analytics requires coordinated development of analytics systems, development and redesign of processes, and development of the digital infrastructure, such as new analytics infrastructure, changes to the digital infrastructure, and linkages between them. It is thus a complex undertaking involving stakeholders with expertise in process management, IT, and data science. We now turn to the implications of this broader conceptualization for research into analytics and process innovation.

3 Research Directions & Challenges

In this section, we present three research directions that have the potential to significantly move forward our understanding of process innovation with analytics. For each of the three research directions we present associated research challenges, examples of research questions, as well as suggest promising angles of inquiry, as summarized in Table 1. The three proposed research directions align with our conceptualization of analytics: the first adopts the infrastructure perspective, the second examines the relationship between process and analytics development, and the third adopts a holistic perspective of how to govern the transformation of the three elements of infrastructure, analytics, and process. In the following subsections, we introduce and elaborate on each of these three research directions.

Research Direction	**Research Challenge**	**Exemplar Research Questions**	**Promising Angles**
Digital infrastructures for process innovation with analytics	Understanding the influence of the existing digital infrastructure	How does the socio-technical data architecture influence analytics development and deployment? How does the organization structure impact collaboration between data scientists and software developers?	Adoption of an information infrastructure lens in qualitative research (e.g., Ciborra et al., 2000; Hanseth & Lyytinen et al., 2010)
	Transition strategies: From constraining to enabling infrastructure	What infrastructure building strategies exist and what are their advantages and disadvantages? How can organizations move from silo-oriented legacy systems to enabling modern architectures?	Contextualization of insights from enterprise architecture (Ross et al., 2006) and classic infrastructure literature (Broadbent et al., 1999). Exploring the potential applicability of *data ecosystems* (Gröger, 2021) and *platformization* (Hanseth & Bygstad, 2018).
Exploring the relationship between process change and analytics development	Understanding the impact of the existing process	How do characteristics of the existing process impact process innovation with analytics? To what extent must the existing process change to enable analytics supported redesign?	Retrospective and longitudinal case studies of process innovation with analytics focusing on the process of *process change* and the benefits realized.

	Constructing integrated methodologies for developing analytics-enabled processes	How should process redesign and analytics development be sequenced? How can mutual adaptation of process and analytics systems be enabled?	Contextualizing and integrating analytics development and process innovation methodologies.
Advancing our understanding of governance of process innovation with analytics	Identifying and understanding successful governance configurations	Which governance configurations are associated with success? How do contextual conditions influence governance? How do governance mechanisms change throughout the analytics and process lifecycle?	Contextualizing existing governance models, e.g., Lightweight IT (Bygstad & Iden, 2017) Case studies and configurational analysis to examine interplay of context, governance configurations, and their impacts.

Table 1: Research agenda for process innovation with analytics.

3.1 Digital Infrastructures for Process Innovation with Analytics

***Why it is important**:* Digital infrastructures play a key role in IT-enabled process innovation in general and the same holds true for analytics, where it impacts both development and deployment. It is well established in digital infrastructure research (e.g., Ciborra et al., 2000) and in analytics research (e.g., Gröger, 2021; Vial et al., 2021) that the digital infrastructure can act as either an enabler or constraint. Research has also provided insights into the characteristics of infrastructures that are generative of innovation, that being decentralized control and loose coupling (Henfridsson & Bygstad, 2013) and proposed more concrete architectures inspired by the platform architecture (Gröger, 2022; Bygstad & Hanseth, 2018). A lot is thus known about the "to-be" of enabling digital infrastructures, but the reality is that most companies are far from this state. What is missing is: 1) knowledge about how to navigate within the non-ideal digital infrastructure that organizations possess, and 2) effective strategies for transitioning towards the "to-be" state. The two research challenges below address these in turn.

3.1.1 Challenge 1: Understanding the influence of the existing digital infrastructure

The existing digital infrastructure, that is the installed base of IT infrastructure, IT systems, developers, and users in an organization, has a large impact on the effort required to conduct process innovation with analytics. IT-enabled Process innovation projects need to interact with the installed base to both access data and functionality present in existing IT systems (Bygstad & Øvrelid, 2020). For process innovation with analytics, historical data is needed in the development phase to develop and evaluate analytical models, whereas fresh data is needed as input to the analytics system during deployment. The deployment phase will often also need to be able to access functionality in existing IT systems to enable acting on analytical outputs, e.g., by updating information in the system or triggering a workflow. A promising angle for further research is to leverage concepts from the information infrastructure literature (e.g., Ciborra et al., 2000; Aanestad & Jensen, 2011; Bygstad & Øvrelid, 2020). This stream of research has developed rich concepts for qualitatively analyzing how the installed base of infrastructure influences development. One area where this lens could be fruitfully applied is *data accessibility*, which has been identified as a key barrier to deploying analytics systems om recent research (Vial et al., 2021). Essentially, deployment requires programmatic access to data and functionality, such as via. APIs or data pipelines, but this functionality is often missing in many organizations. While the importance of data accessibility is non-controversial, a more nuanced understanding of data availability and its impact on the analytics development process is still needed. Different technical resources supporting data accessibility are likely to have subtle effects on the ease of development. At least as important are the processes through which the technical resources are made accessible. Data accessibility aspects includes how resources are discovered, how access is requested and granted, and the ease with which the data is technically accessed, e.g., as facilitated by documentation, software packages, or technical coaching. While research has started suggesting best practices for data accessibility and management under the header of "data democratization", little empirical research has investigated in detail the impacts of these practices on analytics and software development processes. In addition to the processes associated with accessibility, understanding the impact of structural aspects of the installed base remains important. Data scientists and IT developers need to collaborate during development and deployment, but our experience and recent research (Nahar et al., 2022) suggests that this collaboration is challenged by differences in backgrounds, skillsets, and practices. Further research is thus required on how the organization of this relationship influences the development and deployment process.

3.1.2 Challenge 2: Transition strategies: From constraining to enabling infrastructures

Transitioning to digital infrastructures that enable process innovation with analytics requires significant development of the digital infrastructure in most organizations. The change encompasses both the need for new analytics infrastructure, development of the operational digital infrastructure, and linking the two infrastructures. Organizations can approach this transition in several ways. Crucial decisions include the degree of reuse and standardization of analytics infrastructure, where the approach can range from local solutions to one global solution for each of the components of the infrastructure. Another decision concerns when and how to realize the infrastructure. Here the main decision is what to build up-front through infrastructure investments and what to build through projects. On the one hand, infrastructure such as data pipelines can be difficult to justify without a specific use-case. For the same reason, seminal works on IT infrastructure (Broadbent et al., 1999) and enterprise architecture (EA) (Ross et al., 2006) suggests implementing infrastructure as part of projects. On the other hand, recent research (Bygstad,

2017) and our own experience suggests mixing infrastructure and innovation in the same project can be troublesome. Infrastructure and innovation work requires different skillsets and operate at different speed, resulting in tensions and slowing down of innovation. Perhaps because of this, much analytics development does not end up moving the organization towards the to-be architecture and instead develop piecemeal solutions (Gröger, 2018). The question of how to address the transition is thus far from settled and deserves to be revisited in the context of analytics. Empirical research is needed on how successful organizations have approached the transition, along with insights into the challenges faced and benefits and disadvantages of their chosen approach, whether that be ecosystem or platform approaches, or more classic enterprise architecture approaches.

3.2 Exploring the relationship between process change and analytics development

Why it is important: Realizing process innovation with analytics requires both the development of an analytics system offering a set of analytical capabilities and a redesigned process that leverages these capabilities. With traditional IT-enabled process innovation, process design took design precedence and determined requirements for the IT solution (Davenport, 1993). The existing process played only a small role, as the ideal was to start from a blank slate (Grover & Kettinger, 1995). In process innovation with analytics, the relationship between process redesign and analytics is considerably more complicated and nuanced. On the one hand, the analytical capabilities of analytics systems are often uncertain until a prototype has been built (May et al., 2020) and dependent on data from the existing process (Vial, 2023). It thus makes it difficult to redesign processes first and increases the significance of the existing process. On the other hand, some idea of the future process is necessary to guide analytics development. As noted by Swanson (2019) the prioritization of either the process or IT artefact in IT-enabled process innovation is a choice with major implications. However, existing research has yet to 1) address when and whether to pursue incremental or more radical redesigns and 2) sufficiently engage with the issue of how to sequentially organize analytics and process development. One exception is May et al. (2020) which suggests striking a middle ground. The two research challenges below address these two issues.

3.2.1 Challenge 3: Understanding the impact of the existing process

The fact that the as-is process is responsible for generating the historical data that is used to train and evaluate models in analytics development means that it takes on a central role in establishing the process innovation affordances of analytics. Vial (2023) refers to this as the "dual role of data as both output [of processes] and input [to analytics]". However, as-is processes have mostly been designed for efficient execution and without major consideration of generating data that supports analytical development. The result is that certain process innovation affordances of analytics remain impossible to actualize with the data generated by the existing processes. Managers thus need to decide whether to invest in making intermediary changes to their process to collect the data that would enable future analytics-enabled redesign of their processes. For now, research has not examined the extent of "preparatory" change required to existing processes to enable future redesigns and the role that process characteristics play in this. There is furthermore a need for research into whether and when these preparatory changes and investments pay off. Should managers focus their process innovation efforts on incremental redesigns, i.e., those already possible based on the existing processes, or should they focus on potential and make the necessary investments to transform the existing processes? More radical redesigns potentially present managers with a paradox. Whereas the process innovation with analytics is mainly pursued for productivity gains (Kunz et al., 2020, Enholm et al., 2021), the intermediate investments in data collection required to realize more radical redesigns are likely to lower productivity in the short term. Longitudinal case studies are ideal candidates for gaining insights into the process change and investments that are required to realize process innovation with analytics as the innovation process unfolds, whereas retrospective case studies can provide insights into the benefits realized.

3.2.2 Challenge 4: Developing integrated methodologies for analytics-enabled processes

The feasibility of an analytics-enabled process redesign often hinges on the predictive or prescriptive performance that can be realized by the analytics system, but this performance is unknown until a model has been evaluated on a context-specific dataset. High levels of analytical performance might afford automation of a task, whereas lower levels are likely to require different levels of human engagement, or even render the redesign undesirable. Existing research suggests two different approaches to IT-enabled process innovation, both of which do not exactly fit the characteristics of analytics: 1) BPR-era methodologies (e.g., Davenport (1993); Hammer and Champy, 1993), and 2) newer *lightweight* digital process innovation approaches inspired by agile development (e.g., Schmiedel & vom Brocke, 2015; Bygstad, 2017; Bygstad & Øvrelid, 2020). BPR methodologies suggest upfront process design and modelling, which is difficult due to the uncertain capabilities of analytics. On the other hand, the digital process innovation approach recommends using standard and customizable technologies and relying on iterations of mutual adaptation of the technology and process (Schmiedel & vom Brocke, 2015; Bygstad & Øvrelid, 2020). This approach is, however, challenged by the fact that much analytics software remains the result of custom development. While this might change in the future, knowledge is needed as to how to approach the integrated development of processes and their supporting analytics systems. In practice, we mainly see the bottom-up approach where the analytical artefact takes design precedence and process concerns are introduced later. However, senior leaders at our industrial partner have expressed interest in more strategic and process-oriented approaches to analytics. There is a need for prescriptive guidelines and methodologies that consider how, when, and if traditional process tasks such as process modelling, simulation, and organizational prototyping should enter the innovation process and relate to analytics development. Contextualization and integration of existing methodologies from both analytics and IT-enabled process innovation seems a highly promising direction with considerable potential to improve practice. Due consideration of process aspects in these innovation projects has the potential to reduce both the number of analytical systems that are discarded due to low business value and the extent of adjustments required in the deployment process to make the system usable.

3.3 Advancing our understanding of governance of process innovation with analytics

***Why it is important*:** Governance of process innovation with analytics is complex as it comprises a variety of assets ranging from data, models, and IT systems to processes and involves stakeholders from across the organization. Whereas governance of data, IT systems, and processes are all established research domains individually, research into the governance of analytics models (often termed AI or ML governance) is in its early days, although frameworks have started to emerge (e.g., Schneider et al., 2022). Existing research has established that IT-enabled process innovation covers concerns of both IT governance and process governance and has thus suggested the need for relationships between and alignment of process and IT governance (Rahimi et al., 2016). The same holds true for governance of analytics-enabled process innovation with the additional need to establish relationships from IT governance and process governance to analytics governance. There is thus a need to further our understanding of both how to effectively govern analytics and how to align analytics governance with process and IT governance to bring about process innovation with analytics. Governance does, however, not end after successful implementation of a new analytics-enabled process but must consider the analytics-enabled process throughout its lifecycle. We turn to the challenge of understanding what makes for successful governance next.

3.3.1 Challenge 5: Identifying and understanding successful governance configurations

The domain-crossing nature of process innovation with analytics challenges traditional governance frameworks. Early research on IT governance focused on the trade-off between control and innovation, or centralization vs. decentralization (Brown & Grant, 2005) and the distribution of decisions rights between IT and business (Weill & Ross, 2004). As a response to the influx of digital innovations and further demands for agility in recent times, many IT organizations have adopted more complex organizational structures and governance arrangements. The most popular has perhaps been bimodal IT (Gartner, 2014; Haffke et al., 2017), where innovation and execution-focused IT are governed and organized differently. More recently, further approaches such as platform governance have been suggested as means to strike a balance between innovation and control for the case of new *lightweight* digital technologies (Bygstad & Iden, 2017; Gregory et al., 2018). For analytics in particular, platform and ecosystem architectures and governance approaches have likewise been suggested (Gröger, 2021) and are widely implemented in practice. Nonetheless, these models do not cover the full scope of activities and assets involved in process innovation with analytics. Governance of process innovation with analytics should cover both the activities involved including use-case specification, development, deployment, organizational implementation, and operations, as well as the assets involved, such as data, models, IT systems, IT infrastructure, and processes. There is thus a need for further understanding of how the existing governance frameworks and arrangements need to be adapted. Research can contribute by identifying successful governance configurations in leading adopters of process innovation with analytics and exploring how and why these configurations support the adoption process. Furthermore, research should explore how contextual factors impact appropriate governance arrangements. We expect that aspects such as characteristics of the process to be innovated and whether the use-case is automation or augmentation is likely to significantly influence the nature of governance. In this pursuit, both case studies and configurational analysis (e.g., as in Mikalef & Krogstie, 2020) could be applied to further our understanding. Another interesting challenge concerns governance of the analytical artefact once it is in operation in a new analytics-enabled process. In this situation, the analytics performance (data quality and model performance) becomes a key driver of process performance, especially in the case of automation use-cases. How to effectively govern this setup and distribute roles and responsibilities between IT, data science, and process personnel is an interesting and important challenge and it remains to be seen how much established companies can learn from mature software companies, such as Facebook and Google, that have led the adoption of analytics in business-critical settings.

4 Conclusion

In this research note, we reconceptualized analytics in the process innovation context to highlight how it differs from traditional analytics aimed at insights. Process innovation with analytics requires coordinated development of the digital infrastructure and the new process in addition to an analytics system. This increased scope of the analytics process has major implications for the complexity facing organizations looking to adopt analytics for process innovation and calls for contextualization of the existing knowledge bases related to analytics and process innovation to support them in the process.

We drew on our practical experience with the adoption of analytics for process innovation in an ADR study at a large global manufacturer and retailer based in Denmark to present a research agenda for process innovation with analytics. Specifically, we proposed the need for research along three directions: 1) digital infrastructures for process innovation with analytics, 2) the relationship between process change and analytics development, and 3) governance of process innovation with analytics. The literature is full of technically focused research on analytics. What is needed to support organizations in the adoption and value realization process is research that is socio-technical. The combined focus on engineering and business present in the BISE community, along with its focus on delivering practically relevant research, makes it an ideal setting for furthering our understanding of process innovation with analytics. We thus hope that the BISE community will join us in our efforts to better understand process innovation with analytics.

References

Benbya H, Pachidi S, Jarvenpaa SL (2021) Special Issue Editorial: Artificial Intelligence in Organizations: Implications for Information Systems Research. JAIS 22(2):281–303. https://doi.org/10.17705/1jais.00662

Broadbent M, Weill P, St. Clair D (1999) The Implications of Information Technology Infrastructure for Business Process Redesign. MISQ 23(2):159–182. https://doi.org/10.2307/249750

Brown AE, Grant GG (2005) Framing the Frameworks: A Review of IT Governance Research. CAIS 15. https://doi.org/10.17705/1CAIS.01538

Bygstad B (2017) Generative Innovation: A Comparison of Lightweight and Heavyweight IT. Journal of Information Technology 32(2):180–193. https://doi.org/10.1057/jit.2016.15

Bygstad B, Hanseth O (2018) Transforming Digital Infrastructures through Platformization. In: ECIS 2018 Proceedings. p 13

Bygstad B, Iden J (2017) A Governance Model for Managing Lightweight IT. In: Rocha Á, Correia AM, Adeli H, Reis LP, Costanzo S (eds) Recent Advances in Information Systems and Technologies. Springer International Publishing, Cham, pp 384–393

Bygstad B, Øvrelid E (2020) Architectural alignment of process innovation and digital infrastructure in a high-tech hospital. European Journal of Information Systems 29(3):220–237. https://doi.org/10.1080/0960085X.2020.1728201

Ciborra C, Braa K, Cordella A, Dahlbom B, Hepsø V, Failla A, Hanseth O, others (2000) From control to drift: the dynamics of corporate information infastructures. Oxford University Press on Demand

Davenport TH (1993) Process Innovation. Mcgraw-Hill, Boston, Massachusets

Davenport TH (2018) From analytics to artificial intelligence. Journal of Business Analytics 1(2):73–80. https://doi.org/10.1080/2573234X.2018.1543535

Davenport TH, Harris JG, Morison R (2010) Analytics at work: Smarter decisions, better results. Harvard Business Press

Davenport TH, Miller SM (2022) Working with AI: Real Stories of Human-machine Collaboration. MIT Press

Davenport TH, Short JE (1990) The new industrial engineering: information technology and business process redesign. Sloan Management Review 31(4):11–27

Dolata M, Crowston K, Schwabe G (2022) Project Archetypes: A Blessing and a Curse for AI Development. In: ICIS 2022 Proceedings

Dremel C, Herterich M, Wulf J, Waizmann J-C, Brenner W (2017) How AUDI AG Established Big Data Analytics in its Digital Transformation. MIS Quarterly Executive 16:81–100

Dremel C, Herterich MM, Wulf J, vom Brocke J (2020) Actualizing big data analytics affordances: A revelatory case study. Information & Management 57(1):103121. https://doi.org/10.1016/j.im.2018.10.007

Eley T, Lyytinen K (2022) Industry 4.0 Implementation: Novel Issues and Directions. In: Proceedings of the 55th Hawaii International Conference on System Sciences

Enholm IM, Papagiannidis E, Mikalef P, Krogstie J (2021) Artificial Intelligence and Business Value: a Literature Review. Inf Syst Front. https://doi.org/10.1007/s10796-021-10186-w

Gartner (2014) Bimodal IT: How to Be Digitally Agile Without Making a Mess. In: Gartner. https://www.gartner.com/en/documents/2798217. Accessed 23 Aug 2023

Gregory RW, Kaganer E, Henfridsson O, Ruch TJ (2018) IT Consumerization and the Transformation of IT Governance. MIS Quarterly 42(4):39

Gröger C (2018) Building an Industry 4.0 Analytics Platform. Datenbank Spektrum 18(1):5–14. https://doi.org/10.1007/s13222-018-0273-1

Gröger C (2021) There is no AI without data. Commun ACM 64(11):98–108. https://doi.org/10.1145/3448247

Grønsund T, Aanestad M (2020) Augmenting the algorithm: Emerging human-in-the-loop work configurations. The Journal of Strategic Information Systems 29(2):101614. https://doi.org/10.1016/j.jsis.2020.101614

Gross S, Malinova M, Mendling J (2019) Navigating Through the Maze of Business Process Change Methods. In: Proceedings of the 52nd Hawaii International Conference on System Sciences

Grover V, Kettinger WJ (1995) Business process change: Concepts, methods, and technologies. IGI Global

Haffke I, Kalgovas B, Benlian A (2017) Options for Transforming the IT Function Using Bimodal IT. MIS Quarterly Executive 16(2):101–120

Hammer M, Champy J (1993) Business process reengineering. London: Nicholas Brealey 444(10):730–755

Hanseth O, Lyytinen K (2010) Design Theory for Dynamic Complexity in Information Infrastructures: The Case of Building Internet. Journal of Information Technology 25(1):1–19. https://doi.org/10.1057/jit.2009.19

Henfridsson O, Bygstad B (2013) The Generative Mechanisms of Digital Infrastructure Evolution. MIS Quarterly 37(3):907–931

Kim G, Shin B, Kim K, Lee H (2011) IT Capabilities, Process-Oriented Dynamic Capabilities, and Firm Financial Performance. JAIS 12(7):487–517. https://doi.org/10.17705/1jais.00270

Kunz PC, Spohrer K, Heinzl A (2020) The Impact of Business Analytics: A Process-Level Perspective. ICIS 2020 Proceedings :18

Martinez I, Viles E, Olaizola IG (2021) A survey study of success factors in data science projects. In: 2021 IEEE International Conference on Big Data (Big Data). pp 2313–2318

May A, Sagodi A, Dremel C, van Giffen B (2020) Realizing Digital Innovation from Artificial Intelligence. In: ICIS 2020 Proceedings. Forty-First International Conference on Information Systems

Mikalef P, Krogstie J (2020) Examining the interplay between big data analytics and contextual factors in driving process innovation capabilities. European Journal of Information Systems 29(3):260–287. https://doi.org/10.1080/0960085X.2020.1740618

Nahar N, Zhou S, Lewis G, Kästner C (2022) Collaboration challenges in building ML-enabled systems: communication, documentation, engineering, and process. In: Proceedings of the 44th International Conference on Software Engineering. ACM, Pittsburgh Pennsylvania, pp 413–425

Nalchigar S, Yu E (2020) Designing Business Analytics Solutions. Bus Inf Syst Eng 62(1):61–75. https://doi.org/10.1007/s12599-018-0555-z

Phillips-Wren G, Daly M, Burstein F (2021) Reconciling business intelligence, analytics and decision support systems: More data, deeper insight. Decision Support Systems 146:113560. https://doi.org/10.1016/j.dss.2021.113560

Rahimi F, Møller C, Hvam L (2016) Business process management and IT management: The missing integration. International Journal of Information Management 36(1):142–154. https://doi.org/10.1016/j.ijinfomgt.2015.10.004

Raisch S, Krakowski S (2021) Artificial Intelligence and Management: The Automation–Augmentation Paradox. Academy of Management Review 46(1):192–210. https://doi.org/10.5465/amr.2018.0072

Ross JW, Weill P, Robertson D (2006) Enterprise architecture as strategy: Creating a foundation for business execution. Harvard business press

Schmiedel T, vom Brocke J (2015) Business Process Management: Potentials and Challenges of Driving Innovation. In: vom Brocke J, Schmiedel T (eds) BPM - Driving Innovation in a Digital World. Springer International Publishing, Cham, pp 3–15

Schneider J, Abraham R, Meske C, Vom Brocke J (2023) Artificial Intelligence Governance For Businesses. Information Systems Management 40(3):229–249. https://doi.org/10.1080/10580530.2022.2085825

Sedera D, Lokuge S, Grover V, Sarker S, Sarker S (2016) Innovating with enterprise systems and digital platforms: A contingent resource-based theory view. Information & Management 53(3):366–379. https://doi.org/10.1016/j.im.2016.01.001

Swanson EB (2019) Technology as Routine Capability. MISQ 43(3):1007–1024. https://doi.org/10.25300/MISQ/2019/14653

Tarafdar M, Beath CM, Ross JW (2019) Using AI to enhance business operations. MIT Sloan Management Review 60(4)

Tim Y, Hallikainen P, Pan SL, Tamm T (2020) Actualizing business analytics for organizational transformation: A case study of Rovio Entertainment. European Journal of Operational Research 281(3):642–655. https://doi.org/10.1016/j.ejor.2018.11.074

Vial G (2023) Data governance and digital innovation: A translational account of practitioner issues for IS research. Information and Organization 33(1):100450. https://doi.org/10.1016/j.infoandorg.2023.100450

Vial G, Jiang J, Giannelia T, Cameron A-F (2021) The Data Problem Stalling AI. MIT Sloan Management Review 62(2):47–53

Volkoff O, Strong DM (2013) Critical Realism and Affordances: Theorizing It-Associated Organizational Change Processes. MISQ 37(3):819–834

Weill P, Ross JW (2004) IT governance: How top performers manage IT decision rights for superior results. Harvard Business Press

Wirth R, Hipp J (2000) CRISP-DM: Towards a Standard Process Model for Data Mining. In: Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining. p 11

Zhang D, Pan SL, Yu J, Liu W (2022) Orchestrating big data analytics capability for sustainability: A study of air pollution management in China. Information & Management 59(5):103231. https://doi.org/10.1016/j.im.2019.103231

Paper 3: Speeding up Explorative BPM with Lightweight IT: The Case of Machine Learning

Bojer, Casper Solheim, Bygstad, Bendik, and Øvrelid, Egil. (2023). *Speeding up Explorative BPM with Lightweight IT: The Case of Machine Learning* [Manuscript submitted for publication].

