The Journal of Systems and Software 218 (2024) 112185

Contents lists available at ScienceDirect

The Journal of Systems & Software

journal homepage: www.elsevier.com/locate/jss

Towards a common data-driven culture: A longitudinal study of the tensions and emerging solutions involved in becoming data-driven in a large public sector organization

Astri Moksnes Barbala *, Geir Kjetil Hanssen, Tor Sporsem

SINTEF Digital, Trondheim 7034, Norway

ARTICLE INFO

Editor: Dr Alexander Chatzigeorgiou

Keywords:
Data-driven organisation
Data mesh
Public sector development
Data-driven culture
Data culture
Longitudinal case study

ABSTRACT

In recent years, the push to make organizations data-driven has led to data-focused software projects, both in the private and public sectors. The strive for increasing data-driven initiatives introduces a range of new socio-technical challenges, yet there are to date few empirical studies in terms of how data-focused initiatives affect large organizations with significant variations in terms of data needs and usage. This study presents a longitudinal descriptive case study of how data-driven initiatives in the Norwegian public sector cause organizational tensions in a very large, complex organization. We conducted 32 semi-structured interviews over a period of 18 months representing two different data-intensive parts of the organization that had developed incompatible data cultures. Our study shows that these cultural differences create organizational conflicts that hinder data-driven initiatives. The findings also suggest, however, that overcoming these is possible through the strategic, top-down facilitation of a common data-driven culture built on uniting data principles, in turn potentially leading to improved decision-making and enhanced innovation.

1. Introduction

As modern organizations aim to become increasingly data-driven, both in day-to-day operations and strategic decision-making, as well as in software-based service development, managing central data strategies has become more challenging. Enhanced access to higher-quality data and the advent of new analytical technologies provide unprecedented opportunities to improve organizational efficiency and effectiveness. However, data can no longer be regarded as a resource for a select few; it must be managed and utilized strategically across the entire organization (see e.g., Anderson, 2015).

Data thus needs to be a central component in all service development and delivery. For software development teams, transitioning to a data-driven organization has large implications. They can no longer rely on centralized data management teams to extract and transform their data but need to play an active role in collecting and distributing the data produced by their applications. In the same way as movements such as DevOps and BizDev have transformed the way teams work and interact (see e.g., Fitzgerald and Stol, 2017), the increasing use and importance of analytical data is challenging development teams to take on extended responsibilities.

Concepts such as DataOps and data mesh (see e.g., Munappy et al., 2020; Dehghani, 2022) imply that development teams should be responsible for populating the data ecosystem by producing, sharing, and consuming operational and analytical data. This can be termed “data democratization” (e.g., Awasthi and George, 2020) where the idea is that the data value chain needs to be shortened and simplified to fully utilize data. Although there seems to be agreements among both practitioners and researchers about the necessity of such a transformation, we still know little about what the challenges are and how they can be approached in practice within large and complex data-intensive organizations.

With data becoming a central part of the organizational culture, many guides for how companies can create a *data-driven culture* (see e.g., Storm & Borgman, 2020) have been written over the past few years. The literature has however not made clear distinctions between e.g., small companies and larger enterprises, or public and private sector organizations, requiring different approaches to creating data-driven cultures. As previous research has shown, public sector organizations are fundamentally different from private sector ones regarding access to resources, goals, and economic and political constraints (see e.g., Parker and Bradley, 2000), and there is particularly a lack of studies

* Corresponding author.
E-mail address: astri.barbala@sintef.no (A.M. Barbala).

https://doi.org/10.1016/j.jss.2024.112185
Received 28 November 2023; Received in revised form 4 August 2024; Accepted 13 August 2024
Available online 17 August 2024

centralizing how data-driven initiatives have played out in public sector contexts, which arguably present more complex challenges than in the private sector. Therefore, our focus is here on a large public sector organization, and we ask the following research questions:

RQ1: How is public sector organizational culture impacted by the incorporation of data mesh principles for leveraging and managing data?
RQ2: And how can large, complex organizations overcome tensions emerging from the development of conflicting data cultures?

To answer the research questions, we draw on insights from ongoing fieldwork within NAV, short for the Norwegian Employment and Welfare Organization. NAV is the cornerstone of the Norwegian welfare system, responsible for enhancing the workability of the population and providing economic assistance during times when workers cannot sustain themselves. NAV employs thousands of caseworkers who assist citizens with a wide range of needs from the welfare state. Hence, NAV produces a rich portfolio of information systems to support their own caseworkers, citizens, employers, health professionals, family members, etc. NAV's IT department is one of the largest software development organizations in Norway, managing a very large amount of data with a very high level of complexity and criticality for both citizens and society.

NAV manages a third of Norway's national budget, and is also responsible for producing public and official statistics. To enable such statistics, operational data have over the years been extracted from source systems and transformed into a coherent data warehouse model that has had a relatively stable information architecture over time. These extractions and transformations are managed by a centralized unit of data engineers and statisticians.

For parts of the organization, a recurring challenge with the centralized model has been the distance between data producers and data consumers and difficulties with the extraction of data from source systems. These challenges have become increasingly prevalent as the organization aims to increase the quality and quantity of analytical data, and the extent to which it is used to create and improve public services. This development adds to the transition to become a more agile organization in terms of a more distributed system architecture based on microservices and through self-organized domain teams (Vestues et al., 2022). This development has made it more visible that there are differing data needs. Agile domain teams need flexibility in access and use of data, while those producing statistics rely on the stability and controllability of data.

To address the issues arising in creating and improving data-driven services, NAV has begun to implement a new data strategy, where product teams take increased responsibility for extracting and sharing analytical data. In discussing this in a previous paper (Vestues et al., 2022), we found that the transition towards distributed data ownership causes four main challenges for the organization:

1). *Change of control of data extraction and transformation:* Analysts in the statistics section of NAV (Knowledge Department) started questioning how the organization can maintain cross-organizational insights and communicate new data needs if the centralized data management model is replaced by local ownership and domain-specific data products, however, the previous study did not explore these concerns further.
2). *Managing rightful and legal access to, and use of data:* The principle of domain teams providing data as accessible products via a data platform raised concerns about control and compliance with data protection regulations like GDPR. Balancing this with the data mesh's goal of extensive data collection created uncertainty on how to manage a data mesh while adhering to these regulations (see also Barbala et al., 2023).
3). *Creating data products:* Traditionally, software product teams focused on their own application's data, leaving central data engineers and analysts to prepare data for analytics, necessitating added competence and capacity in domain teams. However, the required skillset and cost were unclear. Additionally, creating valuable data products that span multiple domains requires standardization and cross-domain insight, raising questions about the extent of automation support by the data platform, and how teams should learn about data needs in other domains.
4). *Establishing a thriving ecosystem:* A functioning data mesh relies on data owners publishing consumable data products, but the study revealed that it is unclear what incentives software teams have in order to invest time in this process. Questions arose about whether publication should be an organizational obligation or if mechanisms similar to open-source development, where sharing data attracts valuable feedback and improvements could motivate teams to ensure high-quality, understandable data products.

The insights of our previous work thus provided us with a deeper understanding of the implications of becoming data-driven for NAV. The current study builds upon the insights gathered in our previous paper, however with the aim to provide a more nuanced view of how the transition inspired by the data mesh concept plays out in practice and affects the organizational culture when there are different data needs connected both to different regulations and societal purposes. As our data collection has continued since the publication of the previous article, as well as the practitioners have come further along the way in the implementation of a new data ecosystem, we have at this point gained a deeper understanding of the implications of becoming data-driven for a large and complex public sector organization, which NAV is a prime – and early – example of.

The remainder of our paper is structured as follows. Section 2 outlines the theoretical pillars of this study, and presents relevant literature on data mesh, data-driven organizations (DDOs), and data cultures. Section 3 provides context to our case and outlines our methodological framework, empirical data, and analysis method. In Section 4, we present our findings, organized in two sections: Firstly, we tell the stories of the two data cultures we see as having emerged throughout the NAV organization, and second, we present the tensions we encountered due to the data cultural differences. The discussion in Section 5 draws on the previously presented literature to discuss the theoretical and practical implications of our findings. By way of concluding, Section 6 reflects on the relevance of our findings, outlines the study's limitations, and suggests future work.

2. Related work

Although new technologies have introduced a range of innovative theoretical perspectives and methodologies in the field of data management, the concepts of centralized and decentralized data handling have remained central topics of discussion for several decades within Software Engineering and Information Systems studies. For instance, these concepts have been frequently explored within the systems theory research paradigm, which provides a comprehensive framework for understanding the complex interactions and interdependencies within organizations. Systems theory posits that organizations function as interconnected systems, where the structure and flow of information critically influence organizational dynamics. In this context, centralized data handling is often associated with enhanced consistency and control (see e.g., Redman, 1992), while decentralized data handling is linked to increased flexibility and innovation (see e.g., Evans, 2003). Rather than adopting a systems theory approach in its entirety due to it being a large and multifaceted perspective, this paper zooms in on three underlying themes—data mesh, data-driven organizations (DDOs), and data cultures—in order to best discuss and elevate our research findings from a narrower scope. This focus allows us to explore how these dynamics play out in a large, complex organization within the Norwegian public sector, contributing to the ongoing discourse on effective data governance and cultural integration in data-driven initiatives.

