The Beauty and the Beast: Patterns and Anti-Patterns in use of Data

Helena Holmström Olsson [1*] and Jan Bosch [2, 3†]

[1*]Department of Computer Science and Media Technology, Malmö University, Nordenskiöldsgatan 1, Malmö, 205 06, Sweden.
[2]Department of Computer Science and Engineering, Chalmers University of Technology, Gothenburg, Sweden.
[3]Department of Mathematics and Computer Science, Eindhoven University of Technology, Eindhoven, The Netherlands.

*Corresponding author(s). E-mail(s): helena.holmstrom.olsson@mau.se;
Contributing authors: jan.bosch@chalmers.se;
†These authors contributed equally to this work.

Abstract

Companies in the software-intensive systems industry collect vast amounts of data from their products on a continuous basis. While many companies have more data than what they will ever use, the volumes keep growing and the frequency at which data is collected is only increasing. Still, there is little guidance for how to effectively manage and make use of this data and companies struggle with how to turn raw data into customer and business value. In this paper, we study data practices in software-intensive systems companies and we identify beneficial patterns that lead to successful use of data. Also, we identify detrimental anti-patterns that companies should avoid when working with data.

Keywords: data practices, software-intensive systems, patterns, anti-patterns

1 Introduction

Despite decades of data collection from products in the field, companies in the software-intensive systems domain keep facing challenges in relation to data management and use. Although experience and expertise is maturing, the rapidly growing volumes of data, the increasing frequency at which data is collected, and the costs and complexities involved in e.g., data storage and analysis add challenges to this already difficult task. For years, researchers have studied the adoption and use of data-driven development and decision-making practices [1], [2], [3], [4], [5]. In industry, effective use of data is critical and companies across industry domains keep investing heavily in infrastructures and competencies for managing the rapidly growing amounts of data [6]. In our previous research, we studied e.g., the transition towards data-driven development [7], [2], the collection and increasing use of data [8], experimentation practices [6], [9], and how data is the key driver for digital transformation [10], [11]. In a recent paper, Olsson and Bosch explored data collection strategies [12], the use of data in digital product management [13], and the steps companies evolve through when maturing their data practices [14].

However, while there is prominent research on the adoption and use of data in the software-intensive industry [15], [16], [8], [17], [10], [2], [12], [13], [9], we see little guidance for what constitutes best practices when working with data, i.e., beneficial use patterns, and what companies should really try to avoid, i.e., anti-patterns that do more harm than good.

To address this gap, we study the use of data in software-intensive systems companies and we identify beneficial patterns and how these evolve over time. Also, we identify detrimental anti-patterns that companies should avoid. Our research builds on multi-case study research in eight companies that have collected data from their products for decades. However, if compared to the software development practices in these companies, their data practices are not as well-established and still maturing.

In our previous research [18], we identified the key dimensions that help companies reason about cost versus value of collecting data. As the key contribution, we present a data framework in which the dimensions that companies need to consider when working with data are outlined. In this paper, the data framework as presented in our previous work [18] provides the foundation for identifying patterns and anti-patterns that we see in the case companies we studied.

The contribution of this paper is two-fold. First, we identify beneficial patterns and their evolution over time. Second, we identify detrimental anti-patterns that companies should avoid.

The remainder of this paper is structured as follows. In section 2, we present the background for the study. In section 3, we describe the research method we employed during our research. In section 4, we report on our empirical findings. In section 5, we identify the patterns and anti-patterns. In section 6, we discuss threats to validity and in section 7 we conclude the paper and outline future research.

2 Background

2.1 Data practices in the software-intensive systems domain

Continuous delivery of customer value is a key priority for companies in the software-intensive systems domain [19], [6]. As recognized in [20], the ability to deploy software on a continuous basis provides a unique opportunity to learn how systems perform in the field and how they are used by customers. To achieve this, companies adopt data-driven development practices where data is used as the basis for decisions on what to develop, what features to prioritize and how to improve and optimize existing products. For successful adoption and use of data-driven practices, several factors are recognized as important.

