






Data-Driven Development in Public Sector: How Agile Product Teams Maneuver Data Privacy Regulations

Astri Barbala¹ , Tor Sporsem¹ , and Viktoria Stray^{1,2} 

¹ SINTEF Digital, 7034 Trondheim, Norway

{astri.barbala, tor.sporsem, viktorija.stray}@sintef.no

² University of Oslo, 0373 Oslo, Norway

Abstract. Datafication processes, the ongoing strive for making organizations data-driven, have in recent years entailed data-focused software projects and more interdisciplinary teamwork. Simultaneously as agile product teams have been directed towards increased use of data for software development, stronger data protection regulations such as GDPR have further complexified the software developer role, whose responsibilities and expectations now expand far beyond mere coding. Seeking to develop an understanding of how data-intensive product teams in the public sector maneuver the legal hurdles emerging in the wake of data governance, this paper builds on 19 interviews with members of two agile product teams in the Norwegian organizations NAV and Entur. Our findings indicate that including a legal expert in the team can boost confidence in data handling practices and avoid delays in deliveries, but it requires effort to synchronize and overcome interdisciplinary barriers.

Keywords: Datafication · Data Protection · Data Management Regulations · GDPR Compliance · Agile Software Development · Collaboration · Coordination · Interdisciplinary Teams

1 Introduction

Over the last decade, European governments have looked to data-driven technology to enable “the strategic use of data for productive, inclusive and trustworthy governance” [1], meaning that data collection, storage, and sharing are playing increasingly central roles in public administration. The quest for a datafied public sector is largely driven by a vision of both empowered citizens and better and more effective public services, entailing reliable and secure software solutions built with agile methods. However, research has shown that there are several multifaceted constraints involved in becoming data-driven, both technical, organizational, political, and legal (see, e.g., [2–4]).

For software developers, the direct implication of datafication is that product teams must own and manage data as part of their everyday practice, resulting from data mesh structures and the move towards decentralized data storage [5] and democratization of data [6]. Simultaneously, as data science has rapidly made its way into software

development, agile public sector teams are expected to maneuver new team constellations and often ambiguous data management regulations.

A central issue for public sector organizations in the quest of becoming data-driven entails overcoming juridical hurdles effectively and in accordance with applicable laws, specifically to protect citizens' private data. This paper seeks to gain an understanding of how data-intensive public sector product teams can overcome these hurdles and how scenarios can play out differently depending on the team composition; more specifically whether the team includes a legal expert or not. It is well-established that the barriers to the effectiveness of agile teams lie not only at the team level and leadership of the team but also in the organizational and environmental contexts that directly affect team success [7]. However, few empirical studies have to date zoomed in on how data privacy laws affect the day-to-day work of software developers and the importance of team composition and organization in that regard.

To address this gap in the literature, we studied two agile product teams in two central public sector organizations in Norway, NAV and Entur, that were transitioning from centralized to distributed data management. "Product teams" are autonomous teams with the cross-functional competence needed to provide the service or product independently¹. The two agile teams develop data-intensive products, meaning utilizing user data in the software development process, and both software developers and team members from other disciplines were interviewed. We chose to explore the following research questions:

- 1) How are data privacy regulations affecting the day-to-day work of product teams?
- 2) What are the pros and cons of including a legal advisor as part of the team to overcome juridical hurdles?

This paper is organized as follows: In Sect. 2, we introduce public sector datafication and what it entails in the context of agile software engineering. We then explain the increasing importance of juridical understandings for agile product teams, underlining the direct impact data privacy regulations have on present-day software development. Section 3 describes the context and methods before we present our findings in Sect. 4. In Sect. 5, we discuss the main data privacy challenges for agile product teams and discuss how these obstacles can be addressed moving forward. Section 6 concludes and suggests future work.

2 Background

2.1 Defining Public Sector Datafication: Towards Decentralized Data Management

Data has been viewed as the new oil, entailing that enormous value can be extracted from refined data and that just like oil spills, "data spills" can cause tremendous damage [8]. Within software engineering, datafication has also been regarded as an emerging

¹ Both companies are heavily inspired by the Team Topologies book and use "product teams" as synonyms to "stream-aligned teams". See Skelton, M., & Pais, M. (2019). *Team topologies: organizing business and technology teams for fast flow*. It Revolution.

“trending topic” and an aspect of the field’s increasing focus on data [9–11]. Datafication can be defined as “the transformation of social action into online quantified data” [12], and entails “the collective tools, technologies, and processes used to transform an organization to a data-driven enterprise” [13].

Although much research has been conducted on public sector datafication in the last few years in terms of policy implications and citizen rights and participation [4, 14–16], there has to date been a lack of scholarly interest within the software engineering field for discussing how datafication has affected software projects and the software developer role. Systematic data about populations have been collected in the Nordic countries for a long time, partly due to the development of welfare infrastructures [17]. Technology researcher Jathan Sadowski [18] has stated that “data – and the accumulation of data – is a core component of political economy in the 21st century”. Data is thus a potential source of citizen insight and enormous capital and entails that public sector product teams can be defined as particularly data-intensive [19].