2.1. Data mesh: A paradigm shift in data architectures

As the NAV IT department in recent years has adopted data mesh principles, the concept is central to the current case to understand the issues arising in the wake of these technological changes in the organization. Zhamak Dehghani (2020) argues that data should be built and managed around “domains”. She proposes four principles that will enable organizations to manage analytical data at scale: 1) Domain-oriented decentralized data ownership and architecture, entailing that data are owned, managed, and located according to their business or thematic domain. 2) Data as a product, that must have the right level of quality and availability. A data product can be described as consisting of code, data, metadata, and the infrastructure needed to execute the code and store the data. 3) Self-serve data platform where teams can deliver their data products to consumers, e.g., to other teams or data analysts. 4) Federated computational governance, which entails a federated approach to govern and improve the data mesh, Governance can be seen as a shared responsibility between data product owners, their consumers, and owners of the data platform. This also corresponds to how data mesh, according to Dehghani, arises as a called-for paradigm shift both on the technological and organizational level, as a common challenge has been unsatisfactory alignments between the organizational needs and the data architectures instituted (Machado et al., 2022).

Despite increasing attention among both researchers and practitioners, there are to date few peer-reviewed empirical studies that explore how agile data management and data mesh is implemented by organizations, although the scholarship is rapidly growing. Drawing on Dehghani, Machado et al. (Machado et al., 2022) refer to the emerging interest in the concept of data mesh as a paradigm shift that will enable organizations to become “truly data-oriented”. Although finding few instances of how data mesh plays out in practice, the authors present two examples of companies that have implemented a data mesh structure in different ways; Zalando and Netflix, and conclude that migrating from data lake architecture to self-serve infrastructure is possible – and also adaptable for interpretation to best suit the different needs and established technology of different companies. Building on these arguments in a different paper, the same authors (Machado et al., 2022) propose a specific technological architecture for the implementation of a data mesh and demonstrate how this type of data architecture can be useful. Similarly, looking into an industrial case of transitioning from a centralized data warehouse to a decentralized data platform architecture, Loukiala et al. (2021) find that “centralized data platforms stop scaling at a certain point when organization domains are large, heterogeneous, and non-flexible by nature” (p. 11). Additionally, they suggest that a distributed approach to data management provides features that enable complex enterprises’ data platforms to scale. Data mesh might increase analytical capabilities, they argue, and contend that a distributed approach to data management facilitates data quality and enables business domains to work independently.

Due to the youthfulness of the topic, especially in terms of scientific evidence of how data mesh can be implemented in practice, looking to grey literature on the topic is useful to gain more insights into the gap between industry practice and the lack of academic literature. In their literature review, which provides an overview of grey data mesh literature, Goedegebuure et al. (2023) seek to synthesize a comprehensive understanding of data mesh by analyzing how data mesh principles are adopted by industry practitioners. In studying 114 documents, mainly blog posts, the authors conclude that although the benefits of data mesh in comparison to traditional data architectures is more scalable, in the sense that it allows organizations to onboard new data sources and new data consumers more rapidly, centralized monolithic data architectures still prevent companies from being able to leverage the full potential and value of their analytical data.

2.2. Data-driven organizations (DDOs)

Despite Norway’s position as one of the most digitally advanced public sectors in Europe, a comprehensive mapping analysis from 2019 revealed few data-driven technologies in production (Broomfield & Reutter, 2019). However, this is rapidly changing, and the current study seeks to bring novel insights into the status quo for data-driven development in the Norwegian public sector. Although a fairly new topic within Software Engineering research, the last few years have seen several studies of different organizations seeking to become data-driven, both with regard to new development processes (e.g., Olsson and Bosch, 2014), the new roles brought about by the focus on data (e.g., Hukkelberg and Berntzen, 2019) and the central role of continuous experimentation in order to successfully transform into a data-driven organization (e.g., Fabijan et al., 2017). Being data-driven entails leveraging data as a central component in decision-making, strategy formulation, and operational processes. Such organizations prioritize the collection, analysis, and interpretation of data to gain insights, drive innovation, and achieve better outcomes (see e.g., Fischer et al., 2022; Berndtsson et al., 2018).

The concept of data-driven organizations (DDO) has gained significant interest, leading to various interpretations within academic and practitioner literature. Although most agree that a DDO entails that data is treated as a strategic asset, and data-driven decision-making becomes embedded in the organizational culture and practices, the interpretations range from simpler to more intricate understandings. For instance, Schüritz et al. (2017) define a DDO as an organization that simply utilizes data and analysis to drive actions. In contrast, other authors like Thusoo and Sarma (2017), who look into the data infrastructure of companies such as Facebook, Twitter, and eBay, assign multiple characteristics, such as a data-driven culture, data-based decision-making, and technological capabilities in their more comprehensive understanding of a DDO. They argue that a transformation toward becoming data-driven requires not only technical change – a central principle here being to adopt DataOps (see e.g., Munappy et al., 2020) – but also entails an organizational and cultural shift. In attempting to synthesize the different conceptualizations of DDOs, Fischer et al. (2022) suggest a framework proposing that an ideal-typical DDO has integrated both an *outside-in* view and an *inside-out* view. By this, the authors indicate that for an organization to be defined as data-driven, it complements utilizing sourced and analyzed data to create new data products and data-driven administrational change *inside* the organization with also providing *external* data-driven value creation. The latter seeks to leverage the organization’s data capabilities and utilize the innovative opportunities they present to achieve impactful outcomes transcending organizational boundaries. This point corresponds with how the Norwegian government’s strategy document “One digital public sector”[1] outlines a goal for public sector organizations to share and reuse more public data in order to create value and efficiency. As the largest public sector organization in Norway, NAV is thus seen to be a vital piece in these endeavors.

An early outline of how organizations should approach the goal of becoming data-driven, Carl Anderson’s “Creating a Data-Driven Organization: Practical Advice from the Trenches” (2015) offers practical advice and valuable perspectives on how organizations can effectively shift towards a data-centric approach. It provides guidance derived from practical experiences and explores different facets of establishing a data-driven organization, encompassing areas such as data infrastructure, data governance, fostering a data-driven culture, utilizing data analytics, and making informed data-driven decisions. Anderson here underscores the significance of aligning individuals, workflows, and technological resources to maximize the potential of data in enhancing business results. Similarly, Berntsson et al.’s (2018) inquiry into the enabling factors organizations should consider when striving to become data-driven and how they can measure their progress while incorporating advanced analytics, finds that organizations that make decisions and take actions based on data gain substantial competitive advantages over their rivals.

Despite an increased scholarly focus on DDOs, however, non-technical studies providing findings from organizations' experiences with becoming data-driven are still scarce, although more studies are currently emerging. Berntsson, Svensson and Taghavianfar (2020) offer one of the most extensive exceptions here, as they utilize interviews with practitioners from nine different software development companies to identify 49 challenges organizations might encounter when adopting a data-driven approach. Based on their empirical findings, the study identified three significant hurdles that require attention, namely establishing trust in the data used for decision-making, transitioning from intuition-based to fact-based decision-making, and fostering a data-driven culture within the organization.

2.3. Data(-driven) cultures

A focus on *company culture* is central in literature discussing DDOs, and many studies underline, as touched upon in the previous section, that among the crucial factors for organizations to succeed in becoming data-driven is a stimulating culture nurturing the technological initiatives made possible by data (see e.g., Cao and Duan, 2014; Mikalef et al., 2018). The specific nature of an organizational culture built around the various uses of data has in the literature on DDOs been defined as either "data culture" or "data-driven-culture".

A general definition of the data culture concept explains it as encompassing the cultural norms, value systems, and beliefs that shape and guide people's data-related practices (see Bates, 2017). These practices include data production, processing, distribution, and utilization, as well as their endeavors to govern and influence specific forms of data practices through various social and technical approaches. Similarly, applying the concept to organizational phenomena, Kremser and Brunauer (2019) state that a data culture refers to the collective beliefs, values, attitudes, and behaviors within an organization regarding the understanding, interpretation, and use of data. Additionally, it encompasses the mindset and practices that prioritize data-driven decision-making, data literacy, and the integration of data into everyday business operations. Similarly, Marr (2021: 227) suggests that business leaders should aim to "create a strong data culture across the company, with data being recognized as a key business initiative". The data culture term has specifically gained traction in business whitepapers and in the fields of media and communication (Burgess et al., 2022), perhaps due to the fact that the concept can encompass a wide range of practices and cultural factors. Consequently, the term data culture is generic and can be applied to most organizations that gather or use data as part of their operational activities (Kremser and Brunauer, 2019).