The paper has been submitted to *Information Systems Frontiers* and is currently undergoing revisions before a second round of review.

Speeding up Explorative BPM with Lightweight IT: The Case of Machine Learning

Casper Solheim Bojer[1*], Bendik Bygstad[2], & Egil Øvrelid[2]

[1] Department of Materials & Production, Aalborg University, Fibigerstræde 16, Aalborg, Denmark
csb@mp.aau.dk

[2] Department of Informatics, University of Oslo, Gaustadalléen 23 B,
Oslo, Norway

* Corresponding author

Abstract

In the modern digital age, companies need to be able to quickly explore the process innovation affordances of digital technologies. This includes exploration of Machine Learning (ML), which when embedded in processes can augment or automate decisions. BPM research suggests using lightweight IT (Bygstad, 2017) for digital process innovation, but existing research provides conflicting views on whether ML is lightweight or heavyweight. We therefore address the research question *"How can Lightweight IT contribute to explorative BPM for embedded ML?"* by analyzing four action cases from a large Danish manufacturer. We contribute to explorative BPM by showing that lightweight ML considerably speeds up opportunity assessment and technical implementation in the exploration process thus reducing process innovation latency. We furthermore show that succesful lightweight ML requires the presence of two enabling factors: 1) loose coupling of the IT infrastructure, and 2) extensive use of building blocks to reduce custom development.

Keywords: Machine Learning, Explorative BPM, Digital Process Innovation, Lightweight IT, IT Infrastructure

1 Introduction

In the current digital age, companies are faced with an ever-growing plethora of affordances for innovating and improving business processes in both incremental and radical ways. This situation of opportunity abundance has created the need for fast exploration of digital process affordances in addition to the traditional problem-solving focus on existing processes (Rosemann, 2014; Grisold et al., 2019). While much IS process innovation research predates the digital age, exploratory process innovation research has picked up in the BPM community in recent years under the term of Explorative BPM. Explorative BPM concerns opportunity rather than problem-driven process change with one key source of opportunities being those arising from new digital technologies (Grisold et al., 2019). Explorative BPM research has so far focused on digital technologies at a general level and considered the opportunity identification stage (Grisold et al., 2022; Gross et al., 2021), implementation success factors (Baier et al., 2022), and digital capabilities (Kerpedzhiev et al., 2021). Machine Learning (ML) is one technology that offers process innovation affordances by being embedded into business processes and augmenting or automating decisions and actions (Davenport, 2018). However, in practice, companies are struggling to successfully deploy ML (Paleyes et al., 2022; Davenport & Malone, 2021). ML is a complex technology, typically the result of extensive custom development and firmly situated in the domain of IT. It thus seems incompatible with the need for fast exploration and process (rather than technology) focus of explorative BPM.

Previous research has leveraged the concept of *lightweight IT* (Bygstad, 2017) to examine the role of IT in digital process innovation. This research suggests that successful digital process innovation should use lighter engineering and *lightweight IT* (Schmiedel & vom Brocke, 2015; Baiyere et al., 2020; Bygstad & Øvrelid, 2020) that is aligned with the *heavyweight* digital infrastructure, achieved through boundary resources and loose coupling (Bygstad & Øvrelid, 2020). Existing research, however, provides conflicting views of whether ML can be viewed as lightweight IT. One stream of research takes the traditional view of ML and suggests that successful ML requires rigorous and systematic engineering (e.g., Lavin et al., 2022), placing it firmly in the realm of *heavyweight IT*. Another stream of research suggests that recent innovations in ML platforms and tooling will make it possible for users with domain knowledge to build their own systems (e.g., Hutter et al., 2019), thus placing ML in the realm of *lightweight IT*. Having identified this tension and uncertainty in the existing conceptualization of ML, we set out to answer the following research question:

How can Lightweight IT contribute to explorative BPM for embedded ML?

To answer the research question, we examined four embedded ML projects at a large Danish manufacturer on the basis of ongoing engaged scholarship. The projects all attempted to treat ML as lightweight IT, but achieved different degrees of success in doing so. We find evidence of successful lightweight ML, thus showcasing that the traditional idea of ML as a heavyweight technology need not be true. Making use of *lightweight ML* resulted in significantly speeding up the exploratory initiatives compared to heavier approaches. We further show that this speeding up is due to 1) loose coupling, which enables exploratory work to be carried out independently of heavyweight IT, and 2) the extensive use of building blocks, allowing steps of the traditional ML process to be skipped. Our findings thus contribute to explorative BPM by providing insights into how to speed up the assessment and technical implementation phases of explorative BPM (Rosemann, 2014).

2 Background

2.1 Explorative BPM in a Digital Age

Business Processes Management (BPM) is a holistic approach with the goal of ensuring effective and efficient business processes: the coordinated sequences of work involved in delivering products and services (Dumas et al., 2013). Improving process performance is a key concern of BPM and it can be approached using means ranging from incremental (e.g., Lean or Six Sigma) to radical (e.g., BPR). Although IT has long been recognized in BPM as a key enabler of process improvement and innovation (Davenport & Short, 1990), BPM has come under critique for failing to explore (Benner & Tushman, 2003) and take advantage of the many process innovation affordances offered by new digital technologies (Rosemann, 2014; Grisold et al., 2019). One proposed explanation for this failure is that there are significant differences in the underlying logics and assumptions of BPM and digital innovation, with the implication being that BPM should reconsider and update its foundations and methods (Mendling et al., 2020; Baiyere et al., 2020).

Initial work on this revisiting of the foundation has suggested that lighter and more flexible approaches to BPM is required in the digital age. Baiyere et al. (2020) suggested a shift in logic towards light rather than strongly modelled processes, flexible rather than aligned infrastructure, and mindful actors rather than routine followers. The use of lightweight IT for process innovation was suggested by Bygstad & Øvrelid (2020), and they further clarified the need for alignment between the digital infrastructure and lightweight IT to enable successful digital process innovation. Lightweight IT is a knowledge regime characterized by a focus on business-led innovation, competent users, emergent architectures, work process support, and light & standard technologies (Bygstad, 2017). It is juxtaposed with Heavyweight IT, the realm of traditional IT delivery, that is characterized by a focus on transaction processing, systematic engineering, integrated architectures, and proven technologies (Bygstad, 2017). While lightweight IT is focused on innovation in processes, it is nonetheless often reliant on heavyweight IT for access to data and functionality, which can pose challenges for Lightweight IT as the knowledge regimes operate at different speeds. A solution to this challenge has been proposed using the notion of coupling, suggesting that lightweight IT and heavyweight IT should be loosely coupled and aligned rather than integrated, both technically (e.g., in terms of IT architecture) and organizationally (e.g., in terms of developer and user communities) (Bygstad, 2017; Bygstad & Øvrelid, 2020). Coupling signifies the degree and nature of interdependence between two systems. Loose coupling refers to systems exhibiting individual identity and separateness, that are weakly interacting or doing so with minimal dependence, thus exhibiting and requiring little coordination (Weick, 1976). Tight coupling refers to the opposite situation, where systems are strongly interacting and highly interdependent. One of the benefits of loosely coupled systems is that it allows the addition or removal of components to the system without significant impact on other components and the overall system (Weick, 1976).

2.2 Embedded ML to Improve Business Processes

The role of analytics and ML in business process improvement is changing with ML increasingly being embedded in business processes (Davenport, 2018). Traditionally, analytics and ML have been used *offline* in business process work by process analysts or data scientists to discover process insights (Davenport et al., 2010) that could be used for either control or redesign actions. Prime example of this *offline* approach is using either process mining (van der Aalst, 2012) or business process analytics (Lang et al., 2015) to discover bottlenecks in processes, based on which resources can be added or processes can be redesigned. Embedded ML on the other hand, concerns leveraging a ML system as part of the business process at run-time. Improving business processes with embedded ML thus requires the development and implementation of a ML system along with a redesign of the process to take advantage of the ML system. Examples falling under the header of embedded ML from the BPM community includes recent work on Process Forecasting (Poll et al., 2018) and Predictive Process Monitoring (e.g., Teinemaa et al., 2019). Our use of the term Embedded ML in business processes, is however, broader than these two areas as it covers any situation in which a ML system is used as part of the execution of a business process, notwithstanding how it is used.

The role of embedding ML and AI in business processes has been primarily conceptualized as either automating or augmenting tasks rather than the whole process (Benbya et al., 2021; Sedera et al., 2016). However, task-level changes can enable redesigning the process (Enholm et al., 2021; May et al., 2020), or require either addition of new tasks or changes in task composition and structure (Kunz et al., 2022; Grønsund & Aanestad, 2020). The value and impact of ML and AI from a process perspective is in improving process efficiency in terms of increased productivity and reduced errors (Enholm et al., 2021)

2.3 Machine Learning Systems: Heavy or Light?

Machine Learning is a subset of artificial intelligence concerned with using computers to conduct tasks based on learning from data (Enholm et al., 2021). Common tasks include clustering, classification, regression, making recommendations, or prescribing actions. ML is different from traditional software in that 1) its output is probabilistic, with the affordances offered depending on the particular dataset used for learning, and 2) it has models and datasets as key artefacts (Paleyes et al., 2022). As a result, the ML development process differs from traditional software engineering, having more in common with methods from Decision Support Systems. Process models conceptualize ML systems development as an iterative process (Wirth & Hipp, 2000; Martínez-Plumed et al., 2019; Microsoft, 2023; Lavin et al., 2022), concerned with four major groups of activities: 1) *Problem Framing*, where the business problem is understood and translated into a ML problem, 2) *Data Acquisition & Preparation*, where the necessary data is acquired and preprocessed, 3) *Model Development*, where a ML model is trained and evaluated using data, and 4) *Deployment*, where the ML model is turned into a ML system, deployed, and integrated with other systems if required, resulting in a usable IT artefact.

Traditionally, ML research and practice has been narrowly focused on the model, developed by a specialized data scientist or statistician, and has treated deployment as a simple matter of implementation, often done by IT. The main challenge in this view is thus to find the right combination of data and model that yields useful analytical capabilities, measured in terms of quantitative performance metrics. Partly responsible for this focus, is the fact that many ML initiatives are stopped following *Model Development*, due to a failure to obtain satisfactory analytical performance. Following challenges faced in practice in the *Deployment* stage, where models developed either were difficult to integrate with the existing infrastructure or failed to solve the intended business problem (Davenport & Malone, 2021), focus has shifted towards the ML system rather than the model (Ratner et al., 2019; Lavin et al., 2022). It has been recognized that ML models are only a small part of a complex infrastructure required to deploy ML (Sculley et al., 2015). This research stream has taken an engineering perspective and focused on the development of new tools (e.g., Breck et al., 2019), infrastructure (e.g., Phillips-Wren et al., 2021), and systematic methods (Lavin et al., 2022). Calls have likewise been made for increased rigour in ML development, which should borrow from well-established system engineering methods (Lavin et al., 2022).

Concurrently with the calls for increased rigour, other research streams have worked with the assumption that the solution to greater deployment and adoption of ML is to democratize it and allow non-specialists to develop ML systems. The democratization is enabled by technological advancements that automate parts of the ML process (e.g., Hutter et al., 2019; Uzunalioglu et al., 2019) and training of employees (Lefebvre et al., 2021), thus speeding up the development process and making ML accessible for non-experts with greater knowledge of the problem domain. These developments contribute to making ML a lighter technology by reducing the needs for extensive engineering and moving technology development closer to the use domain.

Summarizing our review of the literature, we see that best-practice in digital process innovation is to use lightweight IT and focus on user-needs rather than turning initiatives into large, slow, and heavy engineering projects. ML is increasingly embedded in processes to improve their performance, but the historical use of ML and parts of the current discourse suggest that it belongs to the domain of heavyweight IT with its focus on engineering. However, technological advancements in ML are increasingly making it possible to treat ML as a lighter technology and reducing the need for engineering. Thus, current literature is unclear as to whether ML should be conceptualized as heavy or light, which in turn will have large implications for its role in exploratory process innovation.

3 Analytical Framework

Our review of the literature suggests that ML as a technology has characteristics associated with both lightweight and heavyweight IT. The development culture and application domain points towards lightweight IT, while the competences required and the nature of ML as a technology point towards heavyweight IT. Other aspects are less clear, as data science work is traditionally done outside heavyweight IT delivery either by Center of Excellences, consultants, or in business units, utilizing development methods tailored to ML. To clarify the issue of the role of lightweight IT in embedded ML, we suggest that a more detailed understanding of embedded ML is necessary, which requires opening the black box of the ML development process and examining it in relation to its context. To accomplish this, we propose the analytical framework in Figure 1 to guide our analysis of the cases. The framework builds on and synthesizes our review of the literature on explorative BPM, embedded ML, and lightweight IT.

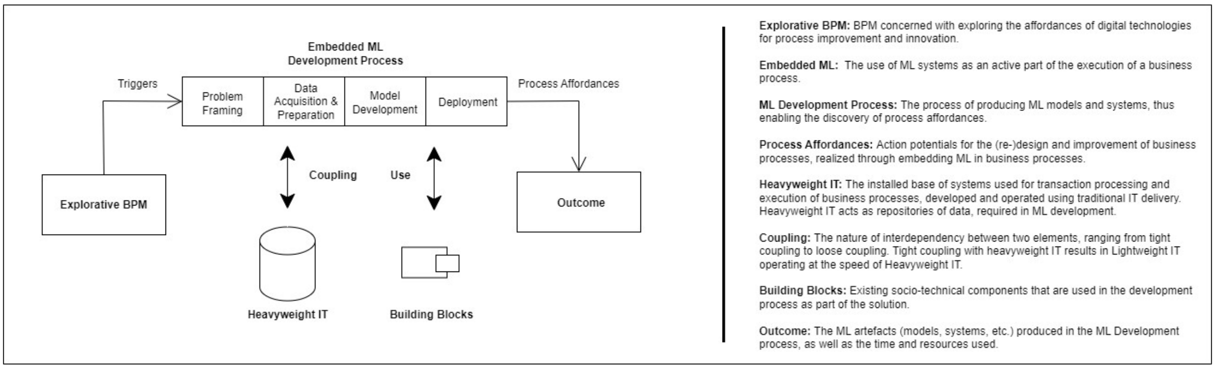


Figure 1. Analytical Framework

As illustrated in Figure 1, we adopt a processual view of embedded ML in the context of explorative BPM. Explorative BPM, concerned with the opportunity-driven exploration of digital technologies (Rosemann, 2014; Grisold et al., 2019), is part of the context and acts as a trigger for the embedded ML development process with the goal of discovering process innovation affordances of embedded ML for a particular use-case. To conceptualize the embedded ML development process in more detail, we leverage the typical four stage development process model of 1) Problem Framing, 2) Data Acquisition and Preparation, 3) Model Development, and 4) Deployment. The embedded development process produces an outcome in the form of ML models and systems, while consuming time and resources. The outcome is a success if the ML artefacts provides process affordances that enable redesign and improvement of the business process.

To analyze and explain the differences in outcome we leverage the concepts of coupling and building blocks. As emphasized in the work of Bygstad (2017) & Bygstad & Øvrelid (2020), the lightweight development process relies on data and functionality in the *heavyweight IT* systems, thus introducing a certain coupling between the heavyweight IT and lightweight IT systems. Tight coupling makes the development process dependent on heavyweight processes and resources, reducing the speed of the initiative to that of the heavyweight systems (Bygstad, 2017). We further leverage the notion of building blocks, which we define as "existing socio-technical components that are used in the development process as part of the solution". One example of a building block is a boundary resource (Ghazawneh & Henfridsson, 2013), which consists of software tools and regulations that provide external access to a platform, including a heavyweight IT infrastructure. Other examples include software-as-a-service, such as that offered by cloud vendors, or other lower level software components. Building blocks are significant, as innovation (including the development of new software systems) mostly happen as a result of recombination of existing components (Arthur, 2009; Henfridsson et al., 2018). Along similar lines, Simon (1996) noted that evolution of complex forms is more likely when leveraging "intermediate stable forms". Relying on configuration of standard technologies to facilitate digital process innovation as suggested in recent literature (Schmiedel & vom Brocke, 2015; Bygstad & Øvrelid, 2020) amounts to leveraging very high-level building blocks in the development process.

In our analytical framework, the *lightweight* or *heavyweight* nature of embedded ML is thus influenced by 1) its context, 2) characteristics of the development process, including the use, and nature of building blocks, and 3) the degree of coupling to heavyweight IT. While the presented framework suggests that embedded ML can be treated as lightweight IT in the right context, we still lack empirical evidence of embedded ML as lightweight IT. Additionally, we need to understand in which context and to what extent treating ML as lightweight IT is possible and, most importantly, whether doing so leads to improved outcomes for the exploratory initiatives.

4 Method

4.1 Research Approach: Engaged Scholarship & Action Cases

The research approach was one of engaged scholarship (Van de Ven, 2007) in the form of a series of action cases (Braa & Vidgen, 1999) aimed at both understanding and intervening. The lead author was engaged with the case company as part of a larger Action Design Research (Sein et al., 2011) project focused on process improvement using embedded ML, which provided access to rich empirical material gathered both as an observer and a designer.The case company is a large manufacturer with a strong process orientation, widely recognized as a world leader in its domain, and thus strongly capable in exploitative BPM. In recent years, the company had made significant investments in exploratory capabilities by establishing new organizational units, investing in IT infrastructure, and running many exploratory innovation initiatives. The lead author's activities made up a small part of these exploratory innovation initiatives and consisted in the participation in a total of six projects. During this participation, the lead author was employed in academia, but was sponsored by the case company. The lead author had an employee ID card, was able to access the organization freely, and collaborated with company employees in the innovation initiatives. The second and third author were not affiliated with the case company and thus played the role of outsiders, who could challenge interpretations and contribute to reducing bias (Robey & Taylor, 2018)

To answer our research question, we selected and analyzed four of these exploratory projects. Selection criteria for the initiatives were that they 1) concern embedded ML, 2) aim to improve process performance, and 3) be conducted outside traditional IT delivery, thus at least attempting to be lightweight. Of the six projects, one was not selected for analysis as it did not end up containing a ML component. Another project was dropped during analysis for reasons of brevity, as the findings were similar to that of the *Engineering Design Rework Reduction* case. Table 1 presents an overview of the cases selected. Below, we elaborate on the case context and our data collection and analyses.

Case	Organizational Unit(s)	Improvement Goal	Data Collection: Role & Materials
Engineering Design Rework Reduction	Manufacturing Engineering & Operations IT	Rework Reduction (Lead Time & Cost)	*Participative Observation:* Daily standups, sprint demos & retrospectives, & technical meetings over 16 weeks. Supported by field notes, documents, and IT artefacts.
Engineering Design Cost Prediction	Manufacturing Engineering	Cost Reduction & Predictability	*Participative Observation/Design:* Four workshops and co-designer of

			architecture & deployment. Supported by field notes and IT artefacts.
Closed-loop Control of Machines	Manufacturing R&D	IT Capability & Quality/ Robustness	*Design:* Several workshops and a two-week sprint where the solution was deployed. Supported by field notes, documents, and IT artefacts.
Data Quality Anomaly Detection	Data Platform Team	Data Quality	*Design:* Sole designer while embedded in Data Platform Team. Supported by IT artefacts and architecture drawings.