In developing digital public services, such as online applications for filing tax returns and unemployment benefits, the Norwegian public sector has increasingly turned towards distributed data management models that are arguably more compatible with modern-day agile software development [5]. Data mesh is defined by Zhamak Dehghani [20, 21] as fulfilling four principles; namely 1) domain-oriented decentralized data ownership and architecture, 2) seeing data as a product, 3) incorporating a self-serve data platform and 4) a turn towards federated computational governance. The direct implication of data mesh for software practitioners is that product teams gain more control and ownership of the data deriving from the team’s designated application, in addition to increased responsibility to share data with other teams. NAV and Entur have been public sector pioneers in turning towards data mesh incorporation, which incentivized us to choose teams from these organizations as our focus points for this study.

2.2 The Legal Aspect of Agile Teams: An Increasing Importance

In recent years, with the advent of data privacy regulations governing how best to collect, store and utilize the value of data, maneuvering the oftentimes ambiguous and little context-specific data management rules has become an increasing challenge for both public and private sector software teams. User-centered laws such as the General Data Protection Regulation (GDPR) launched by the European Union in 2018 have directly affected European software development planning and implementation. As a member of the European Economic Area (EEA), Norway is bound to the GDPR. Additionally, Norwegian law has national adaptations that are tailored more directly to the country [22].

The new and stronger rules on data protection entail that people have more control over their personal data, as well as giving businesses and organizations several obligations in terms of carrying out data protection assessments and maintaining records. These new responsibilities have led software development teams to look outside IT-related professions to fill the emerging roles to become self-organizing [23]. Integrating experts from other disciplines into IT teams to strengthen their agility and interdisciplinary competence is nothing new. However, it has become increasingly important in recent years with more and more complex software projects and team compositions further complicated by data science roles [24]. In addition, the turn to continuous software

deployment means that teams are increasingly expected to deliver faster and faster, going from a couple of releases a year to several a day.

Ensuring compliance with relevant regulations is a vital component of the continuous integration strategy. As pointed out by Klotins et al. [25], regulatory practices are often at odds with agile and continuous principles. Previous research has shown that adhering to GDPR is burdensome for software developers due to both lack of support from institutions and clients, and inadequate online tools [26]. Although continuous compliance is an established procedure within agile literature, there is a lack of empirical studies into how this plays out in practice in data-driven software teams. This study thus seeks to add new insights into how the compliance principle may create congestion in the continuous delivery process due to a limited understanding of data usage legalities.

The importance of juridical competence for software teams has meant that IT departments have started implementing legal professionals into product teams, aiding software developers in maneuvering the legal grey zones in datafication processes. Our present study contrasts one team with another team that did not have a legal advisor available, and we hence seek to provide unique insights in this regard. Although organizational agility is critical to respond sufficiently to the challenges experienced by contemporary software projects, team members with different backgrounds may have different norms guiding them, often proving to be a hindrance to being an effective agile team [27]. Addressing this in the context of a product team incorporating a legal advisor, we also offer novel empirical findings contributing to this literature in software engineering, as this aspect is not discussed in any of the previous literature as presented in the present section. Additionally, the paper adds to the state of the art in that the Norwegian public sector (and especially NAV and Entur) has been a global pioneer regarding incorporating new ideas about organizational architecture in software development, to date still largely untheorized within empirical studies of public sector software engineering.

3 Method and Research Site

To answer the research questions, we utilized a qualitative case study design [28] to tell the contrasting stories of two agile teams in NAV (Team Welfare) and Entur (Team Travel). We chose these cases because both are public companies with a data mesh strategy for becoming data-driven (i.e., decentralizing data ownership to product teams), allowing us to compare two cases within similar contexts. Team Welfare included a legal advisor, which Team Travel did not. We wanted to explore differences in overcoming privacy issues in the two teams and the collaboration between developers and team members with legal expertise. Our goal was to shed light on the challenges faced in the pursuit of becoming a data-driven public sector. Specifically, how to establish and maintain legal standards and procedures for effective data handling.

We kept an exploratory approach as we did not set out to test any specific theory or hypothesis [29]. Further, we hold an interpretive view in this study, comprehending the world and its truths as subjective realities [30]. As per Yin's [28] approach to case study research, case studies are not intended to be statistically generalized, and this can thus be said to be a limitation of case studies and qualitative studies in general. However, the study can be analytically generalized as it establishes an approach to studying software

developers' struggles in maneuvering data protection regulations, an ongoing challenge applicable to software teams worldwide.

Case Descriptions

Both teams apply the four core concepts of Agile: Incrementally designing the software through iterations, instituting ceremonies for inspecting and adapting the product and development process, responding to change collectively, and continuously involving their users [31]. They are structured as typical agile teams with a Team Lead, Product Owner, and team members like developers, testers, and UX designers. Team members were both juniors and seniors, holding between 1–25 years of experience.