When referring to the specific data-related capabilities of organizations striving to become data-driven, however, Kremser & Brunauer (2019) argue that the data culture concept is too generic to grasp the specific cultural factors necessary. Rather, there is a certain quality of data culture better fitting for describing the situation at hand, namely *data-driven culture* – sometimes abbreviated to DDC (e.g., Cao and Duan, 2014; Chatterjee et al., 2021). The data-driven culture concept has gained significant attention in both academic research and practitioner literature and refers to the organizational mindset and practices that emphasize the use of data in decision-making processes and overall operations and can be defined as "a pattern of behavior and practices by a group of people who share a belief that having, understanding and using certain kinds of data and information plays a crucial role in the success of their organization (Kiron et al., 2013: 18). Through our literature searches, we particularly paid attention to three themes in terms of what can contribute to developing a DDC which were suitable for discussing our empirical material, the first being the importance of leadership support: Leadership plays a vital role in promoting a data-driven culture, as leader support, commitment, and active participation in data-driven initiatives set the tone for the organization's data-driven practices (Brynjolfsson & McAfee, 2014). The second central element we paid particular attention to, and that also seemed to dominate the literature, was data literacy and skills: A data-driven culture requires employees to possess data literacy skills, including the ability to collect, analyze, interpret, and communicate data effectively (e.g., Chen, 2020).

Lastly, data accessibility and infrastructure are, naturally, vital to foster DDCs, and were central subjects for our object of study. Organizations need to establish a robust data infrastructure, including data storage, integration, and analysis capabilities, to facilitate easy access to relevant data across the organization (Raghupathi and Raghupathi, 2021). Importantly, however, the full accessibility of data will only be facilitated if the organizational culture underpins cooperation across the company through "cross-functional collaboration" (Yu et al., 2021). Integral here, then, is a focus on seeing the whole supply chain of data across administrative borders. As Berndtsson et al. (2018) state, after looking into the enabling factors for data-driven culture at an international organization in the ferry business; "a data-driven culture will help organizations to have a holistic perspective on their intentions to scale up the usage of advanced analytics". Anderson (2015: 211) also includes other, but related, factors in his outline of what a DDC comprises. For instance, he points to the importance of establishing an "inquisitive, questioning culture", entailing "an atmosphere of healthy debate where one can ask for additional information, challenge assumptions, and discuss recommendations or additional tests", insights central to the current paper.

Adopting a data-driven culture offers various benefits to organizations, and two central factors here are, according to the DDC literature, improved decision-making and enhanced innovation and agility. As argued by e.g., Trieu et al. (2019), data-driven decision-making enables organizations to make informed and evidence-based decisions, leading to higher accuracy and better outcomes. Additionally, by leveraging data, organizations can identify emerging trends, customer preferences, and market opportunities, thereby fostering innovation and enabling rapid adaptation to changing business environments (e.g., Triaa et al., 2016). For instance, in interviewing 456 employees in Indian firms defined as having incorporated a DDC, Chatterjee et al. (2021) also found that the data-driven culture context is crucial for fostering product innovation and competitive advantage in the current business environment.

In terms of challenges, organizational resistance, and management resistance to change, are central factors in the literature discussing company culture and data-driven organizations. For instance, organizational conflicts and a lack of organizational readiness can impede the adoption of a data-driven culture (Farzaneh et al., 2018). To overcome these hurdles, effective change management strategies and communication are necessary, and as e.g., Storm & Borgman (2020) argue, determining the maturity phase of the DDC is vital in order to successfully address the challenges. In their case study of six organizations transitioning into becoming a DDO, these authors also identify that the inflexible nature of organizational structures can pose a significant challenge when it comes to effectively changing employee behavior and hence establishing a thriving organizational culture around the use of data. The current study aims to bring further insights into the reasons and effects of such benefits and challenges.

3. Methodology

3.1. Context: background to the case and previous research on NAV

NAV's IT development has in the last two decades been widely studied due to being seen as a "global pioneer in incorporating new ideas about organizational architecture in software development" (Barbala et al., 2023). The organization has sought to modernize its IT systems since the establishment of the directorate in 2006, yet the development was slow until 2016. Hulda Bernhardt (2022) has investigated the digital transformation taking place in NAV between 2016 and 2020 and found that the main drivers of change within the organization include a redesigned organizational structure to support the formation of teams and product areas, a revised sourcing strategy with insourcing of services, a shift towards a modern application platform and adaptable application architecture, and a transition from waterfall to agile product development. During this transformation, NAV introduced "product teams" holding complete responsibility for the software products provided to end users, for example making digital solutions for citizens to apply for sick leave money.

The latter has been subject to much scholarly interest within the software engineering field, as NAV IT has evolved into an increasingly agile organization where autonomous teams play a crucial role. Investigating how NAV IT has incorporated agile methods in their different teams, Dingsøyr et al. (2022) refer to NAV IT as a “very large-scale agile development program” indicating the complexity of the organization’s many teams and wide-ranging responsibilities. They found that NAV has transitioned from a first to a second-generation large-scale agile development method, where the first phase combined agile methods with project management advice, while the second phase replaced project management advice with current software development ideas. This led to the formation of several – now 12 – autonomous teams organized by product domain. The result, Dingsøyr et al. conclude, was a significant increase in product deployments, from twice a year to daily. The transition from first to second-generation large-scale agile development had a major impact on coordination, where teams decided on their own coordination mechanisms, resulting in fewer intermediaries and reduced dependencies between teams. Vestues et al.’s (2022) case study from NAV also reports on the organization’s agile team constellations, focusing on agile data management practices and the necessity of exploring alternative approaches to manage analytical data for the organization. Whereas Vestues et al.’s investigation focused largely on the nascent phase of introducing distributed data management within the IT organization, the current study has a broader scope and seeks to examine the use and needs of data as seen from different parts of the organization. This paper hence seeks to provide new insights here, offering novel findings from an organizational socio-technical perspective.

Studies have also been conducted inside NAV aiming to gain insight into how quality requirements are captured and classified, as well as to learn from how the agile software development process changes as a consequence of increasing focus on product quality. For instance, Mohagheghi and Aparicio’s (2017) industry report detailed the implementation of quality requirements management practices in NAV over the course of three years, finding that among the main challenges they faced was the complexity of the IT landscape, coupled with functional and technological diversity. Additionally, the organization faces ongoing pressure to deliver new functionalities while effectively managing maintenance costs. To gain better cost and quality control, NAV has because of Mohagheghi and Aparicio’s study increased its involvement in the development phase, aiming to avoid ad-hoc prioritizations.

NAV has also been a subject of interest for scholars looking into how the public sector is becoming increasingly data-driven. Lisa Reutter (Reutter, 2022) followed a NAV data team as part of the division for Data and Insight between January 2018 and May 2019 to study how they attempted to find ways to use data and machine learning to predict sickness absence among citizens, aiming to be of help for caseworkers working in NAV customer support centers. Reutter found that there were organizational constraints within NAV hindering the team’s success, such as a wide administrative scope and a fragmented organizational structure that complicated data access, due to differences in legal mandates between the municipal entities and the central level at NAV. She also found that the team spent a significant amount of time “getting to know” the data before they could start the analysis, due to needing to know how and why the data had been collected. These findings were mirrored in a recent study enquiring into the challenges product teams in NAV meet when they attempt to collect, share, and analyze the data. Looking into the tactics attempted for overcoming the hurdles put in place by GDPR and other relevant data-handling regulations, Barbala et al. (2023) found in their study that teams within NAV also are at the forefront with regards to identifying how new team roles are needed in the strive for becoming data-driven. Studying a team that develops digital applications used by citizens in need of applying for auxiliary aids, the authors found that inserting a legal expert on the team significantly shortened the time to solve privacy-related issues and decreased the deployment time.

Over the years, NAV has become an increasingly digitalized organization and data-intensive IT solutions are increasingly playing a larger and more important part in how the organization provides services to the public. Following this, the IT development has become a larger part of NAV, leading to a need to become a more agile organization. Currently, NAV IT consists of approximately 150 autonomous and cross-functional product teams. The product teams are organized in product areas, related to the main domains and areas of responsibility that NAV has, such as services related to pension management, or services that are supporting employers. These product teams are relatively stable over time where members become domain experts who are in contact with their end users and are constantly updated on laws and regulations.

Following this first agile transition, the IT development organization also had to transform the system architecture to be able to implement and deploy changes more frequently, removing dependencies that would slow down development and hamper the teams’ autonomy. The architecture changed from large monolithic systems to smaller and loosely coupled applications implemented as microservices, which enabled the product teams to increase the deployment rate dramatically from a few times a year to daily deployments.