Table 1. Case Overview.

4.2 Case Context

The case organization is a large Danish manufacturer currently undergoing a digital transformation. With a long history as a manufacturer of physical products, the company has started providing digital products and is working on transforming its operations to leverage digital technologies. An organizational unit has been established to own the digital transformation of operations, consisting of several subunits responsible for specific strategic initiatives. Industry 4.0 is one of these strategic initiatives, aimed at improving productivity and responsiveness in the manufacturing processes. In the Industry 4.0 initiative, data and analytics are seen as offering potential for significant process improvement and has been the subject of several R&D initiatives and pilots. To support this transformation, Operations IT has focused on reducing technical debt and building platforms. As the existing enterprise data platform focused on marketing did not provide appropriate support for data initiatives in manufacturing, Operations IT established a small subunit that developed and operated an operations data platform leveraging modern cloud services. The data platform was envisioned as a self-service platform, where IT teams in Operations IT would deliver data to, enabling data initiatives to work decoupled from the process execution-focused IT infrastructure. The data platform provided access to data, self-service clusters capable of processing big data, and several services to support the machine learning process, such as automated model training and automated deployment. As part of establishing the data platform several data pipelines were built to transport data from the existing digital infrastructure to the data platform, but most data were still siloed in individual systems.

4.3 Data Collection & Analysis

Data was primarily collected using participative observation, although the nature of participation varied from case to case (see Table 1). In one case (*Engineering Design Rework Reduction)* the first author participated as an expert from the organization on ML and active engagement thus consisted primarily of participating in discussions and advising. In the other three cases, the first author participated more actively in the development process with engagement ranging from being a (co-)owner and designer of the development process to participating in parts of the process. The first-hand observation and experiences in the four initiatives by the lead author were documented using field notes, which represented a key source of data. The field notes focused on the rationale and improvement goals of the projects, activities undertaken in the ML development process and the actors engaged, the design of the ML models and artefacts, as well as the surrounding IT architecture. In addition to the field notes, we collected supporting data in the form of relevant documents, presentations, diagrams of IT architectures, and we had access to the IT artefacts and their underlying source code.

To analyze the data, we relied on an abductive process. The aim of abduction is the construction of a plausible and coherent explanation for “unanticipated” or “surprising” empirical findings (Dubois & Gadde, 2002; Sætre & van de Ven, 2021). The process was triggered by the observation that while the four projects had attempted to treat ML as lightweight, they had not all succeeded. To arrive at an explanation, we conducted an analysis of our data where we iteratively refined our analytical framework and explanation based on engagement with literature and our empirical material.

Our analysis process consisted of five steps, summarized in Table 2. First, we read through field notes and documents to construct an initial case description. The initial description focused on 1) the context and motivation of the exploratory initiative, 2) the activities and actors of the development process, and 3) the project outcome. Alongside the case description, we also developed a representation of the IT architecture for each of the projects. Second, we conducted initial analyzes and discussions of the case descriptions and IT architectures. This led us to identify the nature of coupling between heavyweight and lightweight IT as a potential explanation. Third, we constructed an initial version of our analytical framework and reanalyzed the projects using the coupling lens. We compared and constrasted the cases and found that coupling did seem to be an important factor in the success of the projects. But, it was not sufficient to explain the differences in project outcomes. Fourth, we reexamined our empirical material, attempting to identify additional factors that could explain the differences in outcomes. In particular, we focused on differences in project context and the development process. In this process, we identified differences in the technology choices made during the development process as a potential factor: some of the projects required custom development whereas others relied on connecting existing services. We conceptualized this as the use of building blocks and modified our analytical framework, thus arriving at the final version presented in section 3. Fifth, we once again re-analyzed our cases to assess our new explanation and found that it was sufficient in explaining the differences in outcomes. At this step, we also assessed rival explanations to see whether they contained similar explanatory power. As examples, we assessed whether differences in how ML is used (i.e., augmentation vs. automation use-cases) or increased experience with ML could explain the differences in outcomes. In both cases these rival explanations were less supported by the empirical evidence.

Step	Activity	Outcome
1	Reading field notes and documents to develop case descriptions.	Initial case descriptions and IT architecture diagrams.
2	Initial case analysis and discussion	Identification of coupling between heavyweight and lightweight IT as lens.
3.	Construction of initial analytical framework and case re-analysis	Coupling offering explanatory value, but not sufficient on its own
4	Re-examination of empirical material	Identification of building blocks as lens. Modified and final framework.
5	Re-analysis of cases & assessment of explanatory power versus rival explanations	Plausible and coherent explanation of differences in outcomes: Lightweight ML success is enabled by loose coupling of heavyweight IT and lightweight IT, and the use of building blocks.

Table 2: Data analysis process

5 Cases

5.1 Case 1: Engineering Design Rework Reduction

Context

The first initiative took place in a manufacturing engineering department responsible for designing and producing manufacturing equipment. In the current experience-based process, an equipment design is developed by engineers, manufactured, and then subjected to extensive testing to ensure that it can produce products of acceptable quality. Initial designs, however, rarely pass testing, and thus several expensive iterations are required to reach an acceptable design. The existing process was struggling to meet its strategic output targets with the amount of rework identified as a main culprit. The exploratory initiative aimed at reducing rework was initiated as IT managers saw potential in using data to improve the engineering processes, while a vendor had offered to conduct a three month PoC in hopes of becoming a future innovation partner. As a result, a vendor-led initiative was started focused on using data to augment design decisions with the goal of reducing rework. Figure 2 provides an overview of the case by means of the analytical framework.

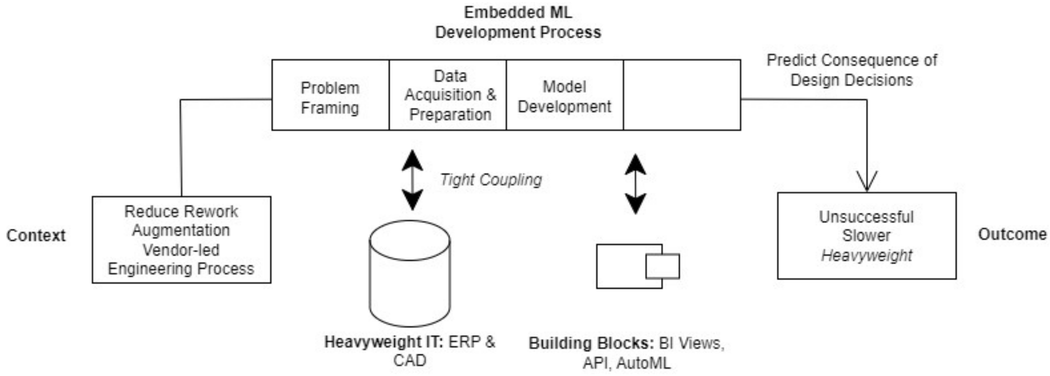


Figure 2. Analytical framework applied to the Engineering Design Rework Reduction case

Embedded ML Development Process

The ML development process was exploratory and focused initially on Problem Framing to identify a good use-case for the data. After several bouts of engaging process stakeholders and subject matter experts, two related use-cases had been prioritized for model development: 1) predicting the number of times a design will fail testing, and 2) predicting the probability of a particular test failure. Initial models were quickly developed using Automated ML and design master data which was readily available, but both models failed to deliver satisfying predictive performance. Concurrently, work was conducted to extract more detailed design data from the CAD system to be used in a second iteration of model development. After considerable delays in extracting CAD data, new models were developed, but the result was still non-satisfactory performance, as there was simply not enough signal in the data.

Coupling to Heayweight IT

While several of the data sources were accessible in a loosely coupled manner, the initiative as a whole was characterized by tight coupling to heavyweight IT due to the CAD data requirements. The design master data was available via. self-service BI views and could thus easily be extracted and updated as required by the participants of the initiative. The CAD data was available via. a proprietary API, which resulted in tight coupling. Access to the API required licenses that had to go through slow IT approval processes. Additionally, as the API was a generic vendor API, it was not designed to extract the data that was relevant for the company's use of the system. Developing the data extraction job was thus complex and required competences from the heavyweight IT team.

Use of Building Blocks

The development process made some use of building blocks in terms of data and model training. Data acquisition and preparation relied on the existing BI views and the CAD data API, but developing the data extraction job for the CAD data API remained a major custom development task. Model development utilized an Automated ML component, which automates the model training part of the process, but leaves the model specific data processing and model evaluation to the data scientist.

Outcome
The initiative was ultimately not successful in building a viable ML model and use-case, even though what was initially scheduled for a 12-week project was extended for another four weeks. It is illustrative of the typical slow and heavyweight nature of ML initiatives, where considerable time, experimentation, and expertise is required to build ML models and discover whether they are useful. While the data scientists did use Automated ML to speed up model development, the speed gained was dwarfed by the slowing down caused by tight coupling to the CAD system, which ended up delaying the whole project.

5.2 Case 2: Engineering Design Cost Prediction

Context
The second initiative took place in the same manufacturing engineering department, however, it had a considerably smaller scope than the first. Initiated by a single engineer, the initiative focused on predicting the costs of making a specific design decision. In the design process, designers have to choose between different equipment design concepts, a decision that is made partly based on an estimate of the cost of the equipment. The existing cost estimate used to choose the design concept was based on heuristics that, while simple, resulted in wildly inaccurate cost estimates. The initiating engineer speculated that ML could improve on the existing cost estimates and as a result set out to explore the potential of ML for accurately predicting design costs. Figure 3 provides an overview of the case by means of the analytical framework.

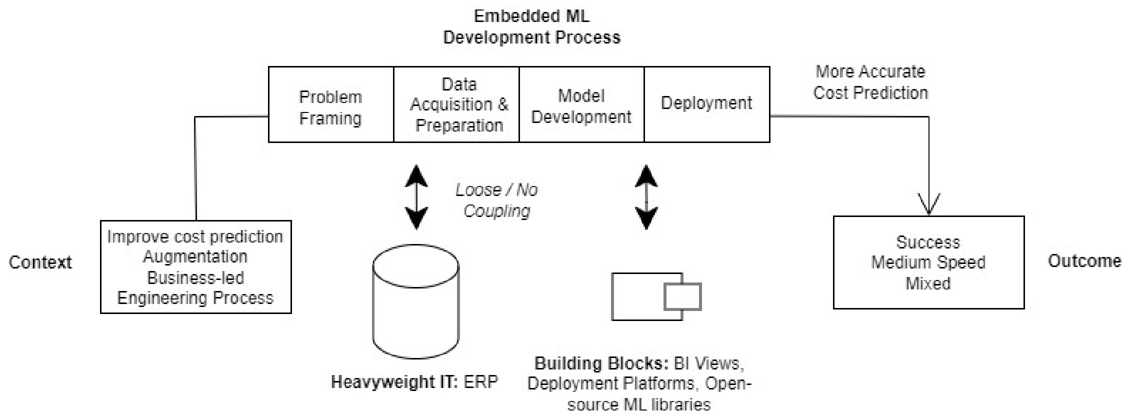


Figure 3. Analytical framework applied to the Engineering Design Cost Prediction case

Embedded ML Development Process
The initiative covered all stages of the ML process. Problem framing was carried out by the engineer in collaboration with colleagues. Afterwards, the engineer single-handedly completed data acquisition and model development on his company laptop using self-service BI tools to extract design master data and historical cost data from existing BI views. After some experimentation, the engineer managed to develop a model that outperformed the existing heuristic by more than 60% and reached out to an IT engineer and a data scientist (the first author) to get support for deployment. The engineer had developed a prototype web application, where users could enter design master data and get cost estimates. After a few architecture workshops, the model was deployed as an API via. a GUI, leveraging functionality in the data platform. Concurrently, the web application was refactored and deployed on an existing data application deployment platform offered by the data science CoE.

Coupling to Heayweight IT
Model development was loosely coupled from heavyweight IT, as the ability to manually extract data via. existing self-service BI tools meant that no support was required from heavyweight IT. Deployment was not coupled to heavyweight IT at all. While atypical, it was possible as the ML system relied on

manual data input rather than data from heavyweight systems. Similarly, the choice to use existing lightweight deployment platforms, meant that the system was deployed outside heavyweight IT.

Use of Building Blocks

The development process leveraged two major building blocks: 1) existing BI views, and 2) the deployment platforms. The presence of BI views that were already in use meant that model development could start with a relatively solid data foundation and thus reduced the efforts needed to extract, clean and preprocess data. Similarly, the existing data platform provided the ability to quickly deploy models as API's, and the lightweight data application hosting platform meant that the infrastructure required for deployment was already in place. Model development also made use of building blocks, but did so using generic lower-level components, including the algorithms, and metrics present in open-source ML libraries.

Outcome

The initiative resulted in the development and deployment of a ML system that was quantitatively evaluated to be considerably superior to the existing heuristic in use. The application was presented to process leadership with positive results and a decision was made for continuing implementation. It also illustrated how a relatively small team was capable of developing and deploying the ML system without significant friction from the heavyweight IT infrastructure. While deployment was lightweight and fast, the process as a whole was not. The custom model development drew on data science expertise and required several bouts of experimentation, resulting in medium speed as the overall outcome.

5.3 Case 3: Closed-loop Control of Machines

Context

The third initiative took place in a manufacturing R&D department responsible for running tests on manufacturing equipment and conducting process R&D. While the manufacturing process was already highly automated and efficient, significant productivity improvements were still expected as part of the company's Industry 4.0 strategy. A roadmap for the usage of data and analytics to further improve the process had been developed. An envisioned future use-case of ML was to optimize the quality of the production process continuously by automatically adjusting machine settings. Early research had shown promising results in using ML to compensate for variations in the material used, thus leading to improved product quality, albeit at a small scale. However, this research took place without the involvement of IT and thus the solution and IT architecture developed was not scalable. With IT expecting future demands for solutions leveraging ML-enabled control, a collaboration with IT was established to explore scalable solutions that were interoperable with the existing IT infrastructure. Figure 4 provides an overview of the case by means of the analytical framework.

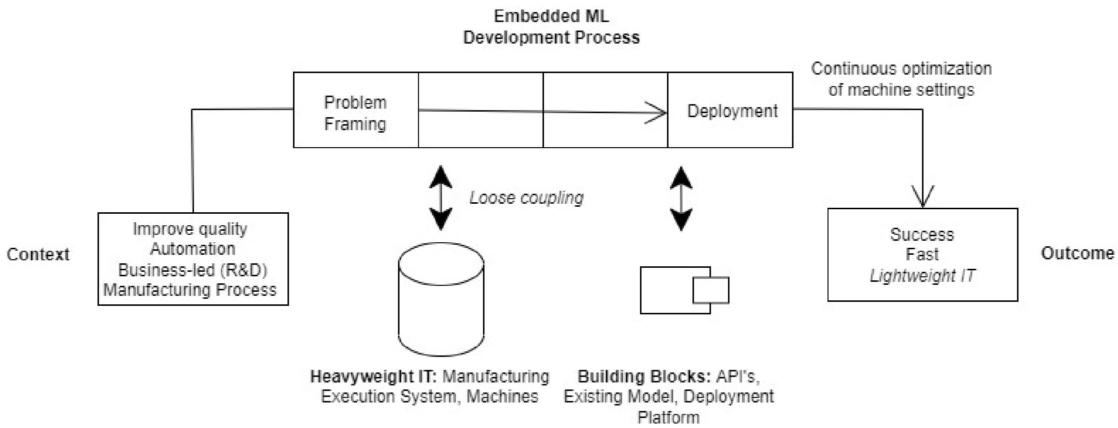


Figure 4. Analytical framework applied to the Closed-loop Control of Machines case

Embedded ML Development Process

The ML development process was focused on deployment rather than model development and made use of the model developed in previous R&D. Using the existing model meant that the initiative could skip data acquisition and model development as illustrated via. the arrow on Figure 4. Deployment used the data platform and standard cloud services configured for the use-case. A data pipeline was developed using SaaS that extracted the necessary data for predictions from the manufacturing execution system (MES) and moved it to the data platform. The model was deployed as a stream processing job on the data platform that received data from the data pipeline, made predictions, and used the predictions to derive new machine settings. The machine settings were then updated automatically via. calls to the MES, that had the ability to update machine settings.

Coupling to Heayweight IT

The initiative was loosely coupled to the heavyweight systems, both in terms of accessing data and functionality. The MES and the ML system were integrated asynchronously, thus ensuring that the MES system was not dependent on the ML system. Accessing data from the MES was possible via. an inhouse API, enabling self-service construction of a data pipeline. The model itself was deployed in a self-service fashion using the data platform and thus had no direct integration with heavyweight IT. Similarly, the MES exposed a machine setting adjustment capability as an API, thus allowing the ML system to control machines through API calls.

Use of Building Blocks

The development process made extensive use of building blocks in terms of both the data, models, and system components. While data pipelines were built as part of the project, they relied on the presence of an existing API that provided access to the data generated by the production process. Similarly, model development relied on an existing model training script that was repurposed on a new dataset to generate the model. Lastly, deployment relied on both the data processing infrastructure and job scheduling provided by the data platform to deploy the model, and existing API's to allow the ML system to interact with the production machines.

Outcome

The initiative resulted in the successful PoC of an IT architecture for closed-loop control of the machines. The development and implementation of the PoC was conducted primarily by two people in the space of a two-week development sprint. This was made possible by using components in the form of the existing model, the data platform, and standard SaaS components. Additionally, the initiative had very few external dependencies thanks to the previous work of turning the MES into a platform by API-enabling it. Although the initiative did not only use standard technologies, as evidenced by the existing custom developed model, it had many of the characteristics of lightweight IT.

5.4 Case 4: Data Quality Anomaly Detection

Context

The fourth initiative took place in IT in the data platform team, which was responsible for operating the data platform and providing access to data for analytical initiatives. Experience with several analytics initiatives using the data platform quickly led to the realization that much of the data in the platform had significant quality problems. The result of these data quality issues was in the best case that significant time had to be spent on data cleaning in analytical initiatives. In the worst case, large parts of the data collected was rendered invalid. Sparked by discussions in the data platform team on the need for a process to monitor and secure data quality, a small scale initiative was started to explore the opportunities of available technology to provide support for a data quality monitoring process. The vision for the initiative was to improve data quality by proactively monitoring data pipelines for issues, using an automated and scalable monitoring solution built on the concept of anomaly detection, followed by detailed manual investigations in case of detected anomalies. Figure 5 provides an overview of the case by means of the analytical framework.

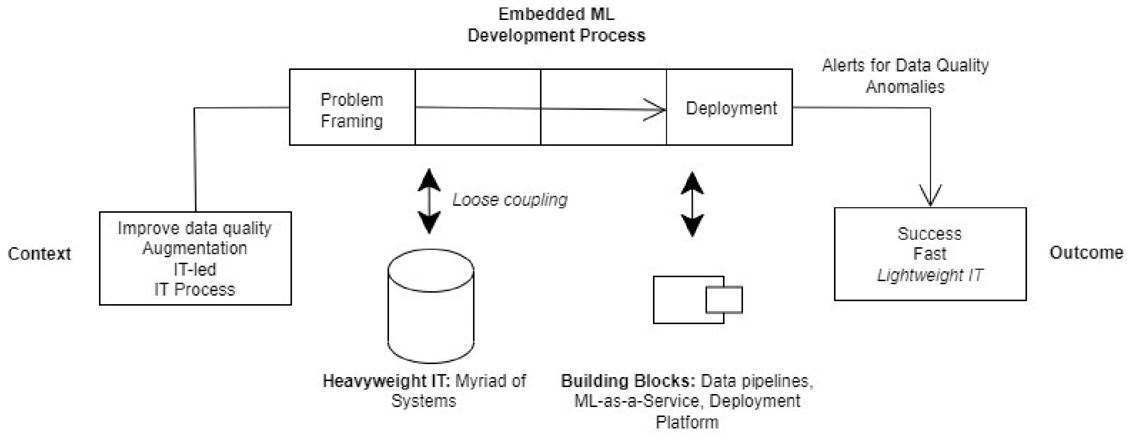


Figure 5. Analytical framework applied to the Data Quality Anomaly Detection case

Embedded ML Development Process

The ML development process focused on deployment and was carried out by a single developer (the first author) with sparring from the data platform team. The development process differed from the other cases, as it made use of off-the-shelf available ML-as-a-Service (MLaaS). Making use of MLaaS allowed the development process to skip the training data acquisition phase, as well as the following model development phase as illustrated on Figure 5 via. the arrow. Deployment consisted of provisioning, configuring, and integrating the MLaaS with the data platform through batch jobs scheduled on the data platform. Deployment started out with a single initial data source and was later expanded. For sending notifications when anomalies were detected, a SaaS alerting solution that was already in use in the organization and compatible with the MLaaS software was used.

Coupling to Heayweight IT

Development of the system took place without direct interaction with the heavyweight IT infrastructure, which was made possible by leveraging the data platform. The necessary data was available in the data platform as the result of previous work to construct data pipelines responsible for moving data from the heavyweight systems to the data platform in a batch or streaming process, thus presenting a case of loose coupling.

Use of Building Blocks

The development process relied extensively on existing components with custom development limited to a simple data aggregation job and component integration. Data was available using the existing data pipelines that were deployed using the data processing infrastructure in the data platform. The ML component consisted of a ML-as-a-Service, thus abstracting away both model development and deployment. System Development was similarly reduced mainly to the configuration of a standard UI provided in the MLaaS.

Outcome

The initiative resulted in the development and deployment of a successful PoC, that was quickly scaled to cover all data sources in the data platform. During its operation, it managed to identify and provide alerts for several data quality incidents that were resulting in data losses and thus allowed the data platform team to take remedying actions faster. The solution was developed in a short timeframe by a single developer through loosely connecting existing components in a non-invasive manner, thus bearing many of the characteristics of lightweight IT.