Team Travel (12 members) develops and runs an app where travelers can search travel routes and buy tickets across transportation modes like buses, trams, trains, subways, ferries, scooters, and city bikes. They are part of Entur, a public company that provides a digital infrastructure to the Norwegian public transport system. For example, components like payment solutions that travel companies can include in their services, thus relieving them from developing a payment solution themselves. Entur has more than 100 developers organized into 20 development teams, and each team is responsible for a specific part of the digital infrastructure. The teams are described as autonomous, meaning they choose freely how they solve their tasks and rely on development methods. Teams include front-end and back-end developers, designers, and product owners. Entur is inspired by the thoughts behind data mesh [21] and is building a data platform where development teams can publish the data they produce. However, the teams are themselves seen as owners of the data.

Team Welfare develops digital applications used by citizens in need of applying for benefits related to special conditions in life. The new web-based application service is replacing a partly paper-based system. The national welfare administration – NAV – is Norway's largest governmental agency, responsible for distributing 1/3 of the federal budget. NAV has 2000 employees, almost 400 product developers, and 150 product teams organized in 10 product areas. NAV has been at the forefront of working towards becoming data-driven and experimenting with the potential of data utilization. By employing over half of the 800 developers, they went from being dependent on large consultancy companies to being in charge themselves. Insourcing was an essential strategic measure in moving towards a DevOps mindset and going from a few releases a year to more than 1500 a week. NAV is moving towards a decentralized data management model [5] heavily inspired by data mesh [21] which imposes ownership for data that applications produce on the product teams.

Research Strategy

We collected data in two rounds of interviews (see Fig. 1). We encountered the privacy issue during the first round of interviews in Team Travel. These interviews were part of a study of development teams' coordination strategies and were reported in a previous paper [32]. Then we explored if privacy issues were also relevant for NAV by interviewing members from different teams and eventually found Team Welfare. The first round of interviews revealed the practical problems of data privacy and gave us an understanding of the context. In the second round, we studied both Team Welfare and Travel's practices for managing data and privacy issues. Interviews were recorded, transcribed, and

coded open-ended using the tool NVivo to identify and structure common codes into larger constructs [33]. Coding and memoing were conducted according to the Constant Comparison Method [34].

To answer the research questions and gain a deep understanding of how data privacy regulations affected the teams’ day-to-day operations, we used interviews to pursue detailed stories and descriptions from our respondents. Interviewing allowed us to probe into *why* privacy regulations changed their way of working and *how*. As this topic has increasingly become more and more relevant because the teams started collecting new forms of user data in the last couple of years, it is still a novel and emerging topic for most team members. Interviewing allowed us a window into their ongoing reflections, and it sparked new thoughts in the respondents. This enabled us to compare different perspectives and reflections.

We also analyzed applicable documents like the team’s strategy documents and Norway’s Digitalization Strategy [35] to explore the research questions thoroughly. For instance, Team Welfare’s strategy documents described their team purpose, goals, and development philosophy.

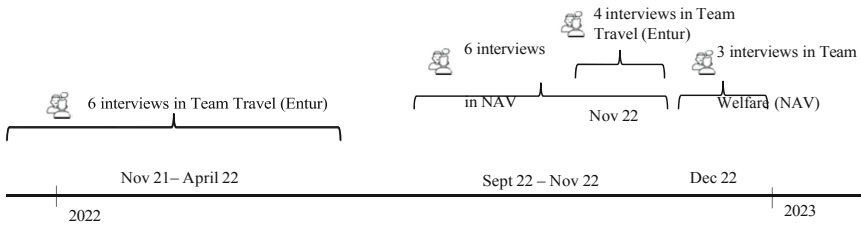


Fig. 1. Timeline for data collection of this study

4 Findings

We found that the two teams encountered the challenge of assessing if data is sensitive in three situations, namely when collecting, sharing, and storing data. These challenges, however, played out differently due to one central aspect: In seeing their developers’ continuous challenges in maneuvering juridical regulations, Team Welfare had met this obstacle by employing a legal advisor assisting both Team Welfare and their “sister team” that worked on another part of the application solution. The two different teams were separate product teams, but still in the same NAV unit, only divided in terms of the different product responsibilities. The legal advisor was particularly engaged in issues dealing with privacy laws and data privacy consequences in their communication with other teams in NAV. Typically, the dialogues took place on Slack, where the different teams exchanged questions, thoughts, and experiences.

However, Team Travel did not have a legal advisor available. They had to evaluate the challenges of data handling independently or seek juridical assistance elsewhere. The following maps out the various challenges public sector product teams encounter regarding data rules and regulations, and the contrasting experiences the two teams reported in this regard.

4.1 Collecting Data

As outlined in Sect. 2.1, collecting various data about the public is a central part of the Norwegian welfare system, and in the quest for a data-driven public sector part of this responsibility now falls on the software teams developing the new solutions.

A central concern in that regard had become an often-occurring obstacle for Team Travel, who struggled to assess what data they were allowed to gather, resulting in a “better safe than sorry” mentality. *“We just want to do it right, which limits us [in what data we collect],”* one developer explained. For example, they did not know if they could gather device IDs from user logs and avoided this by gathering user numbers instead. However, only those users who make a purchase or log into their account hold a user number, meaning the team missed out on data from many users. This case describes how collecting data felt risky, and the team accepted lower data quality to lower the risk. There were also other issues; the effort of interpreting privacy regulations was seen as too high risk, and they instead decided to leave the data uncollected.