3.2. Case study design

To investigate our research questions, we designed a longitudinal descriptive case study (Runeson and Höst, 2009), aiming to illustrate the process of creating a data-driven organization within the public sector. By gathering data over an 18-month span and revisiting the informants at different points in time, our goal was to outline the evolution and transformation of the technology setup and organizational norms. As far as we are aware, no such longitudinal studies on the implementation of a data mesh have been conducted in the public sector, and NAV stands out as the pioneer in Norway. NAV’s journey towards becoming a data-driven public sector company sets a precedent for other public firms considering a similar shift towards data-driven operations. Midway through the research, we presented our preliminary results at the XP 2022 conference, which provided valuable feedback that helped shape the remaining course of our study. Our intent is to continue exploring this topic in the future as NAV progresses further in its transformation into a data-driven organization.

Our unit of analysis is the organizational change in NAV, specifically focusing on the emerging cultures associated with transitioning towards a data-driven organization. The actors in this cultural shift can be categorized into two main groups: (1) The IT department, which is composed of developers, designers, and other roles that develop NAV’s applications and services. They provide IT resources to product teams in NAV’s various product areas, and to the data platform team that aids NAV in the evolution towards becoming data-driven. (2) The Statistics section in the Knowledge Department, whose main task is generating official statistics. These two groups represent different data needs and are composed of unique roles. From the first group, we focused on the product teams, which in NAV typically include roles often found in Agile literature; software developers, designers, product owners and data scientists. The Knowledge Department consists of data scientists, statisticians, data engineers, software developers, privacy experts, and managers. Initially, our study aimed to look into the assumption that the development of a data platform could succeed the long-standing data warehouse. However, as the research progressed, it became clear that

NAV would need to maintain both technologies to realize its vision of enabling product teams to manage their own data and keep producing statistics. This revelation allowed us to inquire into why and how NAV would allow these two solutions to coexist and the emerging interdependencies between them.

3.3. Data collection

Empirical data for the study was gathered from two primary sources: Semi-structured interviews and documents. In order to holistically capture the varying elements of the shift from a centralized to a decentralized data management approach, we selected respondents from two distinct and particularly data-intensive sections of the organization: 1) The Knowledge Department – which produces statistics and insights, and 2) the IT department, where we interviewed people placed in product teams and the data platform team. The rationale for selecting these specific respondents was their comprehensive coverage of the different roles that partake in data analytics within the context of NAV. Roughly, it can be contended that the Knowledge Department acted as data consumers, the product teams served as data producers, whereas the data platform team developed and operated a platform facilitating data exchange, however this was not always the case. A comprehensive distribution of respondents across these diverse team types is provided in Table 1.

A total of 31 semi-structured interviews with 24 different people were conducted as part of the first round of research. Out of these, 17 were audio-recorded and subsequently transcribed. Each interview spanned a duration ranging from 30 to 60 min. In instances where recording was not feasible, one researcher posed the questions while another made meticulous notes. A snowballing approach was employed for participant recruitment, whereby one respondent would recommend another (but the researchers decided whom to approach based on role and context). Predominantly, we were directed towards respondents who were acknowledged to have updated knowledge, competency, and interest in themes pertinent to our study, such as the construction of the data platform, domain teams that are early adopters, data scientists seeking data, managers of groups influenced by the data mesh initiative, etc.

Furthermore, after the analysis of the 31 interviews, we conveyed our discoveries to the Head of Statistics in the Knowledge Department and one (out of two) Principal Engineer at the IT Department, followed by a paired-depth interview with both parties, hence representing the two units of analysis in our study. We thus conducted 32 interviews altogether. The two stakeholders discussed their respective perspectives, identified mutual points of agreement, and agreed on principles for NAV's future direction. We perceive our research as an instrument to articulate dual narratives of the same context in a manner that enabled the case company to interpret and comprehend each other's viewpoints. The paired-depth interview was recorded and transcribed.

Documents served as our second significant data source. The documents analyzed encompassed project steering documents, descriptions of the new data strategy (as proposed by the NAV IT department), online documentation of the data platform (available on GitHub), NAV's IT ambitions, and conference presentations delivered by members of the development organization (such as the Norwegian JavaZone conference and the Data Mesh Podcast[2]). Several of these presentations have been recorded and are publicly accessible online, offering valuable insights into the public-facing aspects of NAV's IT and data strategies.

Table 1
Overview of interviews.

Parts of the organization			
Role		1) Knowledge Department	2) IT Department
	Data scientists		2
	Data analysts	1	
	Developers	5	5
	Privacy coach		2
	Team lead	3	
	Product owners		6
	Managers	3	5
		12	20

3.4. Data analysis

The data analysis can be understood as a three-step iterative process (Pan and Tan, 2011). The initial step involved exploring relevant literature to help us understand the phenomenon of interest. The literature that we found to be especially useful and informative for our study were presented in Section 2. Initially, our point of interest was on the open data literature, but through our fieldwork, we became aware that NAV had a focus on enhancing internal data sharing. The logic was that effective external data sharing necessitates efficient internal data sharing first. As such, we redirected our attention to internal data management, particularly the data mesh concept, and how these factors affected the organizational culture and collaboration across NAV.

The second step of the data analysis involved examining the data through a manual coding process. To ensure the reliability of our findings, we conducted inter-coder reliability checks, where multiple researchers independently coded a subset of the data and compared results to achieve consensus. Additionally, we performed member checking by sharing preliminary findings with a subset of participants to validate our interpretations and refine our themes, which our last interview was an example of, as explained above.

The coding was done manually, with printed transcripts and notes shared among three researchers after an individual round of coding. Some of the codes that surfaced during this process included terms like "data product," "data platform", "ownership" and "cultural tensions due to data-driven initiatives". We then printed three examples of each of the interviews and manually cut out sections of text using scissors and pinned these excerpts in thematical groups on a board. Sections of text that illustrated or explained the data mesh principles were extracted and grouped into categories, each given a descriptive title or code. Some text snippets ended up being coded under several categories, for instance, "data product" was coded both under "IT Department" and "Knowledge Department". The primary categories we ended up with were "Data management for generating statistics" and "Data management for developing software products."

The third and final step saw the merging of the inductively derived codes with concepts drawn from the literature. We noticed a significant overlap between our codes and the principles of the data mesh concept. For instance, this left us with three categories under "Data management for developing software products": "data ownership and products", "data platform", and "data governance". This process helped structure our understanding of the organization's interpretation and adaptation of the data mesh concept. Through this process and informed by some of the literature we had read up on at this point (see Section 2) we also noticed that some of the overlapping codes, although fitting with findings both from our interviews in the Knowledge Department and IT Department, were framed differently by the respective respondents. This part of the analysis led us to the realization that the organization could be argued to consist of two different data cultures, that both surrounded the categories *speed, users, law obligations, product and purpose*, and *distinct language (ontology)* when talking about and conceptualizing data handling, however in conflicting ways. We will return to these findings in 4.3.

In this study, we adopted an interpretive approach, understanding the world and its truths as subjective realities (Oates, 2005). This indicates our recognition of the potential for differing viewpoints among various informants regarding their interpretations of truth. Throughout the coding process, we consistently captured reflections in notes and compared them with the newly emerging codes, a technique known as the Constant Comparison Method (Seaman, 1999). What emerged as the central focus of the study was, as described above, the differences between the IT Department and the Knowledge Department, and to verify our results and gain further insights into these matters we found significant value in the aforementioned paired-depth interview with the two central figures representing the two segments of the organization. This, as well as the presentation of findings to the two departments, acted as *member checking* in order to validate our results.

4. Findings

Our analysis revealed that NAV in the last few years has seen the flourishing of two separate data cultures, emerging from the different needs and uses of data at two separate segments in the organization that are both particularly data-intensive, meaning that leveraging data is central to most of their ventures. These differences had caused friction and challenges in terms of "cross-functional collaboration" (Yu et al., 2021) across the organization, in turn halting the strive towards becoming data-driven. The paired-depth interview with the two leaders representing each culture revealed however that the two different segments in the last year have started outlining some common goals and principles in terms of data leverage. We thus interpreted this as the two organizational segments now working together on strategically building a *common data-driven culture* for the organization, attempting to fuse what had become two seemingly *opposing data cultures*. In order to best reflect the chronological order of the story, the following section is divided into two parts: The first tells the stories of the two "data camps" and the second part presents the tensions that had arisen due to the emergence of the different data cultures in the organization. Our findings from the first round of interviews showed that these cultural differences were due to inconsistencies in terms of the following factors: Speed, law obligations, users, products, and domain-specific data language (i.e., data ontology), which were confirmed by the paired-depth interview with the Principal Software Engineer and the Knowledge Department Manager we interviewed in round 2. We will explain how the different factors differ in detail in what follows.