Summing up

Our four cases dealt with different challenges but all of them relate to the same exploratory process. As shown in Table 3, the outcomes vary, from slow to fast, and we have shown in the previous sections how this can be explained by combinations of coupling, the use of building blocks and the predominance of lightweight versus heavyweight IT.

Cases	Coupling	Use of Building Blocks	Light vs. Heavy	Speed
Engineering Design Rework Reduction	*Tight Coupling*	*Partly*	Heavyweight	Slow
Engineering Design Cost Prediction	*Loose/No Coupling*	*Partly*	Mediumweight	Medium
Closed-loop Control of Machines	*Loose Coupling*	*Extensively*	Lightweight	Fast
Data Quality Anomaly Detection	*Loose Coupling*	*Extensively*	Lightweight	Fast

Table 3. Enabling factors of Lightweight ML and implications for speed.

The cases demonstrated that (i) it is possible for ML to be lightweight, and (ii) the speed of lightweight ML projects was higher. This was illustrated by the *Closed-loop Control of Machines* and *Data Quality Anomaly Detection* cases. Both cases were focused on improving processes and carried out in organizational units responsible for processes. At the same time, they were able to develop and test PoC's of ML systems with few resources and in short time frames. The same was to a lesser extent true of the *Engineering Design Cost Prediction* case, which was primarily developed by a single engineer in the business organization. The speed was, however, lower than in the two preceding cases. On the other hand, the *Engineering Design Rework Reduction* case, was of larger scope. It involved both vendors, IT representatives, and process stakeholders in an intensive process that ended up taking 16 weeks, without major results to show for it.

6 Discussion

In this section we return to the research question, *how can lightweight IT contribute to explorative BPM for embedded machine learning?* ML, when embedded in business processes, has the potential to significantly improve business process performance by augmenting or automating decisions and enabling process redesign (Davenport, 2018; Enholm et al., 2021). Forward looking companies thus need to be able to explore the affordances of embedded ML as part of their explorative BPM activities. While lightweight IT has been proposed as a solution for exploring process innovation affordances of digital technologies in general (Schmiedel & vom Brocke, 2015; Bygstad & Øvrelid, 2020), the literature is unclear as to whether this is applicable to ML, traditionally considered a heavyweight technology.

Our findings suggests that for embedded ML projects to succeed as lightweight IT, requires the co-presence of two enabling factors, i.e., loose coupling between heavyweight and lightweight IT and extensive use of buildings blocks in the development process. We first discuss lose coupling; then the use of building blocks, before we conclude how we contribute to the research on exploratory BPM.

6.1 Loose Coupling

Our case analysis shows how coupling, a property of the IT infrastructure, impacts process innovation latency. Building on previous research suggesting the use of lighter engineering for process innovation with digital technologies (Baiyere et al., 2020; Bygstad & Øvrelid, 2020), we find that tight coupling between lightweight and heavyweight IT reduces speed in exploratory innovation.

In the one case where coupling between heavyweight and lightweight IT was tight (*Engineering Design Rework Reduction*), the need to interact extensively with heavyweight IT processes and developers related to data extraction from the CAD system ended up acting as a bottleneck. This bottleneck significantly slowed down the whole project down and rendered the use of Automated ML tools to speed up model development less impactful. The impact of coupling on speed in this case was particularly clear, as the first modelling iteration, which relied only on data sources that could be accessed independently of heavyweight IT resources, was much faster.

This is in line with the findings of Bygstad & Øvrelid (2020), who identified loosely coupled interaction of lightweight and heavyweight IT as an enabler of successful process innovation. Tight coupling reduces the speed of innovation to that of heavyweight IT, which is slow due to its systematicism and focus on rigour (Bygstad, 2017). Additionally, it increases the scope and complexity of the innovation initiative by requiring the coordination of two widely different knowledge regimes. Lightweight and heavyweight IT systems need to interact as the heavyweight systems contain data and functionality necessary for most process innovations. Thus, the usage of lightweight technologies by themselves is necessary but not sufficient for speed, unless accompanied by either no or loose coupling to heavyweight IT. This has implications for how the role of IT infrastructure in BPM is conceptualized. Supporting the claim of a need for more flexibility in IT infrastructure in digital age BPM (Baiyere et al., 2020), our findings suggests that IT infrastructure needs to take on dual roles. On the one hand, it needs to support efficient process execution as has been its traditional role. But it also needs to act as an innovation platform by exposing data and capabilities that can be leveraged in decoupled innovation activities. In our cases this was achieved through boundary resources that provided loosely coupled access to the heavyweight IT infrastructure, but also by means of new IT infrastructure, such as the data and deployment platforms.

6.2 Use of Building Blocks

Our analysis further suggests that the presence of loose coupling is not enough. Leveraging the concept of building blocks and conceptualizing innovation as recombination allowed us to conduct a more granular investigation of the development process (Henfridsson et al., 2018) and to go beyond the distinction between custom and packaged software. We found that extensive use of building blocks in the development process is required as well, as the data, model, and system component development tasks can otherwise become a considerable development effort. This was partly the case in the *Engineering Design Cost Prediction* case. The project relied on building blocks at the data level in the form of existing BI views, and at system level in the form of the deployments platforms used to deploy the model and application. However, model development relied only on low level building blocks and thus required custom development and experimentation to arrive at an acceptable model. In the *Closed-loop Control of Machines* and *Data Quality Anomaly Detection* cases the use of building blocks for the model component allowed the projects to skip the model development step, thus contributing significantly to speeding up the process.

Examining the building blocks present in our cases in more depth, it becomes clear that they differ in nature and cluster into three categories: 1) boundary resources, 2) developer platforms, and 3) solution components. The boundary resources (Ghazawneh & Henfridsson, 2013), consisting of API's and BI views, were essentially part of heavyweight IT and were what allowed the projects to interact in a loosely-coupled fashion with heavyweight IT. We thus find support for internal boundary resources playing an important role in enabling innovation (Bygstad & Øvrelid, 2020). The developer platforms, namely a cloud data platform and an application hosting platform, provided on-demand access to infrastructural resources, such as storage and compute for deployment and development. This both prevented the innovation team from managing complex infrastructure and prevented tight coupling to the IT infrastructure teams. This finding supports existing research demonstrating the importance of mature data platforms for obtaining value from ML (Reis et al., 2020) and increasing development speed (Anand et al., 2016), and hints at the mechanisms behind these effects. The solution components (consisting of data pipelines, models, and systems) functioned to reduce custom development and skip steps in the ML process. The origin of these components varied. Some components were the result of previous internal development, other components were offered by cloud vendors, and finally, some of the lower-level components were open-source software. On the one hand, this underscores the complexity of ML and that there is indeed much more to ML than the model (Ratner et al., 2019; Lavin et al., 2022; Sculley et al., 2015), as emphasized in the engineering focused stream on ML systems. On the other hand, the presence of building blocks at the model and system level suggests that the increasing technological maturity of ML is indeed making ML a lighter technology, as emphasized in the data democratization and Automated ML research streams (Hutter et al., 2019; Lefebvre et al., 2021).

Thus, we argue that it is the combination of a loose coupling to heavyweight IT and the extensive use of building blocks that allowed for successful lightweight ML and fast exploration of process innovation affordances. It is certainly possible to be successful without the use of building blocks, but the initiatives are bound to be slower due to the increased scope and complexity associated with the development process. On the other hand, it is important to emphasize that the successful discovery of process innovation affordances is not guaranteed by using lightweight ML. Success is a result of the right combination of problem, data, model, and system, something which lightweight ML does not contribute directly to. It does, however, contribute to increasing the speed at which different combinations can be explored.

6.3 Contribution to the Research on Explorative BPM

First, we contribute by providing an empirical investigation of exploratory digital process innovation. Existing research on realizing explorative BPM has focused on support for realizing the opportunity identification stage (Gross et al., 2021; Grisold et al., 2022) and to a lesser extent the implementation phase (Baier et al., 2022), but has so far not addressed the opportunity assessment stage in depth. It is in the opportunity assessment stage that the process innovation affordances of digital technologies are explored in more depth by turning the identified opportunities into proof-of-concepts. This exploration process is particularly critical for digital technologies such as ML, IoT, and big data. Recent research has suggested that these technologies are weakly structured and thus require significant organizational effort to discover their affordances (Eley & Lyytinen, 2022). Our study contributes by providing initial empirical insights into the exploration process for one of these weakly structured technologies.

Second, we contribute by demonstrating how the use of loose coupling and building blocks impacts process innovation latency. Innovation occurs not by inventing something completely new, but rather by recombination of existing technologies and knowledge (Arthur, 2009). Using loose coupling and higher-level building blocks allows for significantly higher speed in the innovation process, achieved by skipping technology development steps, and allows the focus to remain on the process affordances offered by the technology. This is compatible with the existing view of lighter engineering as best practice in process innovation, achieved by configuring standard and flexible technology (Schmiedel & vom Brocke, 2015; Baiyere et al., 2020; Bygstad & Øvrelid, 2020).

We add further nuance to this discussion, by demonstrating that the use of building blocks (rather than standard technologies) also allows for fast innovation. A straightforward implication is that achieving high speed in explorative BPM is dependent on the maturity level of the technology in question, as building blocks are more likely to be readily available for more mature technologies. There is thus a trade-off between innovativeness and speed, and explorative BPM research needs to offer solutions that are able to handle exploration of both immature emergent technologies, as well as the use of more mature technologies in new ways. Zooming in on embedded ML, this means that there is a trade-off between predictive performance and innovation speed, as generic models, embedded in standard solutions, will most likely perform worse than custom developments (May et al., 2020).

To summarize, we show that lightweight IT can contribute to explorative BPM for embedded ML by significantly speeding up the opportunity assessment and technical implementation phases of the innovation process, thus reducing process innovation latency.

6.4 Implications for Practice

Our findings have practical implications for managers looking to explore the process affordances of ML. First, managers need to be aware of the distinction between lightweight and heavyweight ML and adjust their expectations and organization of exploratory activities accordingly. *Lightweight ML* initiatives can be run outside the traditional IT department using fast iterations, e.g., by collaborating with Data Science CoE's, vendors, or data scientists in business units. *Heavyweight ML* initiatives on the other hand are likely to require teaming up with the IT department and adopting a longer-term R&D perspective. All relevant initiatives will not be *lightweight ML,* as ML is still a relatively immature technology with high-level building blocks only existing for specific use-cases, or as in the *Closed-loop Control of Machines* case, as the result of previous data science R&D. Process managers should therefore consider teaming up with internal data science competences to ensure their R&D activities have clear process relevance.

Second, managers looking to achieve fast exploration of the process innovation affordances of ML need to work together with IT and invest in building a loosely coupled IT infrastructure. The ability to deploy ML systems loosely coupled from heavyweight systems requires both self-service data access and self-service ML deployment infrastructure. Self-service data access can be realized by means of API-enabling heavyweight systems or by constructing data pipelines to transport heavyweight data to a data lake or data warehouse. However, as illustrated by the *Engineering Design Loops Reduction* case, vendor provided proprietary API's might not cut it, as they often do not allow or true self-service data access. Self-service deployment platforms can be built internally or as is commonly the case, acquired from cloud vendors (e.g., Databricks, Microsoft Azure, Google Cloud Platform, or Amazon AWS).

6.5 Limitations & Future Research

Our study is not without its limitations. First, the projects addressed were all concerned with opportunity-driven exploratory innovation of business processes, but they did not fit perfectly with the emerging conceptualization of explorative BPM. In addition to its characterization as opportunity-driven, explorative BPM is often associated with radical process innovation focused on novel value propositions (Rosemann, 2014; Grisold et al., 2022). The projects we examined, on the other hand consisted of internally-focused innovations aimed at improving operations and most of the cases leaned towards incremental rather than radical innovation. Nonetheless, we argue the cases deviate considerably from the traditional and reactive exploitative BPM and are thus closer in spirit to explorative BPM. We suggest future research further clarify the concept of explorative BPM, making it clear where opportunity & technology-driven but internally-focused innovation fits, as our empirical engagements show that it is a phenomenon receiving significant attention in practice.

We also acknowledge that our theorizing is based on a selection of projects from our ongoing action-oriented work in a single case organization thus limiting its generalizability. Selecting among projects that we were engaged in facilitated an indepth and detailed understanding of the phenomena under study, including both technical and organizational aspects, and allowed us to go beneath the surface and provide an insider perspective on the use of lightweight ML for process innovation. It does, however, come at the cost of the representativeness of the projects selected. Further research should test whether the co-presence of loose coupling and extensive use of building blocks are necessary for succesful lightweight ML projects in different organizations.

7 Conclusion

In this study, we set out to investigate how lightweight IT can contribute to explorative BPM for embedded ML. We presented and analyzed four cases that attempted to treat ML as lightweight IT with different degrees of success. To assist us in the analysis we relied on an analytical framework that leveraged the concepts of (i) coupling (Weick, 1976), specifically between lightweight and heavyweight IT (Bygstad, 2017), and (ii) building blocks, i.e., the extent of use of existing socio-technical components in the development of solutions.

Our analysis of four cases of embedded ML in a large Danish manufacturer demonstrated that a *lightweight* approach can considerably speed up assessment and technology implementation of ML, thus contributing to the fast exploration of process innovation affordances. The lightweight approach is, however, not always feasible, as it requires the presence of two enabling factors. First, it requires loose coupling between the exploratory initative and the execution-focused organization, which in turn requires a loosely coupled IT infrastructure. Second, it requires that the exploratory initiative makes use of building blocks, thus reducing the need for extensive custom development and engineering. With a lightweight approach the focus is moved from the technology towards the process and its' improvement, and while the existing technological maturity level of ML does not always allow for a lightweight approach, the rapid pace of development suggests that it will be increasingly possible in a near future.

Declarations

Funding

The first author received funding from Manufacturing Academy of Denmark under the MADE FAST research and innovation programme (funding nr.: 9090-00002B) to support the research that led to this article.

Competing Interests

The authors have no competing interests to declare that are relevant to the content of this article.

Data Availability

The data generated during and/or analysed during the current study cannot be made publicly available due to confidentiality agreements made with the participating case organization to protect its commercial interests. Further information and details regarding the data can be obtained from the corresponding author upon request.

References

Anand, A., Sharma, R., & Coltman, T. (2016). Four Steps to Realizing Business Value from Digital Data Streams. MIS Quarterly Executive, 15(4). https://aisel.aisnet.org/misqe/vol15/iss4/3

Arthur, W. B. (2009). The Nature of Technology: What It Is and How It Evolves. Simon and Schuster.

Baier, M.-S., Lockl, J., Röglinger, M., & Weidlich, R. (2022). Success factors of process digitalization projects – insights from an exploratory study. Business Process Management Journal, 28(2), 325–347. https://doi.org/10.1108/BPMJ-07-2021-0484

Baiyere, A., Salmela, H., & Tapanainen, T. (2020). Digital transformation and the new logics of business process management. European Journal of Information Systems, 29(3), 238–259. https://doi.org/10.1080/0960085X.2020.1718007

Benbya, H., Pachidi, S., & Jarvenpaa, S. L. (2021). Special Issue Editorial: Artificial Intelligence in Organizations: Implications for Information Systems Research. Journal of the Association for Information Systems, 22(2), 281–303. https://doi.org/10.17705/1jais.00662

Benner, M. J., & Tushman, M. L. (2003). Exploitation, Exploration, and Process Management: The Productivity Dilemma Revisited. The Academy of Management Review, 28(2), 238–256. https://doi.org/10.2307/30040711

Braa, K., & Vidgen, R. (1999). Interpretation, intervention, and reduction in the organizational laboratory: A framework for in-context information system research. Accounting, Management and Information Technologies, 9(1), 25–47. https://doi.org/10.1016/S0959-8022(98)00018-6

Breck, E., Polyzotis, N., Roy, S., Whang, S. E., & Zinkevich, M. (2019). Data Validation for Machine Learning. Proceedings of Machine Learning and Systems 1, 14.

Bygstad, B. (2017). Generative Innovation: A Comparison of Lightweight and Heavyweight IT. Journal of Information Technology, 32(2), 180–193. https://doi.org/10.1057/jit.2016.15

Bygstad, B., & Øvrelid, E. (2020). Architectural alignment of process innovation and digital infrastructure in a high-tech hospital. European Journal of Information Systems, 29(3), 220–237. https://doi.org/10.1080/0960085X.2020.1728201

Davenport, T. H. (2018). From analytics to artificial intelligence. Journal of Business Analytics, 1(2), 73–80. https://doi.org/10.1080/2573234X.2018.1543535

Davenport, T. H., Harris, J. G., & Morison, R. (2010). Analytics at work: Smarter decisions, better results. Harvard Business Press.

Davenport, T. H., & Short, J. E. (1990). The new industrial engineering: Information technology and business process redesign. Sloan Management Review, 31(4), 11–27.

Davenport, T., & Malone, K. (2021). Deployment as a Critical Business Data Science Discipline. Harvard Data Science Review.

Dubois, A., & Gadde, L.-E. (2002). Systematic combining: An abductive approach to case research. Journal of Business Research, 55(7), 553–560. https://doi.org/10.1016/S0148-2963(00)00195-8

Dumas, M., La Rosa, M., Mendling, J., & Reijers, H. A. (2013). Fundamentals of business process management (Vol. 1). Springer.

Eley, T., & Lyytinen, K. (2022). Industry 4.0 Implementation: Novel Issues and Directions. Hawaii International Conference on System Sciences. https://doi.org/10.24251/HICSS.2022.622

Enholm, I. M., Papagiannidis, E., Mikalef, P., & Krogstie, J. (2021). Artificial Intelligence and Business Value: A Literature Review. Information Systems Frontiers. https://doi.org/10.1007/s10796-021-10186-w

Ghazawneh, A., & Henfridsson, O. (2013). Balancing platform control and external contribution in third-party development: The boundary resources model. Information Systems Journal, 23(2), 173–192. https://doi.org/10.1111/j.1365-2575.2012.00406.x

Grisold, T., Gross, S., Röglinger, M., Stelzl, K., & vom Brocke, J. (2019). Exploring Explorative BPM - Setting the Ground for Future Research. In T. Hildebrandt, B. F. van Dongen, M. Röglinger, & J. Mendling (Eds.), Business Process Management (Vol. 11675, pp. 23–31). Springer International Publishing. https://doi.org/10.1007/978-3-030-26619-6_4

Grisold, T., Groß, S., Stelzl, K., vom Brocke, J., Mendling, J., Röglinger, M., & Rosemann, M. (2022). The Five Diamond Method for Explorative Business Process Management. Business & Information Systems Engineering, 64(2), 149–166. https://doi.org/10.1007/s12599-021-00703-1

Grønsund, T., & Aanestad, M. (2020). Augmenting the algorithm: Emerging human-in-the-loop work configurations. The Journal of Strategic Information Systems, 29(2), 101614. https://doi.org/10.1016/j.jsis.2020.101614

Gross, S., Stelzl, K., Grisold, T., Mendling, J., Röglinger, M., & vom, B. J. (2021). The Business Process Design Space for exploring process redesign alternatives. Business Process Management Journal, 27(8), 25–56. https://doi.org/10.1108/BPMJ-03-2020-0116

Henfridsson, O., Nandhakumar, J., Scarbrough, H., & Panourgias, N. (2018). Recombination in the open-ended value landscape of digital innovation. Information and Organization, 28(2), 89–100. https://doi.org/10.1016/j.infoandorg.2018.03.001

Hutter, F., Kotthoff, L., & Vanschoren, J. (Eds.). (2019). Automated Machine Learning: Methods, Systems, Challenges. Springer International Publishing. https://doi.org/10.1007/978-3-030-05318-5

Kerpedzhiev, G. D., König, U. M., Röglinger, M., & Rosemann, M. (2021). An Exploration into Future Business Process Management Capabilities in View of Digitalization. Business & Information Systems Engineering, 63(2), 83–96. https://doi.org/10.1007/s12599-020-00637-0

Kunz, P. C., Jussupow, E., Spohrer, K., & Heinzl, A. (2022). How the Application of Machine Learning Systems Changes Business Processes: A Multiple Case Study. ECIS 2022 Research-in-Progress Papers, 11.

Lang, K. R., Misic, V. B., & Zhao, L. J. (2015). Special section on business process analytics. Information Systems Frontiers, 17(6), 1191–1194. https://doi.org/10.1007/s10796-015-9610-1

Lavin, A., Gilligan-Lee, C. M., Visnjic, A., Ganju, S., Newman, D., Ganguly, S., Lange, D., Baydin, A. G., Sharma, A., Gibson, A., Zheng, S., Xing, E. P., Mattmann, C., Parr, J., & Gal, Y. (2022). Technology readiness levels for machine learning systems. Nature Communications, 13(1), Article 1. https://doi.org/10.1038/s41467-022-33128-9

Lefebvre, H., Legner, C., & Fadler, M. (2021, September 28). Data democratization: Toward a deeper understanding. ICIS 2021 Proceedings. International Conference of Information Systems.

Martínez-Plumed, F., Contreras-Ochando, L., Ferri, C., Hernández-Orallo, J., Kull, M., Lachiche, N., Ramírez-Quintana, M. J., & Flach, P. (2019). CRISP-DM Twenty Years Later: From Data Mining Processes to Data Science Trajectories. IEEE Transactions on Knowledge and Data Engineering, 33(8), 3048–3061. https://doi.org/10.1109/TKDE.2019.2962680

May, A., Sagodi, A., Dremel, C., & van Giffen, B. (2020, December 13). Realizing Digital Innovation from Artificial Intelligence. ICIS 2020 Proceedings. Forty-First International Conference on Information Systems, Forty-First International Conference on Information Systems.