The wide variety of possible privacy issues developers must consider when gathering data can be described through a question from one of Team Travel’s users. *“A user asked if it was possible for us to identify someone who frequently searched for travel routes from an address to a hospital. In a way, this now becomes sensitive data.”* This was beyond what the team imagined they would have to consider and revealed the additional workload data impedes the development work in form of assessing privacy regulations.

4.2 Sharing Data

A central incentive for organizations to become data-driven is the enormous potential and value-creation in data sharing. The Norwegian digitalization strategy 2019–2025 words it as “[w]e must share and reuse more public data, and we must ensure that regulations are digitalization friendly” [35]. For public administrations, data-sharing initiatives are in theory particularly seen as vital for creating common welfare solutions that function across administrative sectors. In practice, however, sharing data also entails major obstacles for software product teams responsible for creating such data-sharing solutions.

The product owner of Team Welfare explained that *what* they can share and *in what way* they share it was an ongoing conversation in the team, especially when they wanted to share historical data with the managers at the various local welfare centers with the incentive of helping them foresee future requests: *“That’s when a lot of legal stuff came up. If they can dig into this, then they can get down to the individual level. In small cases, you can actually locate which user it is, because there may not be that many with the specific needs within a small geographical area”*. This issue was echoed by a Team Welfare developer who referred to what they used as a rule of thumb, namely how data sharing does not abide by data privacy rules if fewer than five people are identified in the material. This “entanglement problem” was especially significant in smaller municipalities where few people applied for the same kinds of aid.

Simultaneously, the teams did not want to own more data than necessary because they did not want to be responsible for it. Gathering data implies a risk of leakage; the more data a team “owns,” the greater the risk. They rather relied on tapping into other

teams' data than gathering it themselves. *"There is a greater risk of leaking sensitive data if it is stored in many places and spread out,"* one developer in Team Travel explained. Tapping into other teams' data implies a willingness to share. Teams then must assess what data they are allowed to share and with whom. *"We talk to other teams to get the data from them [instead of gathering it ourselves] to avoid spreading [ownership of] the data around,"* the developer continued. The uncertainty resulted in hesitation, which led to time delays. It may take months from issuing a sharing request until the data is shared.

A developer in Team Welfare described how owning data was adding extra burden. They often received data requests and had to assess if they could share the data. Requests typically came from other teams that needed business data to develop their own software and managers that used business data to steer resources. *"But we somehow have to assess this all the time, are we allowed to share [the data] with them?"* a developer asked, illustrating their struggles. This was a new problem for the product team. When data was centralized, someone else made this decision; typically, the central data warehouse team. Now, every product team must interpret privacy rules to determine if sharing complies with privacy laws. *"Who is responsible if you share data with someone and they share it further? Is it the people who originally produced the data or is it the consumer?"* the developer continued. To find answers, they must interpret privacy regulations, which made teams anxious. *"I am terrified of a newspaper headline on an individual's sensitive data being leaked. That would be a disaster,"* the developer said. Hesitation seems to be the result, delaying any team that requests data.

Guidelines and instructions are made to support the teams in assessments. However, guidelines are made generic to comply with various domains and contexts. Thus, interpretation is still necessary, which maintains the hesitance of developers. Luckily for Team Welfare, they had a legal advisor supporting them in making these decisions, shortening the time from request to sharing. In contrast, Team Travel was stuck in a world of uncertainty for every incoming request resulting in time-demanding discussions. These findings correspond with NAV's Current Situation Analysis report from October 2021, where one of the central challenges identified moving forward into becoming data-driven was connected to the sharing of data: *"Regulations prevent data sharing and collaboration. The possibility of increased data access and automation is limited by GDPR regulations, other legislations, and system barriers within NAV, between agencies and other actors. This also makes the data quality too poor."*

4.3 Storing Data

The Norwegian digital strategy for the public sector states that "[o]pen public data shall be made available for reuse for developing new services and value creation in the business sector." This digitalization incentive is yet another aspect of the datafication process that directly creates increased responsibility for public sector software teams.

For instance, Team Travel struggled to conclude if using Google Analytics complied with privacy regulations. Google Analytics supports teams in gathering and visualizing user metrics (i.e., how many clicks on a button) but requires the data to be uploaded to Google's cloud service. Privacy regulations require specific types of data to be stored within Europe, and Team Travel was in doubt if Google stored data on US servers too.

“Even though the new Google Analytics 4 is supposed to be much more compliant and such, we don’t really trust it,” a developer said. They sought answers from the Norwegian Data Protection Authority who told them that they do not consider individual cases, leaving Team Travel in limbo. *“Until we find an alternative tool, we don’t utilize the user metrics.”*

Searching for alternative tools is easier said than done. Data that initially seem harmless can be deceitful when combined with a cloud service. A developer in Team Travel explained that Google Analytics collects cookies on user data which they combine with cookies from other services and build a profile on the user to target advertisement. This means that seemingly harmless and anonymous user metrics turn out to be a potential privacy issue. Team Travel had been searching for alternative tools for more than nine months, with no apparent solution in sight.