First, companies need to short feedback cycles [7], [21], [22]. In practice, this means adopting DevOps practices to ensure rapid and reliable deployment of new functionality [23], [24]. Second, companies need metrics to monitor feature usage, user behaviors and overall system performance. Finally, companies need to adopt techniques such as A/B testing as the basis for evaluating hypotheses and validate what adds value to customers [6]. In literature, companies like Google, Facebook, Amazon, Netflix, Microsoft and Booking.com are often referred to as companies that run extensive A/B tests in their products. However, recent years have shown an increasing uptake of these practices also in the software-intensive systems domain. Several previous studies explored how data-driven practices and experimentation are increasingly adopted and how they provide teams with tools and techniques to help improve decision-making, prioritization efforts and develop software functionality that better meets the needs of customers, e.g.,[17], [25], [26][7], [2], [27], [28], [29], [11].

Although data collection is far from a new phenomenon in the software-intensive systems domain we see that the role and the implications of data are changing. While it was used primarily for diagnostics and trouble shooting, and as an effective mechanism for quality assurance, we see that data is increasingly used as the basis for product innovations, new business models and digital services. Previous research reports on some of the changes we see in how data is used and the increasing importance of data as an asset [29], [9], [6], [25]. Also, it is recognized how data is the backbone for machine- and deep learning technologies [5] as well as for federated and reinforcement learning [30], [31].

2.2 Data framework

In our previous research [18], we developed a data framework with the intention to capture the multiple dimensions that are involved in collection, processing, storing and use of data. The framework, as shown in figure 1, provides a holistic understanding of these dimensions and helps companies when reasoning about business value versus cost of data. In this paper, the data framework as presented in [18] lays the basis for studying and identifying beneficial patterns that we see companies use and how these evolve over time. Also, we use the framework to identify detrimental anti-patterns that companies should avoid when working with data.

The framework consists of an *offering part* and a *data part.* The offering part consists of a business case and a use case. The business case describes the nature of the offering and can have different levels of maturity, ranging from a hypothesis to a validated business case. The use case describes an instance of realizing the business case. The separation between business and use case is important as a business case can often be realized in different use cases. The use case has two dimensions, i.e. the scope and the lifecycle. The scope is concerned with the beneficiaries of the use case ranging from an individual team to the entire business ecosystem. The lifecycle of a use case typically has a strong relationship with the maturity of the business case and

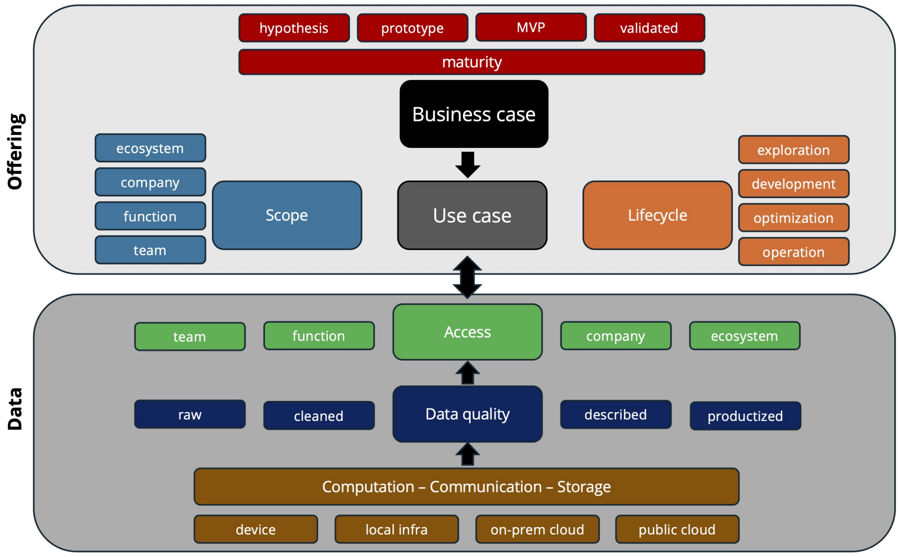

Fig. 1 Data framework as presented in [18].

ranges from exploration to operation [18]. The data part of the framework includes three dimensions, i.e. compute, communication and storage (CCS), data quality and access [18]. The CCS is concerned with where data is processed, stored and communicated. The data quality tends to evolve with the maturity of the use case and the scope of use of the data. As outlined in the data quality dimension, data can be raw, cleaned, described or productized. Productized data is data sets or streams that can be treated (and monetized) as a product. This means the data is guaranteed to be available, cleaned and clearly described and can be used to build commercial data-driven services on top of. There are dependencies between the dimensions we identify and our assessment is that most challenges that companies experience are caused by misalignment between these dimensions.