Mendling, J., Pentland, B. T., & Recker, J. (2020). Building a complementary agenda for business process management and digital innovation. European Journal of Information Systems, 29(3), 208–219. https://doi.org/10.1080/0960085X.2020.1755207

Microsoft. (2023). What is the Team Data Science Process? https://learn.microsoft.com/en-us/azure/architecture/data-science-process/overview

Paleyes, A., Urma, R.-G., & Lawrence, N. D. (2022). Challenges in Deploying Machine Learning: A Survey of Case Studies. ACM Computing Surveys, 3533378. https://doi.org/10.1145/3533378

Phillips-Wren, G., Daly, M., & Burstein, F. (2021). Reconciling business intelligence, analytics and decision support systems: More data, deeper insight. Decision Support Systems, 146, 113560. https://doi.org/10.1016/j.dss.2021.113560

Poll, R., Polyvyanyy, A., Rosemann, M., Röglinger, M., & Rupprecht, L. (2018). Process Forecasting: Towards Proactive Business Process Management. In M. Weske, M. Montali, I. Weber, & J. vom Brocke (Eds.), Business Process Management (Vol. 11080, pp. 496–512). Springer International Publishing. https://doi.org/10.1007/978-3-319-98648-7_29

Ratner, A., Alistarh, D., Alonso, G., Andersen, D. G., Bailis, P., Bird, S., Carlini, N., Catanzaro, B., Chayes, J., Chung, E., Dally, B., Dean, J., Dhillon, I. S., Dimakis, A., Dubey, P., Elkan, C., Fursin, G., Ganger, G. R., Getoor, L., … Talwalkar, A. (2019). MLSys: The New Frontier of Machine Learning Systems (arXiv:1904.03257). arXiv. http://arxiv.org/abs/1904.03257

Reis, C., Ruivo, P., Oliveira, T., & Faroleiro, P. (2020). Assessing the drivers of machine learning business value. Journal of Business Research, 117, 232–243. https://doi.org/10.1016/j.jbusres.2020.05.053

Robey, D., & Taylor, W. T. F. (2018). Engaged Participant Observation: An Integrative Approach to Qualitative Field Research for Practitioner-Scholars. Engaged Management ReView, 2(1). https://doi.org/10.28953/2375-8643.1028

Rosemann, M. (2014). Proposals for Future BPM Research Directions. In C. Ouyang & J.-Y. Jung (Eds.), Asia Pacific Business Process Management (Vol. 181, pp. 1–15). Springer International Publishing. https://doi.org/10.1007/978-3-319-08222-6_1

Sætre, A. S., & Van De Ven, A. (2021). Generating Theory by Abduction. Academy of Management Review, 46(4), 684–701. https://doi.org/10.5465/amr.2019.0233

Schmiedel, T., & vom Brocke, J. (2015). Business Process Management: Potentials and Challenges of Driving Innovation. In J. vom Brocke & T. Schmiedel (Eds.), BPM - Driving Innovation in a Digital World (pp. 3–15). Springer International Publishing. https://doi.org/10.1007/978-3-319-14430-6_1

Sculley, D., Holt, G., Golovin, D., Davydov, E., Phillips, T., Ebner, D., Chaudhary, V., Young, M., Crespo, J.-F., & Dennison, D. (2015). Hidden Technical Debt in Machine Learning Systems. Advances in Neural Information Processing Systems, 28. https://proceedings.neurips.cc/paper/2015/hash/86df7dcfd896fcaf2674f757a2463eba-Abstract.html

Sedera, D., Lokuge, S., Grover, V., Sarker, S., & Sarker, S. (2016). Innovating with enterprise systems and digital platforms: A contingent resource-based theory view. Information & Management, 53(3), 366–379. https://doi.org/10.1016/j.im.2016.01.001

Sein, Henfridsson, Purao, Rossi, & Lindgren. (2011). Action Design Research. MIS Quarterly, 35(1), 37. https://doi.org/10.2307/23043488

Simon, H. (1996). The Sciences of the Artificial (3rd ed.). MIT Press.

Teinemaa, I., Dumas, M., Rosa, M. L., & Maggi, F. M. (2019). Outcome-Oriented Predictive Process Monitoring: Review and Benchmark. ACM Transactions on Knowledge Discovery from Data, 13(2), 1–57. https://doi.org/10.1145/3301300

Uzunalioglu, H., Cao, J., Phadke, C., Lehmann, G., Akyamac, A., He, R., Lee, J., & Able, M. (2019). Augmented Data Science: Towards Industrialization and Democratization of Data Science (arXiv:1909.05682). arXiv. http://arxiv.org/abs/1909.05682

Van de Ven, A. H. (2007). Engaged scholarship: A guide for organizational and social research. Oxford University Press on Demand.

Van Der Aalst, W. (2012). Process Mining: Overview and Opportunities. ACM Transactions on Management Information Systems, 3(2), 1–17. https://doi.org/10.1145/2229156.2229157

Weick, K. E. (1976). Educational Organizations as Loosely Coupled Systems. Administrative Science Quarterly, 21(1), 1–19. https://doi.org/10.2307/2391875

Wirth, R., & Hipp, J. (2000). CRISP-DM: Towards a Standard Process Model for Data Mining. Proceedings of the 4th International Conference on the Practical Applications of Knowledge Discovery and Data Mining, 11.

Paper 4: Developing Analytics Demonstrators for Process Innovation: An Infrastructural Perspective

Bojer, Casper Solheim, and Møller, Charles. (2023b). *Developing Analytics Demonstrators for Process Innovation: An Infrastructural Perspective* [Manuscript submitted for publication].

The paper has been submitted to *European Journal of Information Systems* and is currently under review.

Developing Analytics Demonstrators for Process Innovation: An Infrastructural Perspective

Casper Solheim Bojer[a]* and Charles Møller[b]

[a]*Department of Materials & Production, Aalborg University, Aalborg, Denmark;* [b] *Department of Mechanical and Production Engineering, Aarhus University, Aarhus, Denmark*

Email of corresponding author: casperbojer@gmail.com

Developing Analytics Demonstrators for Process Innovation: An Infrastructural Perspective

The potential of analytics to innovate work processes has led organizations to invest significantly in this technology. To evaluate the value of analytics for process innovation, the development of analytics system prototypes, or analytics demonstrators, becomes essential. Despite the availability of several analytics development methodologies to guide demonstrator development, they lack prescriptive guidance on socio-technical aspects and the deployment phase where models are integrated into the digital infrastructure. This research aims to address this challenge in adopting analytics for process innovation through a collaboration with a large global manufacturer and retailer based in Denmark. We conducted a three-year Action Design Research project, collaborating closely with the organization in the development of their operations data platform and multiple analytics demonstrators. Based on this experience, we propose a development approach and six design principles, adopting a digital infrastructure perspective on analytics. The primary contributions of this study are twofold. First, we provide an approach and design principles for developing analytics demonstrators targeted at process innovation within large organizations with complex digital infrastructures. Second, we demonstrate the significance of adopting an infrastructural perspective in analytics research, emphasizing the importance of digital infrastructure concepts in process innovation with analytics.

Keywords: analytics, machine learning, process innovation, digital infrastructure, design principles, action design research

Introduction

Analytical technologies such as Machine Learning hold immense potential to improve work processes across industries (Tarafdar et al., 2019). Analytics can be embedded in work processes (Davenport, 2018) to augment or automate work (Benbya et al., 2021; Shollo et al., 2022), resulting in improvements to quality and productivity (Enholm et al., 2021; Tarafdar et al., 2019). Realizing these benefits requires not just the implementation of analytics systems, but rather coordinated socio-technical change

(Dremel et al., 2017; Dremel et al., 2020; Tim et al., 2020), including redesign of the work process (Kunz et al., 2020; Enholm et al., 2021). Owing to this potential, companies are increasingly investing in exploring how analytics can improve their processes. As analytics is a weakly structured (Eley IV & Lyytinen, 2022) and data-dependent technology, assessing the value of analytics for process innovation requires development of analytics system prototypes using organizational data, or what we term *analytics demonstrators.*

Several analytics development methodologies exist to support organizations in developing analytical models and systems (e.g., CRISP-DM (Wirth & Hipp, 2000), TDSP (Microsoft, 2023), DST (Martínez-Plumed et al. 2020), BAM (Hindle & Vidgen, 2020)). The newer methodologies attempt to address several shortcomings of the widely adopted CRISP-DM methodology by 1) supporting exploration in addition to well-defined business problems (Martínez-Plumed et al., 2020), 2) addressing how to select analytical opportunities (Hindle & Vidgen, 2020), 3) and considering roles, responsibilities, and to a lesser extent, infrastructure (Microsoft, 2023).

Despite these recent improvements to development methodologies, further prescriptive support is needed for organizations looking to adopt analytics for process innovation. The methodologies are mostly focused on tasks, but lack prescriptive guidance on socio-technical aspects, such as scoping, managing, and organizing analytics development (Vial et al., 2023). Furthermore, they provide little support for the deployment phase, where many organizations struggle in the transition from models to production systems (Lavin et al., 2022; Vial et al., 2021; Davenport & Malone, 2021).

The empirical setting of our research is a large global manufacturer and retailer based in Denmark that faced challenges in adopting analytics for process innovation.

We were engaged by the Vice President of Operations IT, who was facing an increasing demand from operations to support exploratory analytics initiatives aimed at process innovation. However, progress on these initiatives was slow, resulting in an increasingly impatient leadership in operations. At the same time, departments within operations were running initiatives without significant involvement from Operations IT, resulting in solutions that were unacceptable to IT without major rework. Prior to our engagement, the VP had invested in an operations data platform to reconcile the conflicting demands of development speed and scalability. We were tasked with designing an approach for analytics development capable of quickly constructing an *analytics demonstrator* that would allow stakeholders to assess its potential in process innovation, while remaining scalable from an IT perspective.

To develop this approach, we collaborated with the organization in a three-year Action Design Research (ADR) (Sein et al., 2011) project, during which we actively participated in the development of the operations data platform and used it to develop several analytics demonstrators. Drawing on this experience and literature on analytics, digital process innovation, and digital infrastructure, we constructed an approach and design principles for developing analytics demonstrators aimed at process innovation. We then proceeded to instantiate, evaluate, and update the approach and design principles in a final demonstrator.

This paper makes two main contributions to the discourse on analytics for process innovation. First, we provide six design principles and outline an approach for the development of analytics demonstrators in large established companies with complex digital infrastructures. In contrast to existing research, our approach and design principles emphasizes deployment throughout the development process (Davenport & Malone, 2021; Vial et al., 2021). We thus contribute with prescriptive knowledge on analytics development, which has seen surprisingly few contributions in recent years (Hindle & Vidgen, 2018) despite substantial change in the problem and solution domains. Second, we demonstrate the utility of adopting the digital infrastructure lens in analytics research. We found that analytics for process innovation in large established organizations is often highly infrastructural and thus can benefit from the rich concepts developed in this literature.

The rest of the paper is structured as follows. First, we establish the necessary background and motivation for our work. To begin, we review existing literature related to analytics for process innovation and methodologies for analytics development. We then present an infrastructural perspective on analytics, that emerged in the process of our design work and took on a central role in our derivation of design principles. We proceed to explaining our research approach, namely how we adopted ADR to develop design principles and outline an approach for analytics demonstrator development. Afterwards, we present our results by providing insights into the ADR cycles we undertook and formalizing the approach and design principles. This is followed by a discussion, where we highlight our contributions and their implications for research and practice, before finally concluding on our research.

Background

In this section we review existing research related to analytics-enabled process innovation. First, we clarify the role of analytics in process innovation and establish the need for *analytics demonstrators*. We then review and assess research on analytics development. Lastly, we draw on emerging research to conceptualize analytics as infrastructural, a perspective we leverage in our design work. This perspective has so far been largely neglected in the artefact (i.e., model) and task centric analytics research.

Analytics for Process Innovation

Analytics systems are increasingly used to enable redesign and improvement of operational processes in organizations. Whereas analytics was historically used primarily to support strategic and tactical decision making (Badakshan et al., 2022; Davenport, 2018), improvements in technology have led to analytics increasingly being embedded as core parts of operational processes (Davenport, 2018; Tarafdar et al., 2019). Analytics has been used to automatically identify fraudulent credit card transactions (Pozzolo et al., 2014), move from reactive to predictive maintenance (Dremel et al., 2020), automate inspection of quality in manufacturing (Tercan & Meisen, 2022), and more. Research has found that realizing organizational value from analytics requires coordinated socio-technical change (Dremel et al., 2017; Dremel et al., 2020; Tim et al., 2020) and thus amounts to more than just introducing new technology. The process level impacts of adopting analytics-enabled processes can range from incremental to radical change (Kunz et al, 2020; Enholm et al., 2021), however, compared to BPR, the scope of change is considerably smaller (Sedera et al., 2016). In terms of performance, analytics-enabled processes can lead to improvements in quality and efficiency (Enholm et al., 2021, Tarafdar et al., 2019).

To innovate processes with analytics, organizations must engage in extensive experimentation. Analytics is a weakly structured technology (Eley IV & Lyytinen, 2022) and organizations thus need to explore the potential of its' generic affordances in their process context. This exploration includes identifying, assessing, and implementing new digital technologies for process innovation (Rosemann, 2014). Predicting the performance of redesigned IT-enabled process at the design stage is difficult if not impossible, and thus organizational prototyping has been recommended to assess the value of a redesigned process (Davenport, 1993). These prototypes can take on varying degrees of fidelity ranging from simple mockups to pilot studies or simulations. With analytical systems, their data-dependent capabilities mean that assessing their capabilities requires the construction of a technical prototype. In the next section, we thus review methodologies for analytics development.

Analytics Development

Analytics research offers several development methodologies to support the analytics development process. The most popular of these, CRISP-DM (Wirth & Hipp, 2000), was created in the data mining era of analytics and has stood the test of time still being widely applied in research (Schröer et al., 2021) and practice (Piatetsky, 2014; Saltz, 2020). CRISP-DM was developed as a generic methodology that works across development contexts and technical infrastructures (Wirth & Hipp, 2000, p. 30). The methodology conceptualizes development as an iterative process consisting of six phases: 1) Business Understanding, 2) Data Understanding, 3) Data Preparation, 4) Modeling, 5) Evaluation, and 6) Deployment.

In recent years, several analytics methodologies have been proposed as alternatives to or extensions of CRISP-DM. CRISP-DM has been contextualized for specific domains such as manufacturing (Huber et al., 2019), and finance (Plotnikova et al., 2023). Recently, Martinéz-Plumed et al. (2021) presented a refined version called Data Science Trajectories (DST), which covers a wider range of analytics use-cases, amongst others to account for the increasingly exploratory focus of analytics. DST covers the traditional phases of CRISP-DM in addition to phases concerned with exploration and data management. Rather than offering a prescriptive sequence, the methodology suggests analytics projects combine the different phases depending on project needs. Another alternative is the Team Data Science Process (TDSP), which was designed specifically for predictive analytics deployed as part of software applications (Microsoft, 2023). In comparison to the other methodologies, TDSP shares a similarity in the phases considered, but is more opinionated in including a breakdown of roles and responsibilities for the tasks to be performed.

Across these methodologies there is a clear focus on the tasks involved in analytics development and an emphasis on iterative development due to the uncertainty involved. Less emphasis is placed on socio-technical aspects such as managing, organizing, and scoping analytics development projects. An exception is the recent work of Hindle & Vidgen (2020) that addresses scoping. They suggest combining CRISP-DM with top-down business modelling using soft-systems methodology and a simple prioritization approach to select opportunities. Furthermore, the methodologies are artefact-centric, focusing primarily on analytical model building and evaluation with little support for the deployment phase. In CRISP-DM, deployment is merely planned for but considered outside the scope of the analytics project (Wirth & Hipp, 2000), whereas in TDSP, deployment includes turning the model into an API (Microsoft, 2023) but not further technical integration. This lack of support for the deployment phase is problematic, given that organizations are struggling with the deployment process where models are matured into systems ready for production (Vial et al., 2021; Davenport & Malone, 2021). Emerging research is, however, becoming increasingly sensitive to the need for moving beyond an artefact-centric view of analytics. We turn to this emerging view of analytics as infrastructural in the next section.

Analytics as Infrastructural

The importance of infrastructure for realizing value from analytics is increasingly emphasized in research. Early research emphasized the technological analytics infrastructure, which is needed to provide computation and storage resources for analytical applications. The technological analytics infrastructure is one of the building blocks of big data analytics capabilities (Gupta & George, 2016; Mikalef et al., 2018) and an enabler of AI adoption (Enholm et al., 2021) and ML business value (Reis et al., 2020). In production ML systems, the model only amounts to a small component in the overall system with the rest consisting predominantly of infrastructure (Sculley et al., 2015). Owing to the complexity of these systems, significant engineering and research has been undertaken to reduce complexity by developing higher level components and standards (e.g., Kubeflow by Google and MLFlow by Databricks) and architectures for analytics infrastructures building on platform architectures (e.g., Philips-Wren et al., (2021) or Gröger (2021)).

More recently, the focus has expanded to consider the role of the operational digital infrastructure - the enterprise systems and infrastructure responsible for supporting and executing processes. The challenge facing organizations can amount not just to develop and deploy analytical systems, but also to make the necessary changes to their operational digital infrastructure to make this a possibility (Davenport & Miller, 2022). Augmentation use-cases require a mature data infrastructure that enables access to data from transactional systems, whereas automation use-cases furthermore require two-way integration with transactional systems (Shollo et al., 2022). As a result of this need for interaction between the operational digital infrastructure and analytics infrastructure, organizations need to determine how to organize and coordinate work between data scientists and software developers – a process full of challenges due to differences in workflows and skillsets (Nahar et al., 2022). Integration is further complicated by the fact that the operational digital infrastructure in many large established organizations is composed to a large degree of fragmented legacy systems. Data accessibility therefore remains a significant challenge, which can considerably prevent the development process from moving from the modelling stage towards deployment (Vial et al., 2021).

We found that the analytics deployment challenges increasingly documented in research have much in common with the challenges related to the *installed base* of systems as studied in research on digital infrastructures. Digital infrastructure research has a rich history of studying how the installed base of systems, developers, and users influence, and often thwarts, attempts to introduce new IT systems (e.g., Ciborra et al., 2000; Aanestad & Jensen, 2011; Grisot et al., 2014). In this stream of research, the architecture and governance configuration of the digital infrastructure has been found to have significant implications for innovation (Henfridsson & Bygstad, 2013; Bygstad & Øvrelid, 2020), and strategies and design principles (Hanseth & Lyytinen, 2010; Aanestad & Jensen, 2014; Koutsikouri et al., 2018) have been developed to increase success rates of projects that are undertaken within an ultimately non-ideal digital infrastructure. However, this wider perspective of viewing the analytics system in the context of the overall system landscape has so far not been widely adopted in the research on analytics.

Summary and Gaps

To summarize, our review of literature showed that the existing prescriptive knowledge base related to analytics falls short in addressing the increasingly relevant use-case of analytics for process innovation. Table 1 summarizes key differences between traditional analytics and analytics for process innovation. First, existing research on analytics and process innovation assume that the problem is relatively well-defined, whereas analytics for process innovation use-cases often require extensive exploration. While the explorative nature is to some extent dealt with by the iterative nature of data science process models, they focus mainly on task aspects of analytics and do not address socio-technical issues such as scoping, management, and coordination of analytics activities. Second, existing analytics research is artefact-centric, whereas analytics for process innovation is often highly infrastructural in nature. In what follows we attempt to address this gap by developing prescriptive support for the development of analytics demonstrators for process innovation, which consider the explorative and infrastructural nature of the work.

	Traditional Analytics	**Analytics for Process Innovation**
Business Problem	Well-defined	Often unclear
Problem vs. Solution-oriented	Problem-oriented	Solution-oriented
Nature of Data	Existing data from enterprise systems	Existing data from enterprise systems & novel data to-be collected (e.g., IoT)
Nature of Systems	One-off Insights or Daily, Weekly, or Monthly Batch Jobs	Higher frequency – often sub daily. Streaming or batch systems.
Nature of Integration	Weak or no integration	Integrated with Operational Digital Infrastructure

Table 1: Traditional Analytics compared to Analytics for Process Innovation

Research Approach

Our research had the goal of developing a pragmatic solution to the problem faced by our partner organization, while making an academic contribution in terms of generalizable prescriptive knowledge. Design Science Research (Hevner et al., 2004) was thus an obvious fit for a research approach and our setting fell under what Iivari (2015) calls *DSR Strategy 2,* which is concerned with the development of a solution to “a specific problem encountered by a client” (Iivari, 2015, p. 108) and “new, innovative design principles” (Iivari, 2015, p. 110). Due to our close engagement with our partner organization, we adopted Action Design Research (Sein et al., 2011) as our method. ADR is focused on the development of artefacts that emerge from design in an organizational context and the extraction of generalizable design principles.

In our study, we completed five ADR cycles (Sein et al., 2011). Each cycle consisted of the development of an analytics demonstrator, which was built on the data platform of our partner organization. The nature of the artefact developed differed slightly, but always consisted of a prototype of an analytics system or its major components (i.e., the model), as well as the development process used. The ADR project officially began in July 2020, where the lead author was embedded in the operations data platform team to become familiar with the problem domain. After six months of participative observation and development, mostly conducted remotely due to the COVID pandemic, work on the demonstrators started and continued until April 2023. The demonstrator work was separated into two stages: 1) an exploratory phase, consisting of the first four demonstrators, and 2) an evaluation phase, consisting of the last demonstrator.

The exploratory phase began in January 2021 and ended in August 2022. The purpose was to obtain a deep understanding of the problem and solution domains by participating in and running several analytics demonstrators, with the goal of constructing "knowledge-through-action" (Goldkuhl, 2008; Ågerfalk, 2010). Following the BIE stage, we entered the "Formalization of Learning" stage of ADR and derived initial design principles by synthesizing our experiences and grounding them in literature. Findings from research on digital infrastructures aligned with our experiences and provided powerful sense-making concepts.

The evaluation phase took place from January 2023 to April 2023. We relied on a naturalistic evaluation strategy (Venable, Pries-Heje, & Baskerville, 2016), as we wanted to evaluate the utility and applicability of the design knowledge in practice. In this phase, we initiated work on the final demonstrator with the goal of instantiating and evaluating our design principles. Following the evaluation, we entered the last stage of

“Formalization of Learning” and updated our design principles to account for the learnings from our last demonstrator. Figure 1 provides an overview of our research process, while Appendix A1 provides details on how we implemented the ADR principles. In the following section, we elaborate on the research context, before proceeding with details on the individual ADR cycles.