In contrast, Team Welfare felt safe in using their Big Query-based (Google Cloud service) in-house service for visualizing user metrics. They did not have to consider if the service was compliant because legal experts had already approved it. Team Welfare still had to consider privacy regulations when uploading data, like Team Travel, but their legal advisor supported them. The big difference between Team Welfare and Travel was that the legal advisor in Team Welfare sat close to the domain, meaning they knew what data user metrics entail in software engineering. Thus, they were willing to give guidelines on individual cases and take on shared responsibility for uploading the data.

In addition, the laws and regulations for storing data also entail that software teams must meet requirements for continuously storing and documenting their *data usage*, and formally document decision processes behind their practices. However, evaluating whether the data can be labeled as sensitive is an ongoing challenge. Oftentimes, if data is considered harmless and only needed for small decisions and not big deliveries, such evaluations will not be formally archived, but rather “hidden” in informal conversations on Slack or Teams. Underlining the difficulties in evaluating the sensitive or non-sensitive nature of data in the context of storing work logs, the Team Welfare legal advisor said: *“I’m not always quite sure when we must document what. [...] So it’s a challenge coming in from outside, from another place in the world, in a way, and into agile development.”*

4.4 Legal Advisor: A Highway to Autonomy for Data-Intensive Teams?

Although both Team Welfare and Travel encountered various difficulties navigating the different data privacy regulations in their day-to-day work, it was clear that Team Travel experienced more deep-rooted issues, such as not utilizing user metrics due to insufficient legal support. However, inserting a legal advisor into a product team may also come with its own set of challenges. We identified two somewhat intertwined causes of friction in that regard: namely, team synchronizing and interdisciplinary understandings.

Regarding team synchronizing, one central issue was that all the interviewed members on Team Welfare used various communication channels to stay in contact with other team members, other NAV teams and different user groups. Among the most utilized platforms were Teams, Slack and NAV’s internal communication channel for reporting issues, and they had established various group chats with different work topics on the different platforms. Not only did this make it hard to keep track of whom they had talked to and on which channel about which issues; it also made it difficult to track whether

they had documented their data processing according to the rules for how this should be done correctly.

The use of different communication platforms seemed connected to the preferences of people from different disciplinary backgrounds. For instance, Slack was a central communication channel for the technical parts of the team [see 36], whereas the communication with the different case workers took place in closed groups on Teams. Although this point refers to the inclusion of team members from different backgrounds more generally, it was particularly noticeable for Team Welfare, likely due to their team composition being more wide-ranging in terms of interdisciplinary backgrounds.

Additionally, several Team Welfare members pointed out the importance of how synchronizing the team was a longitudinal process, particularly when the team was of an interdisciplinary nature. In the context of the legal advisor, this meant that this person would need to spend a significant amount of time with the team to observe and understand their challenges and everyday practices to truly grasp where assistance would be necessary for the team to operate as efficiently as possible. We will discuss this finding closer in the subsequent section.

In agile software engineering literature, it is well established that there are various barriers to be overcome in interdisciplinary teams for them to become autonomous. That mutual adjustments were challenging was especially articulated by the Team Welfare legal advisor, who said: *“I think [adjustment] is a very big challenge in working interdisciplinary, as we do in the interdisciplinary teams. It’s because we come from different environments. We bring with us very different experiences and knowledge, but we also use words very differently. [...] Then it turns out that we mean very different things when we use the same term”*. The legal advisor mentioned “authorization” and “consent” as often-occurring words, two terms that are central in many contexts related to data handling and storing. They may however have differing connotations for different disciplines.

Although Team Welfare had integrated a legal advisor as part of their team and this person often sat next to the developers when trying to iron out juridical hurdles, the legal advisor seemed to have little understanding of how much the data regulations impacted the developers’ day-to-day tasks. It was clear that the legal advisor seemed to believe the developers solely care about effective technical solutions, not societal responsibility:

There is a very basic thought I get from time to time, that we bring in developers who think it is exciting to work with an idea that ‘we can just make a shopping cart’ where they can sort of get to pick the things they want. And then, the reality is far from that [...] It is, after all, an administrative responsibility that lies at the bottom of everything we do, which must be communicated from scratch.