3 Research method

3.1 Case study research

The objective of this paper is to identify beneficial use patterns when working with data, and to identify detrimental anti-patterns that do more harm than good and that companies should avoid. To study this, we used the data framework as presented in [18] as the basis when exploring how companies work with data.

We employed a case study research approach when studying eight companies in the software-intensive systems domain. Case study research has become a well-established and appreciated method in software engineering research as it allows for exploring complex phenomena in real-life settings where the boundaries between the phenomenon itself and its surrounding context are unclear[32], [33].

3.2 Case companies

Below, we provide a short description of the case companies. The case companies are members of Software Center (www.software-center.se) and they are all operating in the software-intensive systems domain.

- ***Case company A*** is a company manufacturing trucks, buses and construction equipment as well as a supplier of marine systems.
- ***Case company B*** is a food packaging and processing company.
- ***Case company C*** is a company manufacturing network cameras, access control, and network audio devices for the security and surveillance industries.
- ***Case company D*** is a company developing automation and digitalization solutions for process and manufacturing industries.
- ***Case company E*** is a networking and telecommunications company.
- ***Case company F*** is a company developing advanced driver assistance systems and autonomous drive technology.
- ***Case company G*** is a company manufacturing pumps.
- ***Case company H*** is a vehicle manufacturer.

3.3 Data collection and analysis

In this paper, we report on research conducted between September 2023 - March 2025. Throughout this period, data collection was conducted using company workshops and interviews. We engaged in 2-4 hour workshop sessions at company A, B, C and H and in parallel we conducted an interview study with key stakeholders in all eight case companies.

In *company A*, we organized four workshops with the first exploratory workshop on September 12th, 2023, and the last workshop where we summarized discussions on October 11th, 2024. In between, we met on Novenber 14th, 2023 and February 27th, 2024 to discuss data infrastructures, data management and the ways-of-working with regards to collection, analysis and use of data as the basis for new services. Our workshops involved 6 - 15 people representing IT, data and simulation platforms, data science, electro mobility, and architecture. In *company B*, we organized one workshop on March 20th, 2024. At this workshop, we discussed challenges and opportunities associated with data collection and use. This workshop involved a group of 10 people representing e.g., data management, connectivity, architecture, decision science, and automation. In *company C*, we organized one workshop on December 5th, 2024 and the second workshop on May 8th, 2025. In the first workshop we met with people representing architecture, platform development and engineering management and learned about the ways in which data is being used for new services. In the second workshop, we met with a data science team to discuss the efforts required in data labeling, access and governance. In *company H*, we organized one workshop on January

22nd, 2025. At this workshop, we met with three people in manager roles to discuss data management practices and the potential of data products.

In addition to these workshops, we organized one *cross-company workshop* to which we invited all case companies. The workshop was organized on November 15th, 2024. During this workshop, companies shared experiences, discussed challenges and explored best practices. In total, we organized eight workshops and we met with more than 60 people.

In parallel with the workshops, we conducted an *interview study* that involved key representatives from the case companies. The interview study was initiated on October 31st, 2024 and completed on March 7th, 2025. All interviews were conducted online (except one which was conducted in-person) using Microsoft Teams. In total, we interviewed 17 people (6 people from company A, 5 people from company B, 2 people from company C, 1 person from company D, E, F, and G). All interviews lasted for one hour and followed an interview guideline of a semi-structured format. During the interviews, both researchers shared the responsibility of asking questions and encouraging the interviewee to elaborate on the answers by asking follow-up questions.

During data analysis, we adopted an interpretive approach [34]. Following this approach, we revisited the documentation from all workshop discussions to carefully reflect on our insights from these and what implications could be drawn from the empirical material. For the interviews, one of the researchers went through all transcriptions, summarized these and based on the summaries, both researchers discussed the content to identify common themes and patterns among the interviewees.

4 Empirical findings

Below, we present a summary of how data is collected and used within the case companies. Second, we outline some of the challenges the company representatives experience.