4.1. Story of NAV product teams

To maintain short iteration cycles and continuity, the product teams that were introduced during the rehaul of NAV IT, as outlined in Section 3.1, soon required greater control over the applications they developed in the different product areas in NAV. The need for frequent and rapid changes often necessitated modifications to the data structures, which, since the data resided in a centralized data warehouse, required the teams to submit change requests to the data warehouse team. However, with many teams making such requests, a substantial backlog formed, resulting in weeks of waiting before the data warehouse team could implement the necessary changes. Consequently, the progress of the product teams came to a halt, leading to a loss of speed and momentum. Although we had some information about this through our previous study (Vestues et al., 2022), the interviews amongst product teams in the current research further highlighted this issue, and we asked the study participants to describe these challenges in order for us to gain a deeper understanding.

As our interviewees explained, to address this challenge the product teams began retaining the data within their own small databases. By doing so, they maintained control over their data and could make changes as needed to align with the speed of development. Simultaneously, they could provide data to the data warehouse team. This approach was an important step toward empowering the product teams to control their processes completely. However, it also required them to acknowledge the presence of "data users" in addition to their regular user base. This resulted in confusion regarding the main purpose and tasks for the teams. One product owner we interviewed articulated this lack of clarity as such:

> *[I]n the last year, I feel we have started to look at whether this makes any sense anymore. Are we organized in the right way? What should we be as a team? When you ask me now "what do you do", I think many people have a bit of an identity crisis all around. What are we actually doing? What is it that is important for us to monitor in terms of measurements? Things are changing now, in terms of what we actually need to deliver. What is "nice to have", what are "must haves"? It can get a little fuzzy.*

The above quote points to how some members from the IT Department located in product teams experienced a "lack of holistic perspective" (Berndtsson et al., 2018) in the organization, where data-driven initiatives had unclear goals. We found similar signals in several teams we studied, indicating a need to rethink how they were organized and what should be their responsibilities, and ultimately also how they interacted with the wider NAV organization. We used this information when shaping the interview guide for the second interview round as it was crucial for the study to also have the leaders' views on these matters.

The new strategy for transferring ownership of data to product teams was inspired by the principles of data mesh (see Vestues et al., 2022), highlighting the need for viewing data as products leveraged by the product teams. One important argument for seeking inspiration from the data mesh concept is that data should be analyzed by those developing the product because the data will provide them with new insights helping them in developing an even better product. NAV believes that this will make the product teams more data-driven, meaning it will unlock the potential for making decisions based on data about the users.

With the shift in responsibility for managing their own data, the teams required infrastructure that would facilitate data collection, storage, sharing, and analysis. In line with this, a centralized data platform called NADA[3] was established to relieve the product teams from the burden of developing and maintaining such infrastructure themselves. The data platform is hosted in Google Cloud Platform, offers a data catalog and hence becomes a marketplace where data consumers can browse available data products. The primary objective was to streamline the process for product teams to assume ownership of their data. It was important to NAV that the utilization of the data platform was not enforced, but rather made more convenient than external tools so teams could keep their autonomy. *"We are here to support the product teams, not dictate their choices,"* a data scientist in the platform team said.

When presenting these comments to the participants in the last interview, the Principal Software Engineer commented, *"When we move from a few large databases that are rarely updated to many, many databases that are updated frequently, we have to do something to get the data out of the applications, into data warehouses. I think that is the main purpose of the data platform, to create a way to distribute data in a stable manner."* Consequently, the platform served as a means for teams to publish their data, facilitating convenient access to data from all teams through the platform. Additionally, the platform team offers support through tutorials and wikis. The ultimate goal, which also the Principal Engineer underlined, is to simplify the utilization and sharing of data for product teams, eliminating the need for specialized data scientists. However, the product teams are still in the process of understanding the potential uses and management of their data. For instance, knowing what data is needed and which data can bring value for their own and other teams caused confusion among the different teams. A software developer working on a sick leave data product said:

> *The problem is that I don't always know what kind of data I need. But I know who is most likely to know what data I need. [...] It's not that it's insoluble, and I think it's worth trying to get everyone to put operational and analytical data into the platform. To try it and see if it works. I'm just a bit unsure if it actually provides enough business value.*

It was clear during the data collection period that the data platform was a new concept in NAV and the new responsibility of the product teams as data product owners was still unfamiliar. However, as the above quote exemplifies, several of our informants were under the impression that the only way to find the best solutions was to gain practical experience to not only see whether they would achieve the wanted improvements but also to learn about any disadvantages or challenges and potentially discover new and innovative solutions for managing and sharing data.

One of the significant challenges the product teams encountered when they took on the responsibility of managing their own data, was determining whether commercially offered tools complied with GDPR regulations and internal company policies. For instance, they needed to consider if using tools like Google Analytics was acceptable, as there were concerns that data might have been stored in the USA, potentially violating the restriction of storing data outside of Europe (according to EU regulations). To address these compliance concerns, the platform team ensured that the tools and applications provided through their platform met the necessary requirements. As our interviews clearly indicated, this had instilled confidence in the product teams, relieving them of worries about compliance issues. Consequently, valuable time was saved, allowing them to focus on leveraging their data to enhance their products instead of being consumed by data management concerns.

As our analysis showed, the implementation of the data platform NADA brought about a significant shift for both the product teams we studied and the teams responsible for managing the data warehouse at NAV. The product teams gained a newfound level of autonomy as they were no longer dependent on NAV's centralized data warehouse. However, this change not only impacted the product teams but also posed challenges for the Knowledge Department. With the product teams becoming more independent, a crucial challenge arose; namely ensuring a continuous flow of data from the product teams to the data warehouse for statistical purposes, which we were given various examples of in the different teams we interviewed. While the product teams were now responsible for managing their own data, the Knowledge Department still needed to receive data from them to facilitate the creation of comprehensive statistics, but where it became more challenging to control the data sharing. A director on the data warehouse side explained:

> *One cost has emerged that we call "involvement". In the old days, my employees would simply deliver an order to the source [product teams]: "We need this from you!". Then it would be put together and delivered to us. This has become a lot more complicated now.*

This shift in dynamics presented a new hurdle that the Knowledge Department had to address and that was not initially seen as a concern for the product teams.

4.2. Story of the knowledge department

Seven teams in NAV's Knowledge Department are tasked with the crucial responsibility of delivering official statistics on various aspects such as sick leave, pensions, employment rates, and more. These statistics play a vital role in providing information to politicians and the public, as well as supporting internal management functions within NAV. Not only are these statistics mandated by law, but they also serve as the foundation for the Norwegian national budget. Each team functions as a multi-disciplinary unit, specializing in statistics within a specific domain that corresponds to NAV's product areas. As of August 2022, there were five data domains, which were managed by seven teams: Family, Sick leave, Pension, Casework and Task ("Sak"), and Employment. Within each domain, the Knowledge Department receives data from the corresponding product teams. Before incorporating the data into the data warehouse, they meticulously assess its quality and make structural changes to make sure that it complies with historical data. Core roles within the teams include data scientists and data engineers. Their combined expertise contributes to the successful generation of the desired statistics.

What our previous study pointed to (Vestues et al., 2022) became even clearer after the current study's interviews; at first when the product teams were entrusted with the responsibility of managing data, the Knowledge Department regarded it unproblematic, and they were content if the product teams continued to provide the necessary data for statistical purposes. The Head of Statistics at the Knowledge Department reflected on this period, saying, *"Back then, I tried to explain the distinction between operational data that is valuable to product teams and the more strategic data that we require for statistical purposes."* However, a shift in attitude gradually emerged within NAV IT questioning whether the product teams could independently generate the required statistics now that they had direct access to the data. This created much dismay in the Knowledge Department, however, with the Head of Statistics explaining: *"No product team can for example create the Norwegian unemployment rate because you need data from several teams, and you need to define what unemployment is. This is a huge and complex job."* This points to the organizational differences with regard to *data ontology* and is an example of a clear finding in our last round of data analysis, showing variations in the definitions and linguistic representations of different data sets, which we will return to in the next section.