The legal advisor thus doubted the software developers’ ability to grasp the larger picture of their work, which seemed far from the developers’ own stories in that regard. As a Team Welfare developer put it: *“[Juridical issues] is something that everyone should be aware of, it is not like you should always think that this is something that the lawyers on the team handle and are responsible for. Everyone must consciously evaluate what kind of data we store, whether we have the right to handle that data, whether we really need to check the legal basis more carefully before we do anything.”* However, it was also

evident that with proper team synchronizing came mutual adjustments and increased interdisciplinary understandings throughout the entire product team-user chain. And despite the various challenges with data regulations and usage, the Team Welfare product owner underlined the importance of not being blindsided by the juridical hurdles, always keeping in mind how data actually promotes the development of software products: *“I really think that the most important aspect with data for us is knowing that it is the right product that we are making, and somehow finding confirmation of that and seeing that it provides the value that our project is meant to have”*. The same team’s legal advisor concluded: *“For me, I think the keyword is interdisciplinary understanding. At least, that is a keyword for developing good solutions.”*

5 Discussion and Implications for Practice

With public sector organizations globally being in the midst of data-driven “transformation journeys” [37], inevitably, both the software developer role and the composition of product teams are changing rapidly. We here return to our research questions: How are data privacy regulations affecting the day-to-day work of product teams? And what are the pros and cons of including a legal advisor as part of the team to overcome juridical hurdles? In answering these, we discuss the direct implications datafication has on interdisciplinary agile product teams’ practices, focusing first on how laws regulating data collection, sharing, and storing data expand and complexify software developers’ responsibilities.

Data work has not traditionally been a part of the software developer role. Amidst the quest towards becoming data-driven, organizations – both public and private – have added new areas of competence to the developer profession – as well as new liabilities: Carefully considering ethical and juridical aspects of using data now permeate software developers’ day-to-day tasks, and for both teams being responsible for data had become a burden as much as a realm of possibility. Although decentralizing data storage, in theory, gives product teams more control and the ability to self-organize, the context-dependent and ambiguous laws regulating data collection, storage and sharing meant that the potential value-creation of data brought about ambivalent feelings for the teams.

To realize the potential benefits of datafication in the public sector, data needs to directly reflect the public. However, knowing that any minor misuse of personal data could lead to newspaper headlines, developers are forced to settle for fewer data and data of lesser quality than optimal to avoid breaking potential privacy regulations. Consequently, the teams are forced to make decisions and deliver continuous updates based on a skewed picture of their users, i.e., the Norwegian public. Ultimately, the organizations risk investing huge resources in becoming data-driven only to make the day-to-day tasks of developers more demanding and create a false impression of being driven by “real world” data. This corresponds to and adds further layers to the findings of Broomfield and Reutter [14] whose “search for the citizen” in data-driven public administrations find that civil society tends to be obscured in the data and reduced to passive users in datafied public administrations.

Another vital finding in this respect is that the classification of data as either “private” or “sensitive” is not straightforward in a data-intensive software development context.

As the teams pointed out, when faced with large amounts of data that were seemingly risk-free, they oftentimes found themselves in a situation where the *combination* of these data in fact created potential data privacy violation issues. As the study of data democratization is in its early phase [6], more studies are needed here both in order to find better ways of classifying data and in terms of empirical studies on software teams' experiences with data classification.

Zooming in on such situations in the teams we studied, our findings suggest that the team working closely with a legal advisor was better equipped for responding to data privacy regulations. And although there were also challenges involved with inserting a team member whose responsibility was the juridical side of data-driven software development, Team Welfare had slowly but surely developed tactics to make the team more autonomous. For instance, the legal advisor underlined the importance of being "close by" the technical sides of the teams to "grasp" the issues when they emerged, an understanding that was shared by the whole team.

Interestingly, the Team Welfare legal advisor saw it as a large responsibility to educate the technologists about the rules and regulations they must operate within. This responsibility, however, was also well-engrained into the minds of the developers. So, despite the interdisciplinary nature of the team, the understanding of others' ethical responsibility regarding data handling was not communicated that well between the group members from different disciplines, perhaps due to the different terms used by the various disciplines.

Another important point we draw from the findings in this regard is that it is not enough for the legal advisor to join the team for short sessions now and then: It takes time to synchronize the interdisciplinary teams to be as autonomous as possible and for the legal advisor to develop a contextual understanding of the data product teams and their day-to-day practices and challenges. Several informants pointed out how the teams must "mature together" over a longer period of time.

Our study shows that there likely are many benefits of including a legal advisor in data-intensive product teams, as it undoubtedly makes them more confident in their day-to-day data handling practices. However, as previous studies also have pointed to [23], expanding software development teams and including new, formal roles beyond the technical team requires much effort in synchronizing and overcoming interdisciplinary barriers. The additional financial costs of including legal experts must also be examined. We find, however, that the costs of *not* having continuous access to a legal advisor can likely be much larger, considering that the outcome may be week-long – perhaps even month-long – delays in deliveries, as well as mishandling user data. This especially holds true for organizations managing large amounts of data that are difficult to evaluate as sensitive or not, and the stakes are high in terms of potential misuse of the data in question. We thus especially recommend such teams to consider the possibility of including a legal expert as part of the team.