4.1 Collection and use of large volumes of data

All case companies collect data from their products to monitor operational performance such as e.g., downtime, efficiency, utilization, and to avoid breakdown of parts. Company A collects data from test vehicles, from production vehicles, logged data, fleet management data and field test data. In addition, road condition data, weather data and data that reflects the different variants of trucks is collected. In company B, data is collected from the filling machines to help identify stops, the reason for these stops and for analyzing equipment performance. This is reflected upon by one of the interviewees when saying: *"Data is low intensity data e.g., events, status, counters, stop reasons and whatever is needed for monitoring operational performance."* Similarly, another interviewee in company B shared that: *"There are a lot of sensors in the machines that provide us with information about the quality in the machine and the testing patterns that we are doing."* In addition, data is collected for purposes such as simulations, inventory optimization and for understanding the price elasticity of products. One of the interviewees recognized the fact that their data volumes will only grow: *"In the future, we are going to need more capacity for dealing with big data and* *we will need some kind of model to manage a continuous stream of data."* Similarly, another interviewee from the company noted: *"In the pipeline is the new technology where we make use of more intense data sets connected to the usage of AI- embedded applications, e.g., usage of vision, images, streaming of video to understand crashes, creation of digital twins."*

Company C collects video and sensor data to enable diagnostics and optimization of configurations. The data is compressed and is processed only in aggregated and anonymized form. The two interviewees from company C shared how the company is experiencing a transition in which customers are increasingly opening up and allowing for data to be collected from the products. According to the interviewees, this is mainly due to the many device management services they offer and that customers are starting to realize the many benefits of sharing data.

In company D, tracking data, monitoring data, image data and logistics data is collected from thousands of devices to optimize performance of factories. Typically, the company doesn't store data at its own premises but instead lets its customer own and store the data. In some cases however, data from customer sites is transferred via connectors to allow for identification issues at customer sites. In such cases however, the interviewee report on several challenges related to e.g., transferring costs and data privacy.

Company E and F collect large amounts of data and in the interview with a person in company F we learnt that people are very reluctant to getting rid of data. Although the company has limited data storage capabilities the decision on what to keep and what to not keep is a difficult one that people tend to avoid. This is something we noted also in other companies when people often referred to "just-in-case" collection of data i.e., collect everything you can in case there will be a use case for it in the future. In company G, data is collected to monitor performance of products and to allow remote maintenance. In the interview with a person from this company, it became clear that data is only increasing in value and that the products are increasingly viewed as data collection devices.

All case companies store data both in the public cloud and on-premise. This is reflected upon by one of the interviewees when saying: *"The ideal is to have a hybrid concept where we develop services with our on-prem capacity, then scale out to public cloud for computation needs."* Our workshop discussions revealed that compute costs are expensive and in particular, the costs of downloading data. In company B, the interviewees shared that: *"We see a tendency that more and more data ends up in a cloud application."* For now however, the majority of data is kept on the machines at customer sites. One of the interviewees reflects on this when sharing that: *"We have limited storage and limited processing on the machine itself. But there's been a strategy decision taken that we want to move data as soon as possible to other systems on-prem."* With regards to costs, this interviewee concludes that: *"It's basically down to the fact of whether or not we are willing to pay the price for storing everything "just in case" we need it. It's a matter of strategy and how we foresee the future."*

4.2 Challenges when working with data

The primary challenge that people report on is data quality. All case company representatives shared concerns with regard to quality of the data that is used as the basis for development, decisions, innovation initiatives and new services. In company F, the interviewee shared that: *"We collect and keep data without checking quality. We know that if, for example, all camera sensors are not activated the data that is collected will be useless. You need complete sensor data to be able to use it and if one sensor is not activated the entire data set is useless."* Similarly, discussions in some of the workshops in company A revealed that there are data-driven services deployed and in use at customers, but where the data that provides the foundation for the service is not productized. This challenge is touched upon also in the interview with company G in which the interviewee discussed that clean, described, and productized data is always use-case specific: *"There is no such thing as just "cleaned data" as cleaning depends on what you want to use it for."* Related to this, and with the cost perspective in mind, one of the interviewees in company A raised the risk of having low quality data, or lack of competence for analysis, as the basis for decision-making: *"It is also the cost of incorrect conclusions because of wrong analysis, or analytics on incorrect data, or simply misinterpreted data because of unclear semantics of the data."*

A second challenge is the efforts that are spent on accessing data. This is reflected in the following quote from company A: *"Many of the use cases require pulling data from a variety of different sources, and from a variety of different platforms. In general, 30% of our time is spent on getting access to the data we need. After this, another 15 – 30% of the time is spent on cleaning up the data so that we can actually use it."* In company B, one of the interviewees describes a situation in which even more time is spent on this when saying: *"In many cases, 80% of the job is spent on getting access to the data."* Also, several people describe a situation in which lots of efforts are spent on making data available for others. As soon as data is shared across teams, the interviewees report on a situation in which the team that originally collected the data suddenly becomes a "data provider" to other teams using that data. This implies maintaining quality of the data, ensuring the trustworthiness of the data and providing clear semantics to avoid misinterpretations by people who were not involved in the context in which the data was initially generated. As mentioned by the interviewee in company G, problems arise when teams access data but lack domain knowledge. *"Some teams only look at data as data, but have no domain knowledge of the underlying data generator. Therefore, they lack the ability to see if there is a problem in the data."*