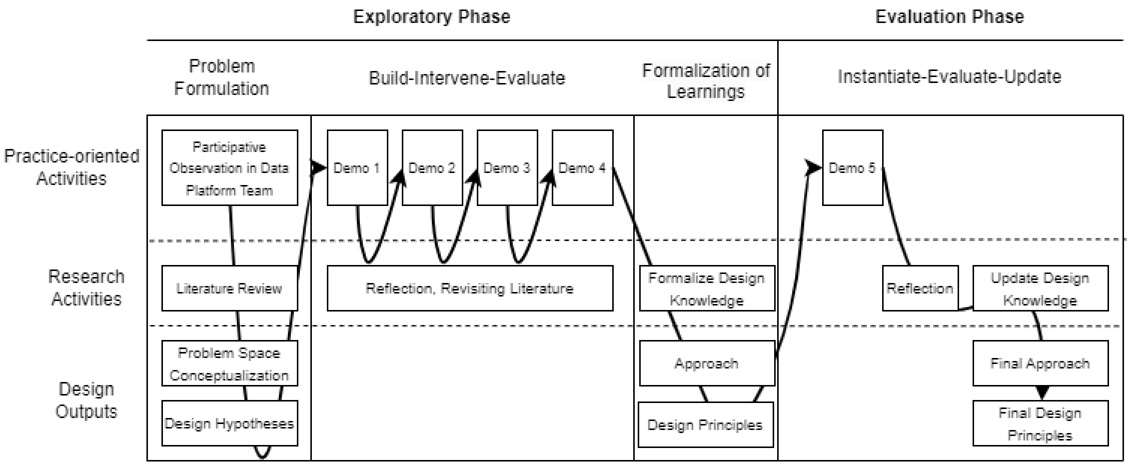


Figure 1: Overview of the research process detailing how ADR was applied in this study.

Research Context

Our partner organization is a large global manufacturer and retailer based in Denmark. The organization has over 25000 employees in more than 100 countries and operates five factories and 900+ stores globally. The organization has a long history of producing physical products but has in recent years increased focus on digital products as part of a digital transformation strategy. The digital transformation strategy includes a transformation of operations with the goal of realizing productivity gains while providing a joyful job experience for employees. To realize the digital transformation an Operations Transformation unit had been established, which was responsible for scouting, assessing, and piloting new digital technologies. Based in operations, this unit was technically supported by a combination of external vendors, research collaborations with multiple universities, and internal support from IT.

We were engaged by the Vice President of Operations IT. The VP was facing an increasing demand from operations to support exploratory analytics initiatives aimed at process innovation, but the existing IT organization and development approach was not up to the task. The VP was keen on supporting digital innovation in operations. He had already established agile ways of working with IT product teams and created a small team to support operations in exploration activities. Additionally, he saw the potential of analytics to significantly improve processes in operations. Significant infrastructure investments had been made in connecting production machinery and equipment and collecting data from them for analytics purposes. A major initiative to connect thousands of machines was still ongoing when we started. Despite the changes created and money invested in analytics, the VP felt that the analytics initiatives so far had failed to demonstrate much value. Progress on projects was slow, and many projects never got further than model building. As a response to this challenge, Operations IT had invested in creating a new team responsible for developing and operating an operations data platform. The purpose of the data platform was to support initiatives relying on data by providing a self-service experience, where IT would not become a bottleneck. IT product teams would become responsible for supplying data from systems they own to the data platform. Innovation initiatives could then access and leverage the data from the data platform without requiring substantial support from IT.

Our research took place within this context as part of the exploration work related to operational process innovation. Our demonstrators had both IT and operations as stakeholders. In addition to the Operations Transformation unit, we collaborated with two other units in Operations, namely the Process Engineering unit responsible for the largest manufacturing process, and the Process Analytics Center responsible for development of the same process. Figure 2 provides an overview of the main organizational stakeholders in the research.

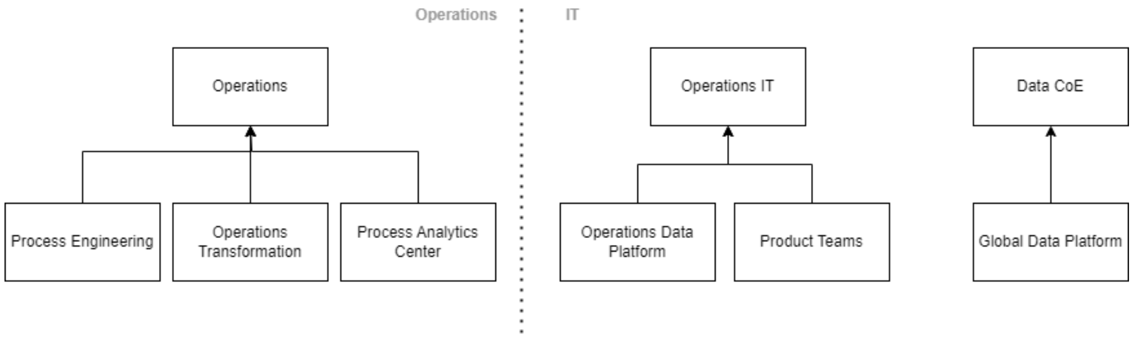


Figure 2: Involved organizational units in the ADR project in Operations and IT.

The ADR Cycles & Engagement

During our research, the first author worked closely with the organization, where (s)he received the role as an employee in the organizational IT systems, had an employee ID card, company computer, and access to internal systems. As such, the first author was largely treated by other employees as an employee reporting to the VP of Operations IT. In the first six months of our research, the first author was furthermore embedded in the operations data platform team and participated in daily standups, team meetings, and architecture and development tasks to get acquainted with the problem domain and technology landscape.

Following this six-month period of participative observation, work started on the ADR cycles. Each of the five ADR cycles we conducted in our research consisted of a single analytics demonstrator. Table 2 presents details on each of the ADR cycles and demonstrators. The first author acted as a designer in each of the demonstrators and took part in both conceptual and development work, sometimes in collaboration with other developers. In all but one of the demonstrators, the artefacts developed were operational analytics system prototypes that leveraged real live data. The exception was the third demonstrator where only an ETL process and a model was developed before the demonstrator came to an end.

Demonstrator	Role	Stakeholders	Artefact	Evaluation
Data Quality Monitoring	Designer of solution while embedded on the data platform team.	Data Platform team as primary stakeholders. They helped scope the problem and provided feedback throughout the development process.	Anomaly Detection System	System deployed live on the data platform for a period of two months.
Process Visualization	Designer of the analytics component of a larger demonstrator in collaboration with the organization and a university. Another researcher and external consultants also participated.	Innovation Managers in Operations Transformation as primary stakeholders. Developers and engineers as secondary stakeholders.	Near real-time Dashboard	Physical prototype demonstration for Innovation Managers. Workshop to disseminate findings to developers & engineers, and three follow-up presentations given.
ML-based Process Monitoring	Co-Designer in collaboration with an industrial PhD & Innovation Manager.	Operations Transformation team as primary stakeholder. IT Product Team and Data platform team as secondary stakeholders.	ML Models	Findings presented to several IT stakeholders & machine engineers. Ended up influencing specifications for future machine purchases.
Machine Parameter Optimization	Part of design team consisting of three data managers, and one external consultant. Responsible for the analytics component together with one of the data managers.	Manager & Data Managers in Process Analytics Center as primary stakeholders.	ML System (Existing Model)	Proof-by-construction. Prototype built, deployed, and tested together with the design team. Findings disseminated to managers. Further work continues in the Process Analytics Center.
Process Productivity	Designer of solution. Team consisted of a product manager	Process Engineers, and Product Manager in Operations	Anomaly Detection System	Prototype deployed and demonstrated for user organization.

	and two user representatives.	Transformation as main operations stakeholders. IT product team as main IT stakeholder.		Results disseminated to IT Teams & Operations Innovation.

Table 2: Description of the analytics demonstrators making up the five BIE cycles.

Results

In this section we present the results of our ADR process. We structure the results according to the two overall phases in our research. First, we present results from the exploratory phase, starting with the initial problem formulation, followed by a presentation of the demonstrators in this phase and end with derivation of initial design principles and outlining our approach. Second, we present the findings from the evaluation phase, where we instantiate the design principles, provide insights into their application, and finish by updating our design principles and approach based on our learnings.

Exploratory Phase: Diagnosis and Design Principles Development

In the exploratory phase, the first author was embedded in and worked closely together with the organization to explore the problem and solution space. In what follows, we detail the evolution our conceptualization of the problem and solution space went through, starting from the initial problem brought to us by the Vice President of Operations IT.

Initial Problem Formulation & Design Hypotheses

We were engaged to develop an approach for quickly demonstrating the potential value of analytics in a scalable manner. The demonstrators would essentially take the shape of *minimum-viable-products* (MVP) that would allow validating both 1) technical feasibility, and 2) value from a user and process perspective. As the result of multiple talks and scoping meetings with the VP, we arrived at the following two requirements for the approach:

- **Meta-Requirement 1 (MR1):** Speed. The approach should allow a small team to quickly build a prototype than can be shown to users.
- **Meta-Requirement 2 (MR2)**: Scalability. The approach should result in prototypes that are built on scalable technologies and fit with the organizational IT landscape.

Throughout the process of scoping the problem, the first author was embedded in the newly established Data Platform team that had been created in Operations IT. The purpose was to obtain an in-depth understanding of the IT landscape of the organization, as well as the potential of the data platform in enabling analytics development. As a result of this process an initial conceptualization of the problem space emerged. An analytics system consists of a data component, a model (or processing) component, and a system component. A demonstrator development approach would have to consider how to frame the analytical problem, and how to build and deploy each of the three components. Additionally, it would have to consider how tasks should be distributed and coordinated among the IT product teams, the data platform team, and the innovation project team. Another aspect concerned whether generic components should be built prior to demonstrator projects as infrastructure, such as a generic data pipeline serving as the data component, or be built in individual projects.

Inspired by the ideas behind the data platform, our own experiences, and the emerging technical literature on machine learning systems, we arrived at initial design principles, or *design hypotheses*. As envisioned in the data platform, generic data pipelines would be prebuilt by the IT product teams and specialized for use-cases as part of the demonstrator. This in turn would speed up development of data components. Model development would be sped up by relying on automated ML and standard models implemented in open-source packages, thus minimizing custom development. System components would be implemented via. rapid application development (RAD) tools, and deployment would happen via. generic functionality in the data platform. The result would be, except for the UI, a demonstrator built using scalable, IT-approved technologies, in a short time due to the use of pre-built components. We then set out to test these, at the time, non-formalized ideas by using them to develop analytics demonstrators.

Build-Intervene-Evaluate: The Four Demonstrators

We conducted a total of four demonstrators, or BIE cycles, in the exploratory phase. For reasons of brevity, Table 3 presents a summary overview of the four demonstrators in terms of 1) the problem addressed, 2) the BIE cycle, and 3) reflection and main learnings. For a more detailed description of the four demonstrators, we refer the reader to Appendix B.

In terms of their characteristics, the four demonstrators were all focused on operational processes, but varied in analytical complexity, ranging from descriptive (*Demonstrator 2)*, to predictive (*Demonstrator 1 and 3)*, and finally prescriptive (*Demonstrator 4*) analytics. The first three demonstrators were augmentation use-cases, while the last was an automation use-case. In all but *Demonstrator 3*, the artefact developed was an operational analytics system prototype rather than just an analytical model. While we originally set out to develop a system prototype in this demonstrator, disappointing results of our model development efforts made us stop the project after model building. The extent of existing data infrastructure also varied in the demonstrators. In *Demonstrator 1*, all the necessary infrastructure was pre-existing. In the other three, some degree of data infrastructure development was necessary.

	Problem	BIE	Reflection
Demonstrator 1: Data Quality Monitoring	Automatically detect and alert in case of data quality problems for operations data.	A data quality monitoring system was developed by the first author using an off-the-shelf anomaly detection SaaS. The solution was integrated with the data platform and consisted of data integration, dashboards, and alerting components. The solution was evaluated via live use on the operations data platform.	Use of the data platform, pre-existing data pipelines and an off-the-shelf system enabled fast prototype development and "live" testing.
Demonstrator 2: Process Visualization	Near-real time visualization and data collection of manufacturing process data using reference architecture.	The BIE cycle consisted of two parts: 1) provisioning infrastructure according to the reference architecture, and 2) developing the visualization solution. Infrastructure in the form of an edge container and a cloud message bus was provisioned as a collaborative effort. The visualization solution was developed by the first author and consisted of a data pipeline, a streaming ETL job on the data platform, and a PowerBI dashboard. The solution was evaluated through presentations to both Innovation Managers and engineers from Operations IT.	Setting up the infrastructure as per the reference architecture took some effort. Once in place, the infrastructure enabled quickly constructing new data pipelines. Together with low analytical complexity and an off-the-shelf system, this made fast prototyping possible.
Demonstrator 3: ML-based Process Monitoring	Leverage ML with real-time machine and sensor data to detect and alert in case of manufacturing process quality problems.	The BIE cycle consisted of two model building iterations. An initial dataset was constructed for model building, which required significant data cleaning was due to data quality issues and resulted in a small dataset. Access to data on element quality for use as labels, required assistance from busy outside product teams resulting in a 1,5-month wait for data access. A first model building iteration using statistical and ML models was disappointing due to our small dataset. A further iteration commenced with a larger dataset, but once again faced a two month wait for data access. The models were evaluated statistically using cross-validation. The demonstrator was stopped due to insufficient performance in the second model building iteration.	Data accessibility and quality resulted in significant delays and prevented user involvement. The nature and maturity of data infrastructure determines the extent of inter-team coordination necessary in demonstrators.
Demonstrator 4: Machine Parameter Optimization	Automatically adjust machine parameters on the fly using ML to improve element quality.	A conceptual system architecture was developed in collaboration with the stakeholders through several workshops. Then, an instantiation was developed as part of a two-week sprint. The instantiation consisted of a data pipeline, reimplementation of the existing model in the data platform, and a rule-based component that adjusted parameters based on model predictions. The instantiation was deployed and evaluated in a proof-of-concept, where it was used briefly to adjust parameters on one machine.	Existing infrastructure in the form of APIs and the data platform made it possible to demonstrate an automation use-case in a very short time frame. However, the analytical complexity of the project, was limited as it used an existing model.

Table 3: Overview of the demonstrators in the exploratory phase including learnings.

Formalization of Learning

Following the four ADR cycles, we reflected on the main challenges we experienced in developing the four analytics demonstrators and updated the problem conceptualization. The challenges were all related to the trade-off between speed and scalability, and they had prevented us from meeting MR1, that is being able to quickly demonstrate value in working prototypes. We formalized this as two main challenges. First, analytics demonstrators inevitably ended up having to deal with data that was not integrated with the data platform. In this case, discovering and accessing existing data turned out to be a significant problem. The challenge turned out to lie in the need to coordinate several independent teams, each with their own priorities and backlog. Furthermore, establishing this missing data infrastructure during analytics initiatives was challenging. The time-consuming nature of building reusable infrastructure meant that there was constantly a tension between innovation and infrastructure work. Second, developing the model component could often become quite complex. Unless the data was already in an acceptable quality and format suitable for model training, data cleaning and preprocessing would take up significant time. This made it more difficult to actively involve users in development. We thus found that there was a limit to the combined infrastructure and analytical complexity of a demonstrator if prototypes were to be developed quickly. Attempting to carry out demonstrators exhibiting both high analytical and infrastructure complexity, such as in *Demonstrator 3*, made it impossible to quickly deliver a prototype to users.

Further reflecting on our learnings, we found them to be in line with findings in the research on *digital infrastructure* or *information infrastructure*. The first challenge could be framed as having to deal with the installed base of digital infrastructure. Drawing on findings from the digital infrastructure literature, we proceeded to formalize our design principles. Where appropriate, we linked principles specific to the development of analytics systems to higher level generic principles from digital infrastructure literature. Table 4 lists the derived design principles along with references to the relevant digital infrastructure literature.

Consolidating our learnings from the four demonstrators, we ended up with the development approach presented in Figure 3. The approach showcases the generic system architecture that emerged from our demonstrators as well as roles and responsibilities associated with the components. Overall, the approach distinguishes between infrastructure and innovation focused analytics work. The ability to develop demonstrators quickly and successfully in a scalable manner requires the presence of infrastructure prior to the fast and iterative development together with the users. As such, following *Problem Framing* and *Scoping*, the approach proceeds to an infrastructure stage, unless the necessary infrastructure is already available. In this stage, the relevant IT Product Teams build reusable data pipelines that move data from enterprise systems into a data platform, where they can be accessed for further use-case specific development. Once the infrastructure is ready, the second stage of demonstrator development can proceed by the project team. Here, specialized data components are built leveraging the data platform, and model and system components are built or configured as per the requirements of the use-case. The data platform thus serves as an integration point between the operational IT infrastructure and analytics development.

Design Principle	Description	Justification
DP1: Use standard components. - **DP1.1**: Leverage existing ML services, Automated ML, pre-trained models, and models in that order. - **DP1.2**: Leverage standards for packaging and deploying models. - **DP1.3**: Leverage existing lightweight solutions for UI.	To speed up initial prototype development, use standard components for data (e.g., generic data pipelines and API's), model (e.g., AutoML, open-source software), and system (e.g., ML-as-a-Service, SaaS, model deployment standards) components when possible.	*Empirical*: Use of ML-as-a-Service (*Demonstrator 1*), PowerBI (*Demonstrator 2)* and model deployment standard (*Demonstrator 4*) reduced development efforts and complexity considerably. *Literature*: Prefer lightweight engineering and IT for digital process innovation (Schmiedel & vom Brocke, 2015; Bygstad & Øvrelid, 2020)
DP2: Prefer self-service data sources.	To speed up initial prototype development, prioritize self-service data sources in problem framing.	*Empirical*: Non-self-service data sources expands the project scope by requiring coordination with other IT teams, exposing the project to delays (*Demonstrator 3*)
DP3: Start simple, demonstrate quick wins. - **DP3.1**: Augmentation before Automation – Visualization before Predictive, and Prescriptive Analytics - **DP3.2**: Small scale projects with few data sources and stakeholders.	To enable fast development and user involvement, start simple in framing the problem and incrementally demonstrate value before increasing complexity.	*Empirical*: Targeting a large complex problem immediately resulted in slow progress and no results to show when we couldn't solve the complex problem within the demonstrator timeframe (*Demonstrator 3).* *Literature*: Delivering simple capabilities as early as possible facilitates adoption and buy-in from stakeholders (Hanseth & Lyytinen, 2010; Grisot et al, 2014)
DP4: Select flexible technologies. - **DP4.1:** Select technologies with configurable rules and thresholds. - **DP4.2:** Select technologies with configurable user interfaces.	To facilitate fast iterations in prototype development, select flexible technologies.	*Empirical:* Iteratively improving the UI and analytics thresholds was very fast with a configurable ML-as-a-Service solution (*Demonstrator 1*). *Literature:* Flexible and configurable technology allows for configuration-in-use thus facilitating mutual adaptation (Bygstad & Øvrelid, 2020)
DP5: Leverage the installed base. - **DP5.1**: Prioritize and leverage data that is already in use. - **DP5.2:** Leverage existing infrastructure and integrations over building new infrastructure	To increase success rate and speed in analytics demonstrators, leverage and build upon the existing IT systems and infrastructure.	*Empirical:* None of the demonstrator would have been feasible without prior infrastructure. Where data was not in use, significant data quality problems were present (*Demonstrator 3).* *Literature:* Building upon the installed base reduces development of new infrastructure and adoption and integration efforts (Hanseth & Lyytinen, 2010).

Table 4: Derived Design Principles following the exploratory phase.

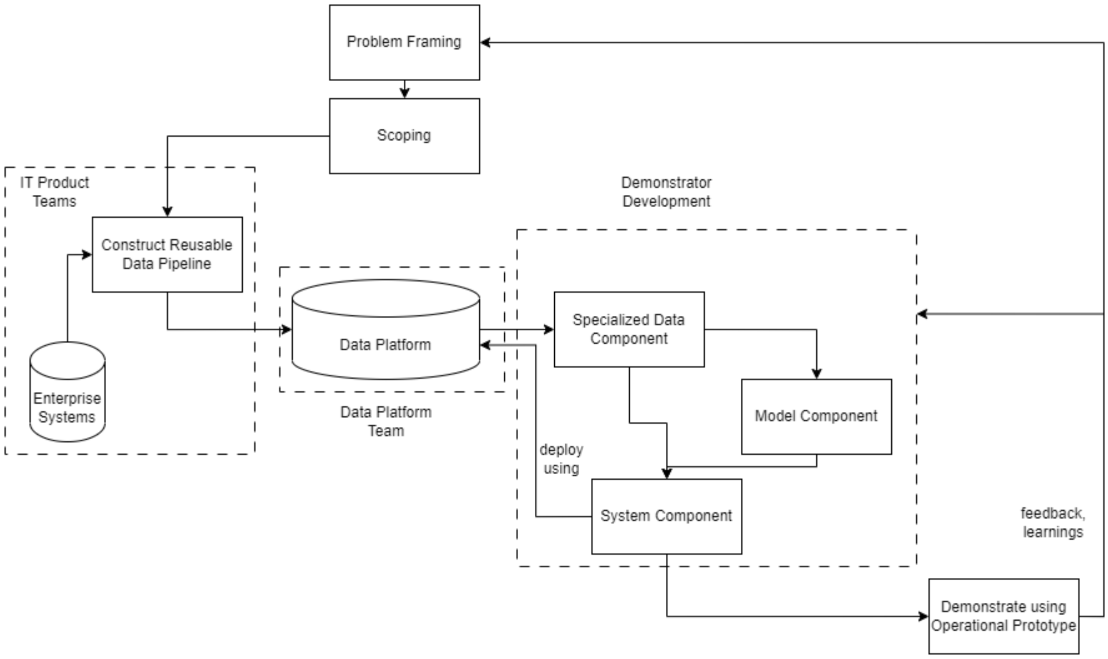


Figure 3: The analytics demonstrator development approach including architecture, and roles and responsibilities.