6 Conclusion and Recommendations for Future Work

The increasing regulations and complexities around datafication processes have expanded the responsibilities and expectations of software developers. To understand how agile teams deal with privacy regulations when becoming data-driven, we investigated two agile product teams. One team had a legal advisor as one of the team members,

and one team did not, keeping this expertise outside the team. Both teams struggled with knowing which data they were allowed to collect and share. We found that having a legal advisor on the team shortened the time to solve privacy issues. The team without this role had time-demanding discussions within the team, and experienced uncertainty and hesitation to act, which resulted in privacy issues taking months to solve. While it was beneficial to have a legal advisor, challenges included coordination and syncing issues among the team members. For example, because they were from different domains, they had different preferences for communication tools (the technical team members preferred to communicate on Slack). Coming from different backgrounds also meant that they had different understandings of terms (e.g., the term “authorization”), which potentially could lead to misunderstandings in the team.

With both public and private sector software projects focusing increasingly on the possibilities of data-driven development, it is vital that also the software engineering research field reflects this impending paradigm shift. Future work should hence monitor closely how datafication affects software developers and their increasingly interdisciplinary teams to pinpoint the implications these changes have on day-to-day practices. Further studies are especially needed addressing how the principle of continuous compliance can create significant hurdles for software development and how this best can be overcome for teams also in various parts of the private sector.

As this study addresses only two teams within the Norwegian public sector, it is hence limited in the sense that it cannot be generalized to other contexts. We thus recommend future work to address these issues with a quantitative or mixed-methods approach, that can show a broader scope of how software developers and product teams tackle data protection regulations. Additionally, as digitalization processes also take place outside Western societies, however receiving considerably less academic attention, future studies addressing how software developers across the democratic world are seeking ways to tackle these issues would be especially welcoming for broadening the scholarship.

Acknowledgments. We wish to thank NAV, Entur and the informants for willingly sharing their experiences. Also, we thank Knowit AS and the Norwegian Research Council for funding the research (grant number 321477).

References

1. van Ooijen, C., Ubaldi, B., Welby, B.: A data-driven public sector: enabling the strategic use of data for productive, inclusive and trustworthy governance. OECD, Paris, May 2019. <https://doi.org/10.1787/09ab162c-en>.
2. Andreassen, R., Kaun, A., Nikunen, K.: Fostering the data welfare state: a Nordic perspective on datafication. *Nord. Rev.* **42**(2), 207–223 (2021). <https://doi.org/10.2478/nor-2021-0051>
3. Dencik, L., Kaun, A.: Datafication and the welfare state. *Glob. Perspect.* **1**(1) (2020). <https://doi.org/10.1525/gp.2020.12912>
4. Reutter, L.: Constraining context: situating datafication in public administration. *New Media Soc.* **24**(4), 903–921 (2022). <https://doi.org/10.1177/14614448221079029>

5. Vestues, K., Hanssen, G.K., Mikalsen, M., Buan, T.A., Conboy, K.: Agile data management in NAV: a case study. In: Stray, V., Stol, K.J., Paasivaara, M., Kruchten, P. (eds.) *Agile Processes in Software Engineering and Extreme Programming*, pp. 220–235. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-08169-9_14
6. Kashyap, B., Shinghal, K.: Data democratisation: the evolution of informed decisions (2019). <https://doi.org/10.13140/RG.2.2.33103.51361>
7. Stray, V., Moe, N.B., Hoda, R.: Autonomous agile teams: challenges and future directions for research. In: *Proceedings of the 19th International Conference on Agile Software Development: Companion*, Porto Portugal, pp. 1–5, May 2018. <https://doi.org/10.1145/3234152.3234182>
8. Hirsch, D.D.: The glass house effect: big data, the new oil, and the power of analogy. *Big Data* **66**
9. Klotins, E., Peretz-Andersson, E.: The unified perspective of digital transformation and continuous software engineering. In: *2022 IEEE/ACM International Workshop on Software-Intensive Business (IWSiB)*, pp. 75–82, May 2022. <https://doi.org/10.1145/3524614.3528626>
10. Jahns, V.: Data fabric and datafication. *ACM SIGSOFT Softw. Eng. Notes* **47**(4), 30–31 (2022). <https://doi.org/10.1145/3561846.3561854>
11. Fernández-Rovira, C., Valdés, J.Á., Molleví, G., Nicolas-Sans, R.: The digital transformation of business. Towards the datafication of the relationship with customers. *Technol. Forecast. Soc. Change* **162**, 120339 (2021). <https://doi.org/10.1016/j.techfore.2020.120339>
12. van Dijck, J.: Datafication, dataism and dataveillance: big data between scientific paradigm and ideology. *Surveill. Soc.* **12**(2), 197–208 (2014). <https://doi.org/10.24908/ss.v12i2.4776>
13. What is Datafication? Definition from Techopedia.com. <http://www.techopedia.com/definition/30203/datafication>. Accessed 03 Feb 2023
14. Broomfield, H., Reutter, L.: In search of the citizen in the datafication of public administration. *Big Data Soc.* **9**(1), 20539517221089304 (2022). <https://doi.org/10.1177/20539517221089302>
15. Misuraca, G., Van, N.C.: AI Watch - Artificial Intelligence in public services. JRC Publications Repository (2020). <https://publications.jrc.ec.europa.eu/repository/handle/JRC120399>. Accessed 03 Feb 2023
16. Ruppert, E., Isin, E., Bigo, D.: Data politics. *Big Data Soc.* **4**(2), 2053951717717749 (2017). <https://doi.org/10.1177/2053951717717749>
17. Tupasela, A., Snell, K., Tarkkala, H.: The Nordic data imaginary. *Big Data Soc.* **7**(1), 2053951720907107 (2020). <https://doi.org/10.1177/2053951720907107>
18. Sadowski, J.: When data is capital: datafication, accumulation, and extraction. *Big Data Soc.* **6**(1), 2053951718820549 (2019). <https://doi.org/10.1177/2053951718820549>
19. *Designing Data-Intensive Applications: The Big Ideas Behind Reliable, Scalable, and Maintainable Systems*: Kleppmann, Martin: 9781449373320: Amazon.com: Books. <https://www.amazon.com/Designing-Data-Intensive-Applications-Reliable-Maintainable/dp/1449373321>. Accessed 03 Feb 2023
20. Dehghani, Z.: *Data Mesh: Delivering Data-Driven Value at Scale*, 1st edn. Beijing Boston Farnham Sebastopol Tokyo: O'Reilly Media (2022)
21. *Data Mesh Principles and Logical Architecture*. martinowler.com. <https://martinowler.com/articles/data-mesh-principles.html>. Accessed 03 Feb 2023
22. Frontpage. Datatilsynet. <https://www.datatilsynet.no/en/>. Accessed 03 Feb 2023
23. Moe, N.B., Stray, V., Hoda, R.: Trends and updated research agenda for autonomous agile teams: a summary of the second international workshop at XP2019. In: Hoda, R. (ed.) *Agile Processes in Software Engineering and Extreme Programming*, pp. 13–19. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-30126-2_2