During our study we noted that there is an on-going discussion in all companies on how to manage data access. One of the interviewees in company A reflects on the risk of having people access what is irrelevant for them when saying: *"Data on deep analytics on energy efficiency and driver behavior, is that something that everyone in the company should have access to? Would they even be able to use it? No."* However, there are also situations in which accessibility across teams and groups is critical. As the most common example, the company representatives mention development of new services as these often involve functions from different parts of the organization as well as data collected by different groups.

Third, we see several problems in relation to storage costs. Especially, companies experience challenges with cloud storage and how to reason about when and what data to store in the cloud. A common pattern is that data is kept in the cloud and on-premise depending on the data type and the purpose for which it is stored. Typically, data is processed on-premise using local infrastructures as these are more attractive from a cost perspective. In one of the interviews, the interviewee reflected on the challenges related to the balance of on-device storage and cloud storage and identified the cost of transferring data and data privacy as the two major ones: *"Compute and data download are expensive tasks and these are the major costs. When you want to crunch the data, that's when the money you pay starts to to ramp up."*

5 Data practices: Patterns and Anti-patterns

5.1 Patterns

Our study revealed several beneficial patterns that lead to successful use of data. Below, we present three patterns that we identified in the case companies.

Exploration of new data-driven services

Description: The companies in our study are collecting vast amounts of data, but that does not automatically convert in new data-driven services. So, business developers and data scientists often work together to explore ideas that may result in new data driven services.
Considerations: As many of the ideas that are explored are likely not successful and will not result in new services, the focus should be at exploring as many ideas as possible at the lowest possible cost per idea. One aspect of this is that often large amounts of data are explored and when conducting this in a public cloud context, the cost can easily become prohibitively high. In our experience, exploration is often best conducted on internal compute infrastructure.
Variants: In addition to internal exploration of data-driven services, companies often have partners that are interested in joint innovation efforts around data. This can concern subsystems delivered by suppliers or partners seeking to combine their own data and the companies data to generate new insights.

Operational data-driven service

Description: A second pattern, at the other end of the lifecycle, are data-services that have been confirmed to deliver value to customers and that are in operation. These services are typically monetized and need to be provided in a reliable and accurate fashion.
Considerations: This is where companies spend significant resources on ensuring data quality as this forms the basis for service quality. Also, in this case, companies tend to operate on top of a public cloud infrastructure to assure reliability.
Variants: Especially in embedded systems companies, some data-driven services can be provided directly by the device used by the customer, e.g. a truck. In this case, no

centralized deployment is required, but data quality and other concerns need to be managed in the device itself in that case.

Collaborative data-driven service

Description: Many of the companies in our study closely collaborate with a set of partners and suppliers. In addition to working on the functionality in the resultant product, there is a natural extension towards collaboration around data as well.
Considerations: As the companies in our study often own the end-customer relationship, this collaboration is often driven by the partners and suppliers having to offer something of value in order to gain access to data concerning the subsystems or solutions that they provide and the context in which these operate.
Variants: In some cases, the companies in our study have even agreed to providing data to suppliers that they themselves do not even have access to, e.g. through encryption. Also, there are cases where a third party collects data directly from devices on customer premises without the companies in our study being involved at all in the data collection nor the associated data-driven services. In this case, many companies are looking for control points to avoid this.

5.2 Anti-patterns

Although the data framework as presented in [18] is helpful when modeling the intended ways of working with data, we have found that the framework is as helpful for identifying anti-patterns. Below, we share three of the most common ones. We number them 1, 2 and 3 and visualize them in figure 2. All anti-patterns reflect one or several misalignments in the data dimensions we identified in our previous work [18], i.e., data quality, data access, computation – communication – storage, scope and lifecycle.