Evaluation Phase: Instantiation & Evaluation of Design Principles & Approach

To evaluate our design principles and the approach, we conducted another ADR cycle where we sought to instantiate them, in effect presenting a "proof-of-concept" evaluation (Iivari et al., 2021). The use-case in this demonstrator was highly exploratory, consisting of a data source and a goal of improving productivity. There was thus no well-defined analytical problem formulation, in line with several of the other initiatives we participated in earlier. Below, we present the problem context and the involved stakeholders, before describing how we instantiated the design principles and approach. We then proceed to evaluate and finally update our design principles based on our learnings.

Problem Context

Machine and sensor data from the manufacturing process remained unused following the failure to develop the process monitoring system in the third demonstrator. Based among other things on the data quality issues we identified in that demonstrator, IT had decided to rebuild the entire data collection infrastructure using custom built software rather than relying on a vendor system. The goal in this last demonstrator was to find an analytics use-case for this data that would improve productivity. The first author teamed up with the Product Manager to identify potential users of an analytical system and identified two process engineers in the Process Engineering department.

Instantiation of Design Principles & Approach

The demonstrator started with efforts to frame the problem. We held an initial meeting with the process engineers and Product Manager to introduce the project. In line with *DP3: Start simple, demonstrate quick wins* all agreed to start with visualization of the data before moving on to more advanced analytics. The initial framing concerned using visualizations of the data to aid engineers in troubleshooting during breakdowns or quality problems. We also used this meeting to get the process engineers to identify additional data sources they would find useful. Initial scoping took place as the first author got to work on an initial design for the first visualization prototype. This involved identifying options for accessing data from the data sources identified by the process engineers. Following *DP2: Prefer data sources with self-service access,* we excluded three potentially useful data sources as integrating them would require significant help and effort from other product teams. The data sources that were selected for consideration were either accessible in the data platform or through an API. The design phase furthermore involved selecting technologies for the system and data components. Given the sensor and machine data to be visualized was high frequency time series data, we needed an architecture that was capable of interactive visualization of time series. Following *DP1: Leverage off-the-shelf components* and *DP4: Use flexible technologies* we selected a SaaS version of a time series database (InfluxDB) and a SaaS version of a highly configurable time series visualization system (Grafana).

Development then started on the data component. The main task involved creating data pipelines to process and transport the data from the data platform into the time series database. Additionally, several wrappers (or "gateways") were built to allow Grafana to access data from the enterprise APIs. In line with *DP5: Leverage the installed base* we leveraged the data platform and existing APIs even though they were less than ideal choices from an architecture standpoint. As an example, development of the wrappers was necessary as the API request plugin in Grafana did not support the authentication mechanism in the enterprise APIs. Another example included the use of existing order master data, which was currently stored at another cloud vendor, requiring the need for the data pipelines to join data across cloud vendors. Developing the data component took some time, during which users were only minimally involved. However, once the data component was in place, developing the first dashboards was a fast endeavor. A simple dashboard was built where the process engineers could search for a particular machine and view data from sensors and machine settings in near real time. Through several iterations based on feedback from engineers, we developed a dashboard that provided an overview of the data of interest for troubleshooting purposes.

In the process of using the dashboards with engineers, we identified significant data quality problems for key sensor and machine settings. These were forwarded to the IT product team responsible for the data collection infrastructure, however, the team was busy on other tasks and only able to squeeze in work towards the end of the demonstrator. Aware of the quality problems, we worked with the process engineering department to identify analytics use-cases in addition to the visualization developed. Throughout this process we followed *DP3: Start simple, demonstrate quick wins*. First, we focused on identifying relatively simple augmentation use-cases. From among three identified alternatives, we proceeded with a maintenance use-case that consisted in monitoring a key sensor measurement for signs of instability, which the engineers hypothesized would be due to the need for maintenance. Second, we decided to start with a simple modelling approach leveraging rule-based statistics rather than using ML or AI. Following exploratory data analysis of the key parameter, the first iteration operationalized instability as the difference between the average value in the last 10 minutes compared to the average in the last hour. This required minimal implementation as it could be implemented as a simple query in Grafana. We proceeded to identify an initial threshold that would catch instabilities we had manually identified.

To evaluate the system, we then used the system to generate alerts for several days and reviewed the results. We learned that too many false alarms were generated in non-running production, e.g., when a changeover had occurred, or when adjustments to the process were made by operators during troubleshooting. In these situations, an operator would already be at the machine, and hence the alerts were not valuable. Another iteration commenced where production status data was integrated with the sensor data, so that alerts were only fired during automated running production. Another use of the alerts commenced showing promising results with much fewer false positives. Following suggestions from the engineers, another iteration implemented another alert that compared the sensor value to an expected value rather than the relative approach of comparing it to the last hour. While more intuitive to the engineers, this exposed incorrect master data, which led to the wrong expected value being joined to the sensor data. At the end of this iteration, we had developed a dashboard along with multiple alerts to monitor the key sensor measurement. Table 5 summarizes how the design principles were instantiated as design decisions in the demonstrator. In what follows, we evaluate the demonstrator and the design principles.

Design Principle	Design Decisions (Instantiation)
Leverage off-the-shelf components to the extent possible	InfluxDB (Time series Database) and Grafana (UI, Alerting) selected as main components.
Prefer data sources with self-service access	Three potentially useful data sources not prioritized in project due to impossibility of self-service access.
Start simple, demonstrate quick wins	Started with visualization and moved toward augmentation with simple rule-based statistics. Four main data sources, two users from one department, and one IT manager involved as main stakeholders.
Use flexible technologies	Selection of Grafana as it allows for simple configuration of alerting rules and thresholds via. drag-and-drop UI design, and low-code queries.
Build upon the installed base of systems when possible	Use of the organization's data platform for development of data pipelines. Existing data pipelines and APIs used for data sources as possible and specialized for the use-case.

Table 5: Design decisions instantiating initial design principles in the final demonstrator.

Evaluation

To evaluate our design principles, we examine and evaluate both 1) the demonstrator and its outcomes, and 2) the applicability and appropriateness of the design principles as instantiated in the demonstrator. First, for the demonstrator we conducted concurrent (Sein et al., 2011) and naturalistic (Venable, Pries-Heje, & Baskerville, 2016) evaluation. Throughout the demonstrator, the first author met most weeks with the end users and presented an updated demo and solicited feedback. When the analytics system had stabilized, we also subjected it to evaluation by using it with real data to generate alerts. The evaluation showed that we were able to detect issues where the engineers would like to be alerted, while keeping the overall number of alerts manageable. The remaining challenge was related to false positives caused by master data quality issues. After the last demo of the system, the users commented:

> "What we have done is good work – and we can see how it is useful. We have a use-case, and an interesting solution that can also be used for monitoring different

> parameters, so that it is generalizable. Now, we need to be able to trust in the data before we invest further resources in using it actively." – Process Engineer

We furthermore presented the findings to other stakeholders throughout the demonstrator with positive feedback. An Innovation Manager in Operations Transformation was keen to figure out how to continue work on the project and was actively looking for resources to continue it. From an IT perspective, the demonstrator had successfully managed to involve users. The Product Manager of the IT team commented:

> "This is great work. In general, we are just starting to get business to look at data [from the machines]. Until now, it has just been numbers to them, but now we are starting to challenge them and getting them to look at data, engage with it, and drive decisions based on data rather than assumptions or emotional decisions." – Product Manager

Second, reflecting on the appropriateness of the design principles, we consider that they supported us and significantly influenced our choice of technologies, our problem framing, scoping of data to include, and the overall development approach. These are all significant design decisions that are not currently addressed by existing analytics methodologies. The use of off-the-shelf components, self-service access data sources, and building upon the installed base in combination with the initial small scope of visualization, meant that we were able to develop a working prototype quickly with real live data. Presentation of the prototype gathered support from our stakeholders and allowed them to better support us in the generation of use-cases for more advanced analytics to build on top of the existing solution. The use-case we decided to pursue was the result of a demo session for the user's department, where one of their colleagues suggested the use-case following a demo of the prototype. The use-case was not one we had considered beforehand and was thus a bottom-up use-case made possible by our iterative development. The use of flexible technology furthermore made it possible to quickly iterate on the user interface and functionality between the weekly demo sessions. In presenting the demonstrator to a Data Scientist from the Data CoE, she commented on the overall approach of our demonstrator (as embedded in the design principles):

> "We have arrived at similar findings [in the Data CoE]. We are also very much working towards as quickly as possible implementing a prototype and assessing whether there is value together with users" – Data Scientist, Data CoE

Formalization of Learnings

The design principles formulated in our exploratory phase were all implemented and contributed to the success of the final demonstrator. Nonetheless, we still ran into unexpected issues related to the tension between infrastructure and innovation work despite employing our design principles. In particular, the data quality proved to be significantly worse than what had been promised by the IT product team responsible for implementing the data infrastructure. Furthermore, they were not able to prioritize fixing the data quality problems until late in the demonstrator. This led to delays in being able to test the solution in use and resulted in the users becoming slightly skeptical about the data. Reflecting on the situation, we found that any demonstrator building on a generic data infrastructure is likely to require adjustments to the infrastructure, unless the infrastructure is very mature. Based on this learning, we arrived at the need for another design principle to ensure that resources were available to make infrastructure adjustments as demonstrators take place:

DP6: Ensure access to infrastructure developers during analytics initiatives.

Discussion

In this paper, we set out to develop prescriptive knowledge about the development of demonstrators for analytics-enabled process innovation in large established organizations. Taking outset in the practical problem faced by a large global manufacturer and retailer based in Denmark, we sought to develop an approach that allows for fast assessment of the process innovation value of analytics, while remaining scalable from an IT perspective. Best practice in digital process innovation research is to leverage simple standard technologies in light engineering processes, while working closely together with users to assess the technologies potential "in-use", thus facilitating mutual adaptation of processes and technology (Schmiedel & vom Brocke, 2015; Bygstad & Øvrelid, 2020). However, we found that demonstrating the value of analytics for process innovation in large incumbent organizations is complicated by 1) the technological complexity of modern analytics, and 2) the complexity of their digital infrastructure, often characterized by legacy and silo systems.

We developed solutions to these problems through an iterative ADR process, where we took part in the development of five analytics demonstrators. Based on our experience and insights from the literature on digital infrastructure, we developed an approach and derived six design principles (see Table 6). Our learnings highlight 1) the importance of being sensitive to the existing digital infrastructure in the early stages of analytics development, 2) adopting a system rather than model focus, and 3) being aware of the need to trade-off infrastructural and analytical complexity. Our findings thus present a prescriptive or *Design and Action* theoretical contribution to the literature on analytics development and digital process innovation in the form of a nascent design theory (Gregor & Hevner, 2013).

	Design Principle
DP1	Leverage off-the-shelf components to the extent possible
DP2	Prefer data sources with self-service access
DP3	Start simple, demonstrate quick wins
DP4	Select flexible technologies to allow for design-in-use
DP5	Build upon the installed base of systems when possible
DP6	Ensure access to infrastructure developers during analytics initiatives

Table 6: Updated and final design principles.

Implications for Research

Our study has implications for existing discourses on analytics and action-oriented design science. First, our findings suggests that analytics development methodologies and design knowledge need to expand in scope to consider analytics in its broader organizational context. Existing development methodologies focus mainly on tasks and neglect important aspects such as how to organize and manage analytics efforts (Vial et al., 2022). Whereas the context independent nature of CRISP-DM was a deliberate choice (Wirth & Hipp, 2000), the challenges exhibited by organizations in realizing value from analytics suggests that further prescriptive knowledge is needed. Our findings highlighted that both the analytics and operational digital infrastructure are important contextual factors with implications for how to organize and manage development. Our findings are thus in line with emerging research suggesting the importance of data accessibility in AI projects (Vial et al., 2021) and considering deployment concerns in the early phases of analytics (Davenport & Malone, 2021). While, our design principles and approach make an initial contribution to this area, significant further research is needed to progress towards a mature design theory for process innovation uses of analytics. We see the creation of updated conceptualizations of both product and process aspects of analytics that consider infrastructural and analytical complexity as necessary steps along this journey.

Second, our conceptualization of the problem of analytics-enabled process innovation as an infrastructural problem suggests a new fruitful avenue of research on the adoption of analytics. By illustrating the usefulness of findings from research on digital infrastructure in our research, we open an avenue for leveraging this lens to investigate in detail the challenges organizations face in realizing value from analytics. Despite the importance of infrastructure in analytics (Shollo et al., 2022; Reis et al., 2020), existing research on analytics infrastructure is mainly technical (e.g., Phillips-Wren et al., 2021; Breck et al., 2019) or practice-oriented (e.g., Gröger, 2021) and does not substantially leverage existing IS theory. Leveraging a digital infrastructure lens to examine how successful analytical infrastructure such as organizational data platforms come into being is one such opportunity for developing research with both practical and theoretical significance.

Lastly, our research has implications for the discourse on action-oriented design science research. Recent research has suggested how DSR can be integrated as part of organizations digital innovation processes (Hevner & Gregor, 2022). Our application of ADR demonstrates how iterations using multiple distinct but related demonstrators in the same organizational context can be a fruitful way of addressing the dual goals of practical and research utility. This use of ADR allowed our research to practically contribute to the organization’s digital transformation process as the demonstrators acted as part of the exploratory innovation process. At the same time, it allowed us to contribute to IT management through insights into how the innovation work should be organized and supported by the digital infrastructure. Our study thus adds to recent literature showing how ADR can play a role in supporting organizational transformations (Chen et al., 2022).

Implications for Practice

Our research also has implications for practice. First, our problem space conceptualization highlights the importance of the operational digital infrastructure or *operational backbone (*Ross et al., 2019*)* in exploring and realizing the value of analytics. Whereas much focus on analytics infrastructure has been around the need for computational capabilities to process "big data" or train and deploy ML models, our work suggests that the socio-technical challenges of having to deal with heterogenous legacy systems in the operational backbone plays a larger part in the infrastructural challenges in analytics. Investing in analytical infrastructure, while necessary, is only the first part of the battle. The remaining challenge consists of transforming the operational backbone and integrating it with the analytical infrastructure - a process, which is not merely technical, but also requires establishing how to organize the collaboration between the two domains. Understanding that analytics process innovation projects can differ significantly in their infrastructural complexity can aid managers in obtaining more accurate expectations of the investments required to realize analytics use-cases and thus contribute to reducing over expectations towards technologies such as AI and ML.

Second, our design principles provide practical guidance and support for data scientists, IT engineers, and innovation managers in scoping and developing use-cases. Our design principles highlight the importance of adopting an incremental strategy where the initial scope is purposefully kept small and simple and slowly expanded as the demonstrators prove valuable together with users. To the extent that complex analytics and models are to be leveraged, off-the-shelf solutions should be preferred if extensive user involvement is desired. Otherwise, the projects might have to take on the character of longer-term R&D rather than demonstrators.

Limitations & Future Research

Our research is not without limitations. An important issue with research such as ours that follows Iivari's *DSR Strategy 2 (*Iivari, 2015*)* concerns the reusability of design principles and prescriptive knowledge in other contexts (Iivari et al., 2021). In our case, we consider in particular organizational characteristics and analytics strategy as important contextual factors that influenced our findings. Considering organizational characteristics, grounding our design principles in the literature on digital infrastructure, suggests that our findings hold for organizations with complex digital infrastructures characterized by heterogeneity and legacy. They should thus be relevant for large established organizations across industries, such as finance, manufacturing, health care, and energy. Despite this grounding in theory, we acknowledge that the design principles were only instantiated in one organization by the first author. Further research could adopt the design principles in different contexts or validate the design principles with external practitioners to test their external reusability (Iivari et al., 2021).

In terms of analytics strategy, our partner organization saw analytics as an opportunity to differentiate itself and thus invested significantly in developing its internal analytics competences, which motivated the need to conduct extensive exploratory work. Future research should investigate whether the design principles and approach remain useful for adopters that have chosen to rely on external vendors for analytics development, e.g., as described in Vial et al. (2023). We acknowledge that our findings are less applicable for organizations that take a more cautious approach towards analytics and wait for enterprise systems vendors to implement analytics as part of their integrated offerings.

Conclusion

We set out to develop prescriptive knowledge for the development of analytics demonstrators aimed at process innovation. Through an ADR process with a large global manufacturer and retailer based in Denmark, we developed six design principles and outlined an approach to meet their goals of building fast and scalable operational prototypes to assess the value of analytics with users. Our contribution moves beyond the task and artefact orientation present in existing methods by considering the socio-technical digital infrastructure in which the analytics system is to be embedded. We found that consideration of the digital infrastructure becomes crucial when analytics is used for process innovation rather than insights, as the digital infrastructure becomes an important force that shapes the development process. A failure to seriously consider the digital infrastructure is arguably partially responsible for the deployment problems that many organizations face in attempting to operationalize analytics. The increasing adoption of analytics for process innovation in practice calls for updated and contextualized conceptualizations. Our work suggests that drawing on digital infrastructure as a lens is a promising starting point for such endeavors.

Declarations

Funding

The work was supported by Manufacturing Academy of Denmark as part of the MADE FAST research and innovation program under Grant number 9090-00002B.

Competing Interests

The authors have no competing interests to declare.

Data Availability

The data generated during and/or analyzed during this study cannot be made publicly available due to confidentiality agreements made with the participating organization to protect its commercial interests. Further information and details regarding the data can be obtained from the corresponding author upon request.

References

Aanestad, M., & Jensen, T. B. (2011). Building nation-wide information infrastructures in healthcare through modular implementation strategies. The Journal of Strategic Information Systems, 20(2), 161–176. https://doi.org/10.1016/j.jsis.2011.03.006

Ågerfalk, P. J. (2010). Getting pragmatic. European Journal of Information Systems, 19(3), 251–256. https://doi.org/10.1057/ejis.2010.22

Badakhshan, P., Wurm, B., Grisold, T., Geyer-Klingeberg, J., Mendling, J., & vom Brocke, J. (2022). Creating business value with process mining. The Journal of Strategic Information Systems, 31(4), 101745. https://doi.org/10.1016/j.jsis.2022.101745

Benbya, H., Pachidi, S., & Jarvenpaa, S. L. (2021). Special Issue Editorial: Artificial Intelligence in Organizations: Implications for Information Systems Research. Journal of the Association for Information Systems, 22(2), 281–303. https://doi.org/10.17705/1jais.00662

Breck, E., Polyzotis, N., Roy, S., Whang, S. E., & Zinkevich, M. (2019). Data Validation for Machine Learning. Proceedings of Machine Learning and Systems 1, 14.

Bygstad, B., & Øvrelid, E. (2020). Architectural alignment of process innovation and digital infrastructure in a high-tech hospital. European Journal of Information Systems, 29(3), 220–237. https://doi.org/10.1080/0960085X.2020.1728201

Chen, K.-L., Lassen, A., Li, C., & Møller, C. (2022). Exploring the value of IoT data as an enabler of the transformation towards servitization: An action design research approach. European Journal of Information Systems, 1–27.

Ciborra, C., Braa, K., Cordella, A., Dahlbom, B., Hepsø, V., Failla, A., Hanseth, O., & others. (2000). From control to drift: The dynamics of corporate information infastructures. Oxford University Press on Demand.

Pozzolo, A. D., Caelen, O., Le Borgne, Y.-A., Waterschoot, S., & Bontempi, G. (2014). Learned lessons in credit card fraud detection from a practitioner perspective. Expert Systems with Applications, 41, 4915–4928. https://doi.org/10.1016/j.eswa.2014.02.026

Davenport, T. H. (1993). Process Innovation. Mcgraw-Hill.

Davenport, T. H. (2018). From analytics to artificial intelligence. Journal of Business Analytics, 1(2), 73–80. https://doi.org/10.1080/2573234X.2018.1543535

Davenport, T. H., & Miller, S. M. (2022). Working with AI: Real Stories of Human-machine Collaboration. MIT Press.

Davenport, T., & Malone, K. (2021). Deployment as a Critical Business Data Science Discipline. Harvard Data Science Review.

Dremel, C., Herterich, M. M., Wulf, J., & vom Brocke, J. (2020). Actualizing big data analytics affordances: A revelatory case study. Information & Management, 57(1), 103121. https://doi.org/10.1016/j.im.2018.10.007

Dremel, C., Herterich, M., Wulf, J., Waizmann, J.-C., & Brenner, W. (2017). How AUDI AG Established Big Data Analytics in its Digital Transformation. MIS Quarterly Executive, 16, 81–100.

Eley, T., & Lyytinen, K. (2022). Industry 4.0 Implementation: Novel Issues and Directions. Hawaii International Conference on System Sciences. https://doi.org/10.24251/HICSS.2022.622

Enholm, I. M., Papagiannidis, E., Mikalef, P., & Krogstie, J. (2021). Artificial Intelligence and Business Value: A Literature Review. Information Systems Frontiers. https://doi.org/10.1007/s10796-021-10186-w

Goldkuhl, G. (2008). What kind of pragmatism in information systems research? AIS SIG Prag Inaugural Meeting, Dec, 14, 3–8.

Gregor, S., & Hevner, A. R. (2013). Positioning and presenting design science research for maximum impact. MIS Quarterly, 337–355.

Gregor, S., Kruse, L., & Seidel, S. (2020). Research Perspectives: The Anatomy of a Design Principle. Journal of the Association for Information Systems, 21, 1622–1652. https://doi.org/10.17705/1jais.00649

Grisot, M., Hanseth, O., & Thorseng, A. (2014). Innovation Of, In, On Infrastructures: Articulating the Role of Architecture in Information Infrastructure Evolution. Journal of the Association for Information Systems, 15(4), 197–219. https://doi.org/10.17705/1jais.00357

Gröger, C. (2021). There is no AI without data. Communications of the ACM, 64(11), 98–108. https://doi.org/10.1145/3448247

Gupta, M., & George, J. F. (2016). Toward the development of a big data analytics capability. Information & Management, 53(8), 1049–1064. https://doi.org/10.1016/j.im.2016.07.004

Hanseth, O., & Lyytinen, K. (2010). Design Theory for Dynamic Complexity in Information Infrastructures: The Case of Building Internet. Journal of Information Technology, 25(1), 1–19. https://doi.org/10.1057/jit.2009.19

Henfridsson, O., & Bygstad, B. (2013). The Generative Mechanisms of Digital Infrastructure Evolution. MIS Quarterly, 37(3), 907–931.