24. Hukkelberg, I., Berntzen, M.: Exploring the challenges of integrating data science roles in agile autonomous teams. In: Hoda, R. (ed.) XP 2019. LNBP, vol. 364, pp. 37–45. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-30126-2_5
25. Klotins, E., Gorschek, T., Sundelin, K., Falk, E.: Towards cost-benefit evaluation for continuous software engineering activities. *Empir. Softw. Eng.* **27**(6), 157 (2022). <https://doi.org/10.1007/s10664-022-10191-w>
26. Alhazmi, A., Arachchilage, N.A.G.: I'm all ears! Listening to software developers on putting GDPR principles into software development practice. *Pers. Ubiquit. Comput.* **25**(5), 879–892 (2021). <https://doi.org/10.1007/s00779-021-01544-1>
27. Stray, V., Fægri, T.E., Moe, N.B.: Exploring norms in agile software teams. In: Abrahamsson, P., Jedlitschka, A., Nguyen Duc, A., Felderer, M., Amasaki, S., Mikkonen, T. (eds.) *Product-Focused Software Process Improvement*, pp. 458–467. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-49094-6_31
28. Yin, R.K.: *Applications of Case Study Research*. SAGE (2011)
29. Runeson, P., Höst, M.: Guidelines for conducting and reporting case study research in software engineering. *Empir. Softw. Eng.* **14**(2), 131 (2008). <https://doi.org/10.1007/s10664-008-9102-8>
30. Oates, B.J.: *Researching information systems and computing*. Sage (2005)
31. Baham, C., Hirschheim, R.: Issues, challenges, and a proposed theoretical core of agile software development research. *Inf. Syst. J.* **32**(1), 103–129 (2022)
32. Sporse, T., Moe, N.B.: Coordination Strategies When Working from Anywhere: A Case Study of Two Agile Teams. arXiv, 08 April 2022. <https://doi.org/10.48550/arXiv.2204.03978>
33. Technische Universität München et al.: Grounded theory methodology in information systems research. *MIS Q.* **41**(3), 685–701 (2017). <https://doi.org/10.25300/MISQ/2017/41.3.02>
34. Seaman, C.B.: Qualitative methods in empirical studies of software engineering. *IEEE Trans. Softw. Eng.* **25**(4), 557–572 (1999). <https://doi.org/10.1109/32.799955>
35. K. moderniseringsdepartementet: Én digital offentlig sektor: Digitaliseringsstrategi for offentlig sektor 2019–2025. Regjeringen.no, 14 June 2019. <https://www.regjeringen.no/no/tema/statlig-forvaltning/ikt-politikk/digitaliseringsstrategi-for-offentlig-sektor/id2612415/>. Accessed 06 Feb 2023
36. Stray, V., Moe, N.B., Noroozi, M.: Slack me if you can! Using enterprise social networking tools in virtual agile teams. In: 2019 ACM/IEEE 14th International Conference on Global Software Engineering (ICGSE), pp. 111–121 May 2019. <https://doi.org/10.1109/ICGSE.2019.00031>
37. Stettina, C.J., van Els, V., Croonenberg, J., Visser, J.: The impact of agile transformations on organizational performance: a survey of teams, programs and portfolios. In: Gregory, P., Lassenius, C., Wang, X., Kruchten, P. (eds.) XP 2021. LNBP, vol. 419, pp. 86–102. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-78098-2_6

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