The unclarity of responsibilities and different uses and requirements of data thus created major tensions between the NAV IT product teams and the Knowledge Department. For instance, it ultimately led to a growing skepticism regarding the necessity of the Statistics section amongst some product teams. A manager at the department described this change, saying, *"It manifested as a lack of willingness to share data,"* indicating that some product teams hesitated to collaborate and share their data with the data teams. Members of the Knowledge Department reported that data extraction activities were often postponed or neglected by the product teams, as their priority was focused on development tasks that directly provided value to their own users. A deputy manager at the Knowledge Department also pinpointed how they got the impression that employees at NAV IT viewed them as not up-to-date or open to innovation and new technological solutions: *"[The attitude is that] what's new is exciting, so we are becoming legacy."*

In the past, when all data had to pass through the data warehouse, the Knowledge Department held a gatekeeping role and had the authority to instruct the product teams to make necessary changes. However, with the product teams no longer dependent on the data warehouse, this dynamic had shifted. Instead, the product teams could now publish their data on the NADA data platform, where NADA would be solely responsible for providing the platform rather than ensuring data quality. Consequently, the Knowledge Department was preparing to increasingly meet challenges such as getting incorrect or inadequate data that failed to meet their specific requirements, which would be especially detrimental if they soon would be forced to source data from NADA. This was a concern that both leaders and software developers clearly articulated in our interviews. As the data mesh structure had not been in place for more than a few months at the time of our interviews, this scenario was yet to be determined.

Expressing the uncertainty of the data mesh structure for their statistics operations, a Knowledge Department manager said, *"We don't know what [data] we're getting, and in many cases, we get the wrong data. This happens because [the data sharing system] is set up in a way that no longer meets our requirements and needs."* In this new setup, the Knowledge Department was required to request data from the product teams without any leverage or incentives to enforce specific data quality standards. The manager highlighted the magnitude of this challenge, explaining, *"I only have 20 people in my [data warehouse] teams to cover >150 development teams."* This limited capacity made it difficult to demand a certain level of data quality from the product teams who transfer data using the NADA data platform. Commenting on this, the deputy manager added: "*[...] The data mesh platform does not support our needs in seeing the information cross sectionally."*

When creating statistics, the Knowledge Department relies on historical data, meaning they must collect the same data consistently to analyze trends over time. The data must remain consistent across long periods, not allowing anyone to make changes to previous years to ensure reliability. A manager highlighted the importance of data consistency by explaining, *"When we put data into the data warehouse, we lock it to prevent any modifications."* They cannot allow the product teams to make any changes to this data, which they do by keeping their data warehouse and thus the control over the data. This ensures that data remains reliable for creating statistics. Due to the critical nature of maintaining control over data and its structure, the Knowledge Department could not solely rely on the NADA data platform as they need to have full control over the data within their domain to ensure its trustworthiness. While each NAV IT product team may share data in their preferred structure on the platform, the Knowledge Department must secure its interests to fulfill its mission. This presents a significant challenge to NAV IT's data mesh-inspired solution, as it does not function for the crucial data operations of the Knowledge Department.

Nevertheless, there also seemed to be agreements among several employees in the Knowledge Department that modernization was also needed on their part, which several interviewees mentioned. As an example, a manager in the statistics team said:

> *We must find an appropriate way to modernize the data warehouse. And that includes more than finding a new storage medium. [But] there is an entire ecosystem that must be taken care of and sources that are located in many places. [...] If NADA meets all the requirements, we don't know. No one has actually looked at it, but it has not been selected based on the requirements we have for a storage medium. Maybe NAV is a type of organization where we must have both [solutions].*

This quote points to an opinion that soon would become a widely held consensus amongst both segments of NAV, as we will examine in the following.

4.3. Two differing data cultures: organizational tensions and emerging solutions

The table below presents what our analysis of the interviews identified as cultural differences between the two segments of NAV, the product teams connected to NAV IT and the Knowledge Department, rooted in their differing uses of data. We hence interpret these as making up two *conflicting data cultures* in the NAV organization, and our findings clearly show how tensions had arisen due to these differences. These tensions were both directly articulated, as exemplified in the previous section, and found indirectly as differences when coding the interviews, as we noticed that, despite similar topics, these themes were framed differently in the two segments of the organization. As the second part of our data collection revealed, however, representatives from NAV IT and the Knowledge Department had recently been in talks after realizing they would have to address and ease the existing frictions between them for the NAV organization to successfully continue its data-driven journey. The following will detail some of these tensions rooted in the factors outlined in Table 2, which is inspired by findings in a Master thesis one of the authors supervised in 20,23[4].

Contrasting time horizons (a) led to a significant difference in what data was valuable to the product teams and the Knowledge Department respectively, with their differing objectives playing a crucial role. The Knowledge Department needs consistent data across years and even decades to deliver reliable statistics, while the product teams rely on short development iterations; days or weeks, often having to change the data they collect to facilitate the constantly changing software. As they continuously deploy code, they collect data to monitor user reactions, and the data they base their development on will often change.

As our interviews and subsequent analysis clearly showed, this disparity in time horizons created agitation, as ensuring data consistency demands resources from the product teams without providing them with direct benefits back. Developers, who primarily focus on creating services that meet user needs, often remained unaware of the data requirements for statistical analysis, a concern for the Knowledge Department. Investing time and cognitive effort to understand these needs was seen as a distraction from their primary goal of developing software for end users. Consequently, developers perceived this as time wasted and prioritized their attention and efforts elsewhere. Convincing product teams to keep collecting and sharing data that is not necessarily valuable to them became the job of the Knowledge Department.

Table 2
Sum-up of differences between the two data cultures causing organizational tensions.

Factors	Product Teams	Knowledge Department
a) Speed	Inspired by agile methodologies; short development iterations to serve continuous deployment.	Long-term planning is crucial to ensure consistency across years and decades for reliable statistics.
b) Law obligations, product and purpose	No specific law obligation but provide user-friendly and correct data for building and improving digital public services.	Obliged by law to deliver official and public statistics to governmental organs.
c) Users	Other teams in Product Areas in NAV providing services to the public.	Governmental organs, news media, researchers, etc.
d) Language (Data ontology)	Technological understanding of data use and data products.	Legal and societal understanding of data use and data products. Due to the importance of correct statistics, they need specific linguistic representation of the data they use.

Terms like "inside data" and "outside data" appeared as a way of grasping this difference in time horizons. Inside, the product team could make whatever changes they wanted to their data as long as they kept the outside data they offered to other teams through the data platform stable. *"You had to maintain two [data] models with two different rates of change, which is a complicated issue,"* the Principal Software Engineer we interviewed explained.

Unclarity about each other's roles and mandates (b and c) was another central difficulty causing friction. We found this to be particularly clear in the interviews with leaders in NAV: Maintaining both a data warehouse and a data platform posed challenges and raised questions about their coexistence in the organization. The introduction of the NADA data platform was perceived by some as a potential threat to the data warehouse's role. The fundamental question emerged: Why should NAV maintain both a data platform and a data warehouse? This perceived threat intensified the tension between the data platform and NAV IT on one side, and the Knowledge Department on the other. The tension led to members of the Knowledge Department feeling unwanted and unrecognized by the product teams, while the product teams sought to avoid any involvement. Consequently, finding a way to fulfill NAV's common mission of providing high-quality services to Norwegian citizens became increasingly challenging.

This unclarity about roles and seeing how each party fitted into the wider NAV puzzle was further complicated by *their differing data language and definitions* (d): Where the product teams are mainly made up of technologists such as developers, designers, and others with tech backgrounds, the Knowledge Department consists of employees whose mandate requires a legal and societal understanding of data use and data products. For instance, was this difference something we found in our last round of analysis, when noticing that words connected to law and regulations were common when talking about data handling in the Knowledge Department, but not amongst NAV IT employees, who oftentimes rather used words from popular tech books, such as *Inspired* by Marty Cagan, when talking about data use, sharing and storing. For the Knowledge Department, use of language had to be specific: Due to the importance of correct statistics, they need specific linguistic representation of the data they use. As the Head of Statistics described it, *One example is the unemployment statistics in NAV, which is one of Norway's most important statistics. [...] And they are made following very specific criteria and definitions, which are established in Norwegian law. Among other things, it uses data from report cards, because there are several definitional prerequisites for whether you should be defined as "unemployed" or not. A few years ago, we experienced a product team saying "we don't see the point of report cards, because we don't need them. We are not interested in knowing whether the user has applied for any jobs." If we did not know this, the unemployment statistics would not have been possible to produce afterward.*

Similarly, few incentives existed for product teams to publish data on the data platform to share with others across the organization. To initiate the adoption and utilization of the data platform, NAV implemented a strategy they called "cake-driven development." This involved rewarding product teams with cake when they published a data product on the platform. Pictures were taken and shared on common Slack channels, creating attention, and encouraging other teams to follow suit. We were shown evidence of this from a software developer working on unemployment products, who showed us how these celebratorily Slack posts could look in order to create optimism and enthusiasm amongst the teams. However, there were limited real incentives for product teams to invest time in customizing the data they published to align with the needs of data consumers, such as the Knowledge Department. Consequently, product teams found themselves relying on data from other teams, which was rarely tailored to *their* specific requirements. This necessitated reaching out to the respective teams, requesting the data, and specifying their individual needs. As a result, the concept of a data platform intended to replace the need for extensive coordination between teams utilizing each other's data was undermined.