Raw data for productized services (1)

Description: Many services are initially developed with one or a few customers in a prototype fashion. At that stage, the quality of the data and the resultant service is not prioritized. When the service proves to be successful, it is rapidly scaled and commercially deployed and ensuring data quality may easily become deprioritized.
Consequences: As a commercial, customer-facing service is based on raw data that is not cleaned nor described, the service may easily present data that is entirely inconsistent and irrelevant. This easily decreases the value that customers associate with the service.
Mitigation: The mitigation of this anti-pattern is obvious: the quality of the data needs to be improved by putting it through a data pipeline that removes incorrect records, completes incomplete records and provides consistency checking over a window of data in order to ensure that the service always provides relevant and complete information.

Ungoverned data access (2)

Description: Several interviewees described cases where data access was ungoverned, meaning that data collected by and intended for a specific team was used at a company-level and by teams and functions for which it was never intended.
Consequences: Teams that collect data for their own purposes can easily stop collecting data when they don't need it anymore or to change the semantics or structure of the data as their needs change. If others are dependent on that data and the original team is not aware of this, it causes failures and downtime. If the team is aware of it, it may not be interested in becoming a service provider for others in the company as this requires significant effort.
Mitigation: In general, we see companies move towards a binary solution. Either data is provided as a data product and its quality, semantics and availability are guaranteed. Alternatively, data is only available by the party that initiated its collection and any deviations to the rule are only operationalized in mutual agreement.

Excessive cloud cost (3)

Description: We were exposed to a number of cases where the company received excessive cloud computing bills that took them by surprise. For instance, when conducting exploratory analysis of large amounts of data in a public cloud setting, the cost for computation can easily become quite large quite quickly.
Consequences: The first, obvious consequence is high cost that outweigh the potential value of use cases that are explored or under development. The indirect consequence is that engineers and data scientists become reluctant to explore services and use cases due to the high associated cost. This is obviously not desirable as all of our case study companies prioritized developing and monetizing data-driven services.
Mitigation: The primary mitigation strategy is to ensure access to two compute environments. An internal cloud that is purchased outright by the company and where there are no usage-based cost. Here engineers and data scientists can run any exploratory use case that they want without worrying about cost. Secondly, a public cloud where commercially deployed services are hosted and where up-time and response times are critical and can be guaranteed by cloud providers.

6 Threats to validity

The validity of a study implies the trustworthiness of the results, which is divided into construct, internal, external, and reliability [33]. To address construct validity, we made sure to define key concepts and terminology in all our interactions with company representatives. In this way, we mitigated potential misunderstandings and we invited for a discussion around the topic and terminology of the study. To address internal validity, we engaged with key stakeholders and experts who were able to provide an understanding of the complexity of the phenomenon we studied, i.e., people with responsibility for data collection, data processing and analysis, data storage and service development. To address external validity, we used our empirical cases to inductively

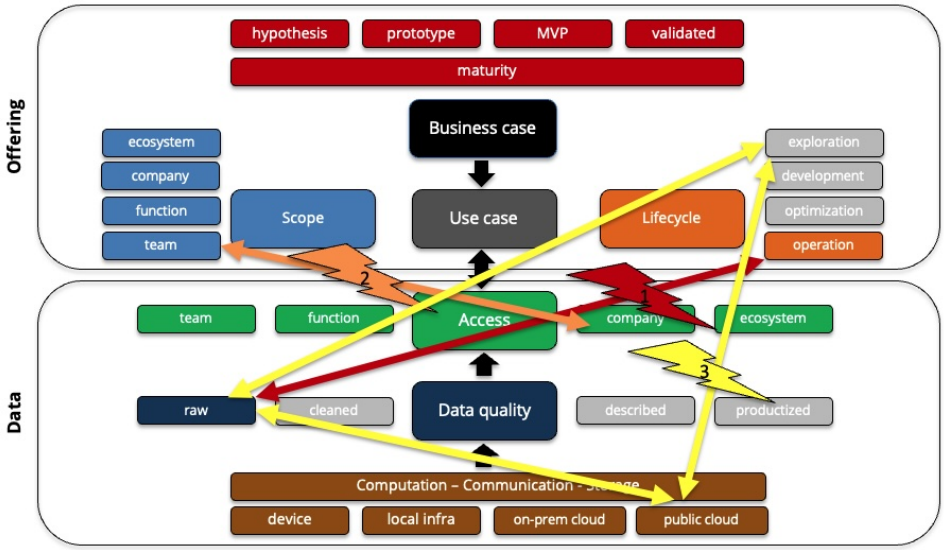

Fig. 2 Anti-patterns identified in the case companies: (1) Raw data for productized services, (2) Ungoverned data access and (3) Excessive cloud cost.

derive our findings with the intention to provide value for companies that have common characteristics as the companies we studied. Finally, to address reliability, we applied established methods and practices for data collection and analysis and we made sure our results were frequently reviewed and approved by company representatives.