Hevner, A., & Gregor, S. (2022). Envisioning entrepreneurship and digital innovation through a design science research lens: A matrix approach. Information & Management, 59(3), 103350. https://doi.org/10.1016/j.im.2020.103350

Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design science in information systems research. MIS Quarterly, 28(1), 75–105.

Hindle, G. A., & Vidgen, R. (2018). Developing a business analytics methodology: A case study in the foodbank sector. European Journal of Operational Research, 268(3), 836–851. https://doi.org/10.1016/j.ejor.2017.06.031

Huber, S., Wiemer, H., Schneider, D., & Ihlenfeldt, S. (2019). DMME: Data mining methodology for engineering applications – a holistic extension to the CRISP-DM model. Procedia CIRP, 79, 403–408. https://doi.org/10.1016/j.procir.2019.02.106

Iivari, J. (2015). Distinguishing and contrasting two strategies for design science research. European Journal of Information Systems, 24(1), 107–115. https://doi.org/10.1057/ejis.2013.35

Iivari, J., Rotvit Perlt Hansen, M., & Haj-Bolouri, A. (2021). A proposal for minimum reusability evaluation of design principles. European Journal of Information Systems, 30(3), 286–303.

Koutsikouri, D., Lindgren, R., Henfridsson, O., & Rudmark, D. (2018). Extending Digital Infrastructures: A Typology of Growth Tactics. Journal of the

Association for Information Systems, 1001–1019. https://doi.org/10.17705/1jais.00517

Kunz, P. C., Spohrer, K., & Heinzl, A. (2020). The Impact of Business Analytics: A Process-Level Perspective. ICIS 2020 Proceedings, 18.

Lavin, A., Gilligan-Lee, C. M., Visnjic, A., Ganju, S., Newman, D., Ganguly, S., Lange, D., Baydin, A. G., Sharma, A., Gibson, A., Zheng, S., Xing, E. P., Mattmann, C., Parr, J., & Gal, Y. (2022). Technology readiness levels for machine learning systems. Nature Communications, 13(1), Article 1. https://doi.org/10.1038/s41467-022-33128-9

Martínez-Plumed, F., Contreras-Ochando, L., Ferri, C., Hernández-Orallo, J., Kull, M., Lachiche, N., Ramírez-Quintana, M. J., & Flach, P. (2019). CRISP-DM Twenty Years Later: From Data Mining Processes to Data Science Trajectories. IEEE Transactions on Knowledge and Data Engineering, 33(8), 3048–3061. https://doi.org/10.1109/TKDE.2019.2962680

Microsoft. (2023). What is the Team Data Science Process? https://learn.microsoft.com/en-us/azure/architecture/data-science-process/overview

Mikalef, P., Pappas, I. O., Krogstie, J., & Giannakos, M. (2018). Big data analytics capabilities: A systematic literature review and research agenda. Information Systems and E-Business Management, 16(3), 547–578. https://doi.org/10.1007/s10257-017-0362-y

Nahar, N., Zhou, S., Lewis, G., & Kästner, C. (2022). Collaboration challenges in building ML-enabled systems: Communication, documentation, engineering, and process. Proceedings of the 44th International Conference on Software Engineering, 413–425. https://doi.org/10.1145/3510003.3510209

Phillips-Wren, G., Daly, M., & Burstein, F. (2021). Reconciling business intelligence, analytics and decision support systems: More data, deeper insight. Decision Support Systems, 146, 113560. https://doi.org/10.1016/j.dss.2021.113560

Piatetsky, G. (2014, October 28). CRISP-DM, still the top methodology for analytics, data mining, or data science projects. KDnuggets. Retrieved 3 August 2023, from https://www.kdnuggets.com/crisp-dm-still-the-top-methodology-for-analytics-data-mining-or-data-science-projects.html

Pozzolo, A. D., Caelen, O., Borgne, Y.-A. L., Waterschoot, S., & Bontempi, G. (2014). Learned lessons in credit card fraud detection from a practitioner perspective.

Expert Systems with Applications, 41(10), 4915–4928. https://doi.org/10.1016/j.eswa.2014.02.026

Reis, C., Ruivo, P., Oliveira, T., & Faroleiro, P. (2020). Assessing the drivers of machine learning business value. Journal of Business Research, 117, 232–243. https://doi.org/10.1016/j.jbusres.2020.05.053

Rosemann, M. (2014). Proposals for Future BPM Research Directions. In C. Ouyang & J.-Y. Jung (Eds.), Asia Pacific Business Process Management (Vol. 181, pp. 1–15). Springer International Publishing. https://doi.org/10.1007/978-3-319-08222-6_1

Ross, J. W., Beath, C. M., & Mocker, M. (2019). Designed for digital: How to architect your business for sustained success. MIT Press.

Saltz, J. (2020, November 30). CRISP-DM is Still the Most Popular Framework for Executing Data Science Projects. Data Science Process Alliance. https://www.datascience-pm.com/crisp-dm-still-most-popular/

Schmiedel, T., & vom Brocke, J. (2015). Business Process Management: Potentials and Challenges of Driving Innovation. In J. vom Brocke & T. Schmiedel (Eds.), BPM - Driving Innovation in a Digital World (pp. 3–15). Springer International Publishing. https://doi.org/10.1007/978-3-319-14430-6_1

Schröer, C., Kruse, F., & Gómez, J. M. (2021). A Systematic Literature Review on Applying CRISP-DM Process Model. Procedia Computer Science, 181, 526–534. https://doi.org/10.1016/j.procs.2021.01.199

Sculley, D., Holt, G., Golovin, D., Davydov, E., Phillips, T., Ebner, D., Chaudhary, V., Young, M., Crespo, J.-F., & Dennison, D. (2015). Hidden Technical Debt in Machine Learning Systems. Advances in Neural Information Processing Systems, 28. https://proceedings.neurips.cc/paper/2015/hash/86df7dcfd896fcaf2674f757a2463eba-Abstract.html

Sedera, D., Lokuge, S., Grover, V., Sarker, S., & Sarker, S. (2016). Innovating with enterprise systems and digital platforms: A contingent resource-based theory view. Information & Management, 53(3), 366–379. https://doi.org/10.1016/j.im.2016.01.001

Sein, M. K., Henfridsson, O., Purao, S., Rossi, M., & Lindgren, R. (2011). Action Design Research. MIS Quarterly, 35(1), 37–56. https://doi.org/10.2307/23043488

Shollo, A., Hopf, K., Thiess, T., & Müller, O. (2022). Shifting ML value creation mechanisms: A process model of ML value creation. The Journal of Strategic Information Systems, 31(3), 101734. https://doi.org/10.1016/j.jsis.2022.101734

Tarafdar, M., Beath, C. M., & Ross, J. W. (2019). Using AI to enhance business operations. MIT Sloan Management Review, 60(4).

Tercan, H., & Meisen, T. (2022). Machine learning and deep learning based predictive quality in manufacturing: A systematic review. Journal of Intelligent Manufacturing, 33(7), 1879–1905. https://doi.org/10.1007/s10845-022-01963-8

Tim, Y., Hallikainen, P., Pan, S. L., & Tamm, T. (2020). Actualizing business analytics for organizational transformation: A case study of Rovio Entertainment. European Journal of Operational Research, 281(3), 642–655. https://doi.org/10.1016/j.ejor.2018.11.074

Venable, J., Pries-Heje, J., & Baskerville, R. (2016). FEDS: A Framework for Evaluation in Design Science Research. European Journal of Information Systems, 25(1), 77–89. https://doi.org/10.1057/ejis.2014.36

Vial, G., Jiang, J., Giannelia, T., & Cameron, A.-F. (2021). The Data Problem Stalling AI. MIT Sloan Management Review, 62(2), 47–53.

Wirth, R., & Hipp, J. (2000). CRISP-DM: Towards a Standard Process Model for Data Mining. Proceedings of the 4th International Conference on the Practical Applications of Knowledge Discovery and Data Mining, 11.

Appendix A: Application of ADR Principles

Action Design Research (Sein et al., 2011) distinguishes itself from other DSR methods by its emphasis on action and emergence of the artefact through in context design. In their seminal paper, Sein et al. (2011) presented seven principles of Action Design Research that applications of the method should live up to. In Table A1 we present the seven principles together with an elaboration of how we applied the seven principles in our research.

ADR Principle	Application in our Research	Concrete Example(s)
Problem Formulation		
#1: Practice Inspired Research	Research driven by a concrete need in a large manufacturing company: approach to quickly develop and assess analytical systems aimed at operational process innovation.	The VP of Operations IT was involved in defining the overall problem and in approving each of the individual use-cases.
#2: Theory-ingrained Artefact	Use of three areas of theory: 1) design knowledge on analytics (architectures, process models), 2) digital infrastructure, and 3) IT-enabled process innovation.	Digital infrastructure was discovered as a useful theory base and was ingrained in the later development initiatives via. the developed design principles.
BIE		
#3: Reciprocal Shaping	Framing of the problem and solution spaces evolved throughout the research project because of IT artefacts' development and evaluation context.	Initially, the problem was seen as mainly technical, i.e., having the right technical infrastructure in place would allow new more iterative development approaches. Throughout the demonstrators we discovered the importance of the socio-technical architecture, such as its impact on coordination and collaboration with other teams.
#4: Mutually influential roles	Practitioners involved in all the research initiatives although in	Collaboration with Analytics Innovation

	different roles (co-designers, problem owners, experts)	Manager in the *Mould Process Monitoring* initiative as a co-designer. Data Platform team as stakeholder and problem owners in the *Data Quality Anomaly Detection.*
#5: Authentic & concurrent evaluation	Artefacts designed together with organizational actors & confronted with the organizational context throughout the research process.	Throughout the demonstrators we presented progress and artefacts in "Demo" sessions to stakeholders. Examples include standup meetings, more formal demo presentations and dissemination meetings in the end of projects.
Reflection & Learning		
#6: Guided Emergence	The approach and design principles evolved based on learning from application in the analytical initiatives in the company. Several of the design principles were not evident in the first artefacts developed.	The initial initiatives tackled use-cases with both high infrastructural and analytical complexity. The principle of starting simple, and developing quick wins emerged from challenges we observed with this approach.
Formalization of Learning		
#7: Generalized Outcomes	An approach and a set of design principles for prototyping analytical systems aimed at process innovation.	We generalized the problem to a class of problems (large established organizations looking to develop internal analytics competences for operational improvement) and extracted six design principles and a generic architecture.

Table A1: ADR Principles and their application in our research

Appendix B: Description of Demonstrators in the Exploratory Phase

In this appendix, we provide further details on the four demonstrators we conducted as part of the exploratory phase that led to the creation of the approach and design principles. For each of the four demonstrators, we structure our description in terms of 1) the problem, 2) the BIE cycle, and 3) reflection on learnings from the cycle.

Demonstrator 1: Data Quality Monitoring

Problem: The first demonstrator was carried out together with the Data Platform Team. The team had integrated several manufacturing source systems with the data platform using either batch or streaming data pipelines to support future analytics systems. Examples of data sources included manufacturing execution systems, quality information systems, warehouse execution systems, and more. Early exploratory analyses of data from the pipelines, however, showed significant problems with data quality for multiple of the data sources. Examples of data quality problems included missing data (e.g., from certain parts of the process, or data missing due to connectivity problems in certain plants), or erroneous content. Manual monitoring of data quality for the many data sources was not feasible, so instead a demonstrator was scoped with the goal of automatically detecting and alerting in case of data quality problems.

BIE: The first author worked on developing a data quality monitoring (DQM) solution to demonstrate to the Data Platform Team. The DQM problem was formulated as an instance of anomaly detection. The first BIE cycle focused on developing a solution to detect missing data for all data flowing into the platform from one of the manufacturing execution systems. As the DQM system was limited to data that was already integrated with the data platform, development of the data component consisted only of integrating the data with the anomaly detection components. For the model and system components, an off-the-shelf system with anomaly detection functionality was used.

The solution ran once per hour and evaluated all incoming data from the last hour to detect anomalies in the volume of data ingested. If an anomaly was detected, an alert was sent to a Teams channel for investigation. The first BIE cycle quickly developed a proof-of-concept of the DQM and trialed it with real data for the first major data source. As the result seemed promising to the data platform team, a decision was made to continue the work and scale the DQM to all the data sources on the platform. A second BIE cycle commenced which generalized the system developed. This included automatically provisioning the data integration, monitoring dashboards, and alerts for all data sources on the platform to prevent manual setup as in the PoC version. The solution was deployed on the data platform for a period of around two months, where it managed to catch multiple incidents with missing data to the satisfaction of the Data Platform team. During the deployment, the alerting thresholds had to be adjusted to reduce the number of false alarms. The results were not only positive, however, managing and investigating all the alarms from the Data Platform team's side proved to take up significant time. Even then, determining whether an alarm was an actual problem almost always required forwarding the issues to personnel from the data source systems, who would know whether the anomaly observed was expected or not. In the Data Platform team, we came to the realization that the DQM should be offered as a service to the source system teams, and not be operated by the Data Platform team. At the end of the two months, a decision was made to postpone further development of the DQM, as convincing the IT product teams to use the system was politically difficult at the time. While data pipelines were officially the responsibility of the IT product teams, the reality was that they lacked bandwidth to operate and own the data pipelines and hence the data platform team ended up doing it. A decision was thus made to revisit the DQM once the IT product teams had taken over data pipeline responsibility.

Reflection: The demonstrator proved how the use of the data platform and pre-existing data pipelines together with off-the-shelf systems allowed for very fast development of a prototype that could be tested with real data. The DQM solution met the requirements of both speed and scalability, however, the initial scoping of the project towards the Data Platform team as the main "customers" meant that important stakeholders in the IT product teams were not sufficiently involved and the project thus failed to attract support from them.

Demonstrator 2: Process Visualization

Problem: The second demonstrator was initiated by Innovation Managers from the Operations Transformation department. The Innovation Managers were running an R&D project together with colleagues from university and a machine vendor to develop a new and improved version of one of the production processes. The machine vendor had taken part in the development of a new open reference architecture for Industry 4.0 and the Innovation Managers wanted to test the feasibility of the architecture by deploying it to collect and visualize near real time data from the R&D process setup on the University premises, while integrating it with the organization's data platform.

BIE: The first author collaborated with the principal investigator and consultants from the vendors to deploy the data collection infrastructure from the R&D setup in accordance with the reference architecture. This included the deployment of a pre-built container from the vendor to an edge device connected to same wireless network as the R&D setup, as well as the deployment of a cloud message bus, which would receive the data collected from the edge device. Once the infrastructure was deployed, building the data pipeline consisted of two steps. The first was configuring an adapter module (using OPCUA) that was running on the edge device to read data from the machines. This data was then automatically transported to the cloud message bus. The second was to build a streaming ETL job that processed the data from the machines and prepared it for visualization, while also saving it in the data platform. In the proof-of-concept, a simple PowerBI dashboard was connected to the streaming ETL job and used to visualize the status of the production process in near-real-time. The first author demoed the system to Innovation Managers and presented on the architecture for engineers from Operations IT. The response was overwhelmingly positive, and both the IT engineers and Innovation Managers went looking for an upcoming project where they could leverage the architecture.

Reflection: The demonstrator showcased how the right architecture and infrastructure enabled quickly constructing a data pipeline to collect data from new production machinery. The fact that the modelling part was limited to simple aggregation of data (show the most recent value) and an off-the-shelf system was used meant that the development went very quickly once the data collection infrastructure was in place. Getting the data collection infrastructure in place, was, however, slightly more challenging, but we were aided by support from the consultants and the fact that we were operating in an R&D setting with little to no controls in place. Had the implementation been in the organization using operational machines, setting up the data collection infrastructure would have been subject to IT governance and required significant involvement from Operations IT, meaning that implementation would have been much more time consuming.

Demonstrator 3: ML-based Process Monitoring

Problem: The third demonstrator was a collaboration between an Innovation Manager from Operations Transformation and the first author. Part of the Industry 4.0 strategy was to move towards proactively identifying issues in the highly automated production processes by using analytics rather than relying on periodic inspections to detect problems. When we started the project, an hourly output control was used to identify potential output problems, while lower frequency periodic inspections of samples were used to control quality. In case of quality problems, significant productivity loss in terms of scrap could thus occur. The goal of the demonstrator was to test whether we could develop an analytical monitoring concept that leveraged sensor and machine data to detect quality problems and alert operators. As mentioned in the section on *Research Context*, investments had already been made in connecting machines and data collection. This project was thus supposed to be one of the first advanced analytics use-cases that leveraged the new large-scale data collection setup from the production machines.

BIE: The BIE cycle started with an exploratory data analysis of the machine and sensor data that was already available in the data platform. The analysis quickly led us to realize that significant data quality issues were present in our data. The result was a substantial amount of time spent cleaning the dataset and a substantially smaller dataset than we had imagined going into the project. Meanwhile, we also talked to various stakeholders to figure out what data to use as a signal of production quality. We identified and settled on the database of historical quality issues as registered in the quality information system (QIS). As data from the QIS was not integrated with the data platform, discovering exactly what data was available required support from QIS IT engineers. This in turn introduced some delays in our development as we waited for the IT engineers to prioritize our request. Once granted, our access was limited to manual extraction through an existing BI view. On gaining access to the data, we learned that the identifiers used to identify batches in the QIS and the production data differed and that data from the warehouse execution system would be needed to link the production data with the quality data. Obtaining access to the warehouse data required further discussions with IT and process stakeholders. Ultimately, it was decided that access would be provided by the BI team who would develop a view containing the data we needed. After waiting more than 1,5 months for data, development could finally move to cleaning and integrating data. The integrated dataset was, however, much smaller than we initially expected due to various data quality problems. We proceeded to experiment in parallel with models based on multivariate statistical process control (MSPC) and machine learning (ML) to detect quality issues based on the process data. The results were, however, disappointing. In evaluating our model's performance, we found that there was not enough data to properly validate and assess model performance with the signal available in the data. We decided to try to obtain a larger dataset by updating all our data sources with the most recent two months of data collected. Obtaining the updated datasets required further rounds of negotiation with multiple IT teams and after two months our request had finally been prioritized. Another modelling iteration commenced, but the model evaluation results were still inconclusive. While we managed to develop models that proved promising for detecting certain issues, the models developed were overall less powerful than expected and significant further work would be required to have a feasible and scalable monitoring setup. The project was stopped after this modelling cycle, and the results were disseminated to a group of stakeholders from IT management.

Reflection: The demonstrator ended up in the exact situation we were trying to avoid. Complexities related to accessing, cleaning, and preprocessing data resulted in model building being postponed significantly. This also meant that close user involvement in this part of the project was not feasible. Ultimately, the project ended before users were significantly involved. It did still, however, provide important learnings as to the feasibility of the use-case. Most importantly, the demonstrator opened our eyes to the significant role that the data infrastructure played in determining the extent of inter-team coordination required for analytics projects.

Demonstrator 4: Machine Parameter Optimization

Problem: The fourth demonstrator was a collaboration between the data team in the Process Analytics Center (PAC) and the first author. Concerned with improving quality of the same manufacturing process as in the process monitoring use-case, PAC had been conducting R&D on an alternative method to reduce scrap. Rather than detecting when quality issues might have occurred, this approach would use ML to adjust machine settings in a closed control loop to optimize quality. The closed loop control required access to more granular sensor measurements from machines, which were only available from machines in the process test center that PAC owned. Early-stage R&D in PAC had shown promising results, but the solution had been built in a non-scalable way. The goal of the demonstrator was to reimplement the ML solution in a scalable way. This would involve an integration of the manufacturing execution system (MES) responsible for machine connectivity and data collection, with the data platform and deployment of the ML system.

BIE: The demonstrator consisted of two BIE cycles conducted together with the PAC data team. In the first, several workshops and meetings were held between the PAC data team and the first author to understand their use-case and existing IT architecture. As a result of these workshops, we developed a conceptual architecture for the solution. The PAC data team wanted to implement the architecture to evaluate it in practice. The second BIE cycle concerned the implementation and was carried out together with a Data Manager from PAC as part of a two-week sprint in the PAC data team. The first step was to develop the data component, which consisted of a data pipeline that extracted the granular sensor data from the manufacturing execution system (MES) to the data platform. As the MES was API-enabled and under the control of PAC, developing the data pipeline was a relatively simple matter without any delays to gain access. The model component was likewise rebuilt to be made compatible with the data platform requiring a significant refactoring of the code. Lastly, we built the system component, which integrated the data pipeline, the ML model, and the MES. The system leveraged the streaming data pipeline, fed the data to the ML model, and implemented a rule-based control logic to update machine settings via. a call to an API in the MES. Within the two-week timespan we had successfully developed and integrated the ML system with the surrounding infrastructure. An evaluation in the form of a small proof-of-concept took place, where we deployed the system for a limited amount of time and validated that it successfully made predictions and updated settings on the machines. The demonstrator was viewed as a success by the PAC data team, as we had managed to develop and validate the architecture. Rigorous designed experiments to compare the quality of the solution to the existing process was to be carried out later, as it required acquiring capacity on both the machines to run the tests as well as the quality department, who had to thoroughly examine the output. Before investing in this, the PAC data team wanted to further improve the model, as they were aware of several shortcomings in the current model.

Reflection: The demonstrator showcased that it was possible to conduct a demonstrator focused on an automation use-case in a very short timeframe. The demonstrator included the development of a new data pipeline and the development of integrations but made use of an existing model. Implementation was thus mostly a matter of integrating components with the existing IT infrastructure, which was made easy due to the presence of APIs for the MES. Speed was further facilitated by having no dependencies outside of the PAC data team removing any potential delays due to coordination issues.

ISSN (online): 2446-1636
ISBN (online): 978-87-7573-607-2

AALBORG UNIVERSITY PRESS