As the paired-depth interview we conducted with the Head of the Statistics division in the Knowledge Department and the Principal Software Engineer in NAV IT revealed, however, a common understanding of the need to solve the tensions between the two sections top-down had evolved over the past year. *"We have transitioned from childish disagreements to having mature conversations and finding common ground,"* stated the Head of Statistics. Despite their differences, the two representatives emphasized their shared commitment to contributing to NAV's overarching vision. They both agreed on the coexistence of the data platform and the data warehouse, recognizing the vital roles played by each in facilitating speed and continuity for product teams and generating essential statistics for the Norwegian public.

5. Concluding discussion: towards a common data-driven culture

As our findings demonstrate, there exist inherent and fundamental differences between the developer community within NAV IT and the Knowledge Department, as described in Section 4. These differences have given rise to distinct and contradicting data cultures within the two organizational sections. We raise the question of whether these challenges are underestimated or overlooked in the existing literature on data mesh, data-driven development, and data platforms, where there seems to be an overemphasis on the belief that implementing technology alone can resolve an organization's wish to utilize data and developing better software even faster. We thus underline the importance of utilizing a holistic socio-technical framework to discuss the subject matter, considering the defining characteristics of data in shaping organizational culture and strategies.

Our findings offer new insights in terms of what can contribute to developing a data-driven culture, especially for the following three topics, which also was the focus in our literature review as presented in Section 2: *Leadership support, literacy and skills,* as well as *data accessibility and infrastructure.* As the last two factors are fairly self-explanatory, we, for instance, see how the establishment of the NADA platform already has contributed to new and innovative ways of sharing data between actors in NAV, and how it is vital to inhabit knowledge about how to best exploit the value of this data, we will in this section mainly focus on how *leadership involvement* can be vital for solving organizational hurdles in large-scale data-driven project. We argue, however, that data literacy and skills as well as data accessibility and suitable infrastructure are vital in order for leadership initiatives to be fruitful, which also were clearly articulated in some of the quotes presented previously in the paper.

As the last part of our analysis revealed, the dissonance between the two emerging and conflicting data cultures made the Knowledge Department conclude that product teams do not have sufficient *data literacy* needed for a holistic organizational data-driven operation (see Kremser and Bruneuer, 2019), even though they are rigged as domain-oriented product teams. Producing statistics on e.g., the unemployment situation at a national level requires an overview of the whole organization that specialized product teams do not have, as explained in Section 4.3. Also, producing such statistics and insights requires strict control of the data quality where data must be stable over time to ensure that recent data can be used in combination with historical data, e.g., to show development and trends correctly over time.

The interviews within the product teams pointed to that the data platform that was established for their purposes had functioned fairly well, where teams eventually offered data products via this marketplace, and where others found and used them. This, however, seemed to have become a one-sided process where the needs of the operational part of the organization outweighed the needs of the analytical part, i.e., the Knowledge Department, even though both these are clearly dependent on access to data. The new data platform was developed in parallel with the established data warehouse, but as members from both sections pointed to in the interviews, these two were poorly coordinated, both technically and organizationally. This created a level of mistrust between the two sides. This story thus illustrates well the consequences of not being able to establish cross-functional collaboration (Yu et al., 2021). This level of mistrust became evident over time for both sides and has created a joint understanding that both data units – platform and warehouse – must co-exist since they cater to different data needs.

That central leaders in the two segments of the organization had decided to start talks to ease the tensions between them points to a central argument in the literature about data-driven culture, namely the vital role of leadership in supporting and committing to data-led initiatives (see e.g., Brynjolfsson & McAfee, 2014). This became clear to us only in the last interview round, due to choosing to interview the two leaders together, leading to them oftentimes having a conversation between each other, not only merely answering the researchers' questions. The solutions emerging because of leadership engagement in the two parts of NAV can hence be described as *a top-down reaction to the growth of a bottom-up data culture*. For instance, the norms and practices of data management and sharing within the product teams were shaped by the autonomy given to them. As these teams were responsible for developing software for their users, they were also granted freedom in dealing with data. This autonomy was further enhanced when they became independent of the data warehouse. Consequently, the product teams developed their own practices of not prioritizing data sharing, as it did not directly contribute to their development tasks. This observation underscores the potential challenges that can arise in an organization where autonomous teams have accountability for delivering products or services. When data sharing is approached from a bottom-up perspective, there is a risk that teams prioritize their own needs above those of others, resulting in various tensions between different parts of the organization. To address these tensions, top-down initiatives may be necessary to establish guidelines and promote a more balanced approach to data sharing and collaboration, as we concluded after analyzing the second interview round with the two leaders. This mirrors the findings in recent research by David Holger Schmidt and Dirk van Dierendonck (Schmidt et al., 2023) who developed a competency framework for organizational leaders, arguing that data-driven leadership must "combine task-oriented competencies with relationship-oriented competencies". It also points to what Brynjolfsson and McAfee (2014) have named a "management revolution", indicating how data, and a shared perception of what data is and how it can be used, must be central to leadership decision-making to succeed with data-driven initiatives.

To further link our findings to recent literature, it can be argued that the data-mesh-inspired data culture in the NAV IT department and the product teams emerged as a result of seeing the importance of *democratizing data* (see e.g., Dehghani, 2020), where the teams were given the additional responsibility of providing data as products. Yet, as some of the interviewees articulated, it was unclear who would be the *consumer* of such products. This unfolded as a bottom-up process with deliberately no direct leadership or support beyond offering a new data platform (NADA) where teams were encouraged to share data and subsequently rewarded. This loose and facilitation-oriented leadership approach allowed the teams to be creative and to decide for themselves what the data products should be, but it also enabled the two data cultures to emerge through a bottom-up process. As became evident through the interviews, however, this created a top-down reaction from the Knowledge Department side that was not included sufficiently in the leadership of this process, consequently leading to their needs as data consumers were missed.

Hence, the present study also adds important lessons into how data mesh-inspired ecosystems may play out in practice in organizations where there exist different data needs. Although the data mesh literature (e.g., Dehghani, 2020) suggests that a data platform can replace a data warehouse and become the central component of an organization's existing software ecosystem, we can from our findings from the context of a large, complex public sector organization suggest that data mesh principles may benefit from a more nuanced view: Companies that aim to transition from a data warehouse to a data platform may find it necessary to adopt both solutions, which both the Head of Statistics the Knowledge Department and the Principal Engineer at the IT Department agreed on in our last interview. This might particularly be true for organizations that have obligations to produce statistics, whether due to legal requirements (like governmental organizations, banks, etc.) or managerial demands for data-driven decision-making. Such companies need an organizational structure that maintains control over their data and allows for its management to align with statistical needs. Our case study thus suggests that companies relying on autonomous product teams for software development should anticipate the implementation of both a data platform and a data warehouse. The data platform serves as a catalyst for enhanced development continuity and speed, complementing the existing infrastructure rather than replacing it. In the context of large governmental companies like NAV, our findings indicate that a juxtaposition of both solutions is necessary.

Additionally, our analysis revealed the emergence of two distinct data cultures in different phases, as explicated in Section 4.3. As our interviewees pinpointed, initially there had been a chaotic phase where product teams grappled with an understanding of what it meant to manage their own data, how they could do it, and what technology they needed. Simultaneously, the Knowledge Department faced the challenge of adapting to the loss of its gatekeeping role and influence over the product teams.

In the second phase, NAV IT began to explore ways to effectively utilize the product teams' data, including analysis and sharing practices. The introduction of the NADA data platform supported these efforts within the product teams. However, the Knowledge Department encountered issues with declining data quality and had to invest more time in convincing the product teams to provide the necessary data. This led to growing frustration on both sides. These findings were connected to the longitudinal nature of the study, since we were able to follow the evolution of the implementation and use of NADA. As became evident in the last phase of the data collection, top-level managers from both sides eventually converged on a shared strategy of implementing top-down data requirements. As these requirements are put into action, at the time of writing, there is a recognition that the data platform and the data warehouse serve distinct purposes and must coexist. This recognition seems to grow from ongoing dialogues between the managers that enable insight into each other's data needs – something that had not been clearly communicated between them at an earlier stage. Hence, even though the rapid implementation of the new data platform did cause some disharmony, it has also been a fruitful learning process. This mutual recognition marks the beginning of a new phase aimed at consolidating their coexistence and finding ways to harmonize their operations in NAV.

As previous studies also have reported (see e.g., Anderson, 2015), building a data-driven culture must be done with the whole organization in mind. The analysis of our interviews, particularly with regards to the four factors discussed in 4.3, clearly underlines how organizational conflicts, and a lack of organizational readiness, can impede the adoption of a data-driven culture. Due to the many and multifaceted actors involved, counting many different departments and disciplines in organizations such as NAV, it can be argued that a holistic, strategic vision is especially important in very large and complex public sector organizations, where leveraging data is a requisite for being able to successfully serve citizens for the future.