7 Conclusions and future work

Effective use of data is key for companies in the software-intensive systems domain. For these companies, data from products in the field lays the basis for diagnostics, product improvements, optimization efforts, innovation initiatives and new data and AI -driven services. However, while there is prominent research on the adoption and use of data in the software-intensive industry, we see little guidance for what constitutes best practices when working with data.

In this paper, we study data practices in software-intensive systems companies and we identify beneficial patterns that lead to successful use of data. Also, we identify detrimental anti-patterns that companies should avoid when working with data. Our research builds on multi-case study research and shows that despite years of data collection, data practices are still maturing. In future research, we aim to further explore and validate the patterns and anti-patterns we identified. In addition, we seek to broaden our research to involve companies in other domians.

Acknowledgements. We would like to thank the Software Center company representatives that participated in this study.

References

[1] Svensson, R.B., Feldt, R., Torkar, R.: The unfulfilled potential of data-driven decision making in agile software development. In: Agile Processes in Software Engineering and Extreme Programming: 20th International Conference, XP 2019, Montréal, QC, Canada, May 21–25, 2019, Proceedings 20, pp. 69–85 (2019). Springer

[2] Olsson, H.H., Bosch, J.: The hypex model: from opinions to data-driven software development. In: Continuous Software Engineering, pp. 155–164. Springer, ??? (2014)

[3] Maalej, W., Nayebi, M., Johann, T., Ruhe, G.: Toward data-driven requirements engineering. IEEE software **33**(1), 48–54 (2015)

[4] Munappy, A.R., Mattos, D.I., Bosch, J., Olsson, H.H., Dakkak, A.: From ad-hoc data analytics to dataops. In: Proceedings of the International Conference on Software and System Processes, pp. 165–174 (2020)

[5] Munappy, A.R., Bosch, J., Olsson, H.H., Arpteg, A., Brinne, B.: Data management for production quality deep learning models: Challenges and solutions. Journal of Systems and Software **191**, 111359 (2022)

[6] Fabijan, A., Dmitriev, P., Olsson, H.H., Bosch, J.: The evolution of continuous experimentation in software product development: from data to a data-driven organization at scale. In: 2017 IEEE/ACM 39th International Conference on Software Engineering (ICSE), pp. 770–780 (2017). IEEE

[7] Olsson, H.H., Alahyari, H., Bosch, J.: Climbing the" stairway to heaven"–a mulitiple-case study exploring barriers in the transition from agile development towards continuous deployment of software. In: 2012 38th Euromicro Conference on Software Engineering and Advanced Applications, pp. 392–399 (2012). IEEE

[8] Olsson, H.H., Bosch, J.: Towards continuous validation of customer value. In: Scientific Workshop Proceedings of the XP2015, pp. 1–4 (2015)

[9] Fabijan, A., Dmitriev, P., McFarland, C., Vermeer, L., Holmström Olsson, H., Bosch, J.: Experimentation growth: Evolving trustworthy a/b testing capabilities in online software companies. Journal of Software: Evolution and Process **30**(12), 2113 (2018)

[10] Bosch, J., Olsson, H.H.: Digital for real: A multicase study on the digital transformation of companies in the embedded systems domain. Journal of Software: Evolution and Process **33**(5), 2333 (2021)

[11] Olsson, H.H., Bosch, J.: Living in a pink cloud or fighting a whack-a-mole? on the creation of recurring revenue streams in the embedded systems domain. In: 2022

48th Euromicro Conference on Software Engineering and Advanced Applications (SEAA), pp. 161–168 (2022). IEEE

[12] Olsson, H.H., Bosch, J.: All data is equal or is some data more equal? on strategic data collection and use in the embedded systems domain. In: 2023 49th Euromicro Conference on Software Engineering and Advanced Applications (SEAA), pp. 319–327 (2023). IEEE

[13] Olsson, H.H., Bosch, J.: Strategic digital product management in the age of ai. In: International Conference on Software Business, pp. 344–359 (2023). Springer

[14] Olsson, H.H., Bosch, J.: How to get good at data: 5 steps. In: 7th International Workshop on Software-intensive Business, p. (2024). ACM

[15] Berndtsson, M., Forsberg, D., Stein, D., Svahn, T.: Becoming a data-driven organisation. In: 26th European Conference on Information Systems (ECIS2018), Portsmouth, United Kingdom, June 23-28, 2018 (2018)

[16] Kiron, D.: Lessons from becoming a data-driven organization. MIT Sloan Management Review **58**(2) (2017)

[17] Mattos, D.I., Dakkak, A., Bosch, J., Olsson, H.H.: Experimentation for business-to-business mission-critical systems: A case study. In: Proceedings of the International Conference on Software and System Processes, pp. 95–104 (2020)

[18] Olsson, H.H., Bosch, J.: Dealing with data: Bringing order to chaos. In: 2024 50th Euromicro Conference on Software Engineering and Advanced Applications (SEAA), pp. 350–355 (2024). IEEE

[19] Fagerholm, F., Guinea, A.S., Mäenpää, H., Münch, J.: The right model for continuous experimentation. Journal of Systems and Software **123**, 292–305 (2017)

[20] Dakkak, A., Mattos, D.I., Bosch, J.: Success factors when transitioning to continuous deployment in software-intensive embedded systems. In: 2021 47th Euromicro Conference on Software Engineering and Advanced Applications (SEAA), pp. 1–9 (2021). IEEE

[21] Bosch, J.: Speed, data, and ecosystems: the future of software engineering. IEEE Software **33**(1), 82–88 (2015)

[22] Dakkak, A., Bosch, J., Olsson, H.H., Mattos, D.I.: Continuous deployment in software-intensive system-of-systems. Information and Software Technology **159**, 107200 (2023)

[23] Zhu, L., Bass, L., Champlin-Scharff, G.: Devops and its practices. IEEE Software **33**(3), 32–34 (2016)

[24] Amaro, R., Pereira, R., Silva, M.M.: Capabilities and practices in devops: a multivocal literature review. IEEE Transactions on Software Engineering **49**(2), 883–901 (2022)

[25] Liu, Y., Mattos, D.I., Bosch, J., Olsson, H.H., Lantz, J.: Size matters? or not: A/b testing with limited sample in automotive embedded software. In: 2021 47th Euromicro Conference on Software Engineering and Advanced Applications (SEAA), pp. 300–307 (2021). IEEE

[26] Mattos, D.I., Bosch, J., Olsson, H.H.: Your system gets better every day you use it: towards automated continuous experimentation. In: 2017 43rd Euromicro Conference on Software Engineering and Advanced Applications (SEAA), pp. 256–265 (2017). IEEE

[27] Fabijan, A., Olsson, H.H., Bosch, J.: Data-driven decision-making in product r&d. (2015). Springer

[28] Fabijan, A., Olsson, H.H., Bosch, J.: Early value argumentation and prediction: an iterative approach to quantifying feature value. In: Product-Focused Software Process Improvement: 16th International Conference, PROFES 2015, Bolzano, Italy, December 2-4, 2015, Proceedings 16, pp. 16–23 (2015). Springer

[29] Olsson, H.H., Bosch, J.: Going digital: Disruption and transformation in software-intensive embedded systems ecosystems. Journal of Software: Evolution and Process **32**(6), 2249 (2020)

[30] Zhang, H., Li, J., Qi, Z., Lin, X., Aronsson, A., Bosch, J., Olsson, H.H.: 5g network on wings: A deep reinforcement learning approach to uav-based integrated access and backhaul. arXiv preprint arXiv:2202.02006 (2022)

[31] Zhang, H., Bosch, J., Olsson, H.H., Koppisetty, A.C.: Af-dndf: Asynchronous federated learning of deep neural decision forests. In: 2021 47th Euromicro Conference on Software Engineering and Advanced Applications (SEAA), pp. 308–315 (2021). IEEE

[32] Verner, J.M., Sampson, J., Tosic, V., Bakar, N.A., Kitchenham, B.A.: Guidelines for industrially-based multiple case studies in software engineering. In: 2009 Third International Conference on Research Challenges in Information Science, pp. 313–324 (2009). IEEE

[33] Runeson, P., Höst, M.: Guidelines for conducting and reporting case study research in software engineering. Empirical software engineering **14**(2), 131–164 (2009)

[34] Walsham, G.: Interpretive case studies in is research: nature and method. European Journal of information systems **4**(2), 74–81 (1995)