6. Limitations and future research

This case study is subject to several limitations, three of which are the most central ones. Firstly, due to NAV being a very large and complex organization with approximately 22 000 employees, gaining a proper overview of the different actors in the data ecosystems is challenging, and hence it can be said to be a limitation of the study that only a fraction of the employees in the organization was interviewed.

Secondly, as we are here focusing on a Norwegian context, both the organizational cultures, applicable law obligations and regulations, as well as the nature of the digitalization efforts are likely to be country-specific and not generalizable to other countries' public sector organizations.

Similarly, and lastly, the nature of the case study method represents methodological limitations, as case studies are not designed for statistical generalization (see Runeson and Höst, 2009). However, it's worth highlighting that this study allows for *analytical* generalization: It offers an approach for investigating how the emergence of desynchronized data cultures within the same organization causes tensions and ultimately contravenes data-driven initiatives, an enduring challenge that likely applies to large, complex organizations globally seeking to become increasingly data-driven.

We will continue to follow NAV's data-driven initiatives. For future studies, we aim to enquire more deeply into how two of the desired benefits of a data-driven culture, namely *improved decision-making* and *enhanced innovation and agility*, play out in the organization following their recent focus on establishing a common, data-driven future.

Notes

1. https://www.regjeringen.no/en/dokumenter/one-digital-public-sector/id2653874/
2. https://daappod.com/data-mesh-radio/
3. The name NADA means 'NAV Data' but is also a tongue-in-cheek name indicating that the platform holds no data itself but is rather built as a marketplace where users can announce and identify where to find data from other sources.
4. "Exploring Institutional Complexity: The Interplay Between Views on Data, IT Usage, and Collaboration in a Large-Scale Agile Environment" (2023), Master thesis by Ingeborg Westborg and Steel Sofie Hovde Bragnes, where one of the authors was a co-supervisor.

CRediT authorship contribution statement

Astri Moksnes Barbala: Writing – review & editing, Writing – original draft, Visualization, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Geir Kjetil Hanssen:** Writing – review & editing, Writing – original draft, Visualization, Validation, Project administration, Formal analysis, Data curation, Conceptualization. **Tor Sporsem:** Writing – review & editing, Writing – original draft, Visualization, Project administration, Methodology, Formal analysis, Data curation, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data that has been used is confidential.

Acknowledgements

This project was supported by the Research Council of Norway under grant number 321477. We are grateful to NAV who granted us permission to conduct interviews, review internal documents, participate in company events, and interact with employees.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.jss.2024.112185.

References

Anderson, C. (2015). Creating a Data-Driven Organization: Practical Advice from the Trenches (1st edition). O'Reilly Media.
Awasthi, P., & George, J. (2020). A case for Data Democratization. AMCIS 2020 Proceedings. https://aisel.aisnet.org/amcis2020/data_science_analytics_for_decision_support/data_science_analytics_for_decision_support/23.
Barbala, A., Sporsem, T., & Stray, V. (2023). Data-Driven development in public sector: how agile product teams maneuver data privacy regulations. In C. J. Stettina, J. Garbajosa, & P. Kruchten (Eds.), Agile Processes in Software Engineering and Extreme Programming (pp. 165–180). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-33976-9_11.
Bates, J. (2017). Data cultures, power and the city. In R. Kitchin, T. P. Lauriault, & G. McArdle (Eds.), Data and the City (1st ed., pp. 189–200). Routledge. https://doi.org/10.4324/9781315407388-14.
Berndtsson, M., Forsberg, D., Stein, D., & Svahn, T. (2018). BECOMING A DATA-DRIVEN ORGANISATION. Research-in-Progress Papers. https://aisel.aisnet.org/ecis2018_rip/43.
Berntsson Svensson, R., & Taghavianfar, M. (2020). Toward Becoming a Data-Driven Organization: challenges and Benefits. In F. Dalpiaz, J. Zdravkovic, & P. Loucopoulos (Eds.), Research Challenges in Information Science (pp. 3–19). Springer International Publishing. https://doi.org/10.1007/978-3-030-50316-1_1.
Burgess, J., Albury, K., McCosker, A., & Wilken, R. (2022). Everyday Data Cultures (1st edition). Polity.
Cao, G., & Duan, Y. (2014). A path model linking business analytics, data-driven culture, and competitive advantage. https://uobrep.openrepository.com/handle/10547/332972.
Chatterjee, S., Chaudhuri, R., Vrontis, D., 2021. Does data-driven culture impact innovation and performance of a firm? An empirical examination. Ann. Oper. Res. https://doi.org/10.1007/s10479-020-03887-z.
Chen, Y., 2020. Data-Driven decision-making literacy among rural community college leaders in iowa: the role of leadership competencies. Community College J. Res. Practice 44 (5), 347–362. https://doi.org/10.1080/10668926.2019.1592032.
Dehghani, Z. (2022). Data Mesh: Delivering Data-Driven Value at Scale (1st edition). O'Reilly Media.
Dingsøyr, T., Bjørnson, F.O., Schrof, J., Sporsem, T., 2022. A longitudinal explanatory case study of coordination in a very large development programme: the impact of transitioning from a first- to a second-generation large-scale agile development method. Empirical Softw. Eng. 28 (1), 1. https://doi.org/10.1007/s10664-022-10230-6.
Evans, E. (2003). Domain-Driven Design: Tackling Complexity in the Heart of Software. Addison Wesley Professional, 1st Edition.
Fabijan, A., Dmitriev, P., Olsson, H.H., & Bosch, J. (2017). The evolution of continuous experimentation in software product development: from data to a data-driven organization at scale. Proceedings of the 2017 IEEE/ACM 39th International Conference on Software Engineering (ICSE), 770–780. https://doi.org/10.1109/ICSE.2017.76.
Farzaneh, M., Mozaffari, F., Ameli, S.P., Karami, M., Mohamadian, A., & Arianyan, E. (2018). Designing an organizational readiness framework for big data adoption. Proceedings of the 2018 9th International Symposium on Telecommunications (IST), 387–391. https://doi.org/10.1109/ISTEL.2018.8661073.
Fischer, H., Wiener, M., Strahringer, S., Kotlarsky, J., & Bley, K. (2022, December 12). From Knowing to Data-Driven Organizations: review and Conceptual Framework.
Fitzgerald, B., Stol, K.-J., 2017. Continuous software engineering: a roadmap and agenda. J. Syst. Softw. 123, 176–189. https://doi.org/10.1016/j.jss.2015.06.063.
Goedegebuure, A.A. (n.d.). Data Mesh: systematic Gray Literature Study, Reference Architecture, and Cloud-based Instantiation at ASML.
Hukkelberg, I., & Berntzen, M. (2019). Exploring the Challenges of Integrating Data Science Roles in Agile Autonomous Teams. In R. Hoda (Ed.), Agile Processes in Software Engineering and Extreme Programming – Workshops (pp. 37–45). Springer International Publishing. https://doi.org/10.1007/978-3-030-30126-2_5.
Kremser, W., & Brunauer, R. (2019). Do we have a Data Culture? In P. Haber, T. Lampoltshammer, & M. Mayr (Eds.), Data Science – Analytics and Applications (pp. 83–87). Springer Fachmedien. https://doi.org/10.1007/978-3-658-27495-5_11.
Loukiala, A., Joutsenlahti, J.-P., Raatikainen, M., Mikkonen, T., & Lehtonen, T. (2021). Migrating from a centralized data warehouse to a decentralized data platform architecture. In L. Ardito, A. Jedlitschka, M. Morisio, & M. Torchiano (Eds.), Product-Focused Software Process Improvement (pp. 36–48). Springer International Publishing. https://doi.org/10.1007/978-3-030-91452-3_3.
Machado, I.A., Costa, C., Santos, M.Y., 2022. Data Mesh: concepts and principles of a paradigm shift in data architectures. Procedia Comput. Sci. 196, 263–271. https://doi.org/10.1016/j.procs.2021.12.013.
Marr, B. (2021). Data Strategy: How to Profit from a World of Big Data, Analytics and Artificial Intelligence. Kogan Page Publishers.
Mikalef, P., Pappas, I.O., Krogstie, J., Giannakos, M., 2018. Big data analytics capabilities: a systematic literature review and research agenda. Info. Syst. E-Business Manage. 16 (3), 547–578. https://doi.org/10.1007/s10257-017-0362-y.
Mohagheghi, P., Aparicio, M.E., 2017. An industry experience report on managing product quality requirements in a large organization. Inf. Softw. Technol. 88, 96–109. https://doi.org/10.1016/j.infsof.2017.04.002.
Munappy, A.R., Mattos, D.I., Bosch, J., Olsson, H.H., & Dakkak, A. (2020). From Ad-Hoc data analytics to DataOps. Proceedings of the International Conference on Software and System Processes, 165–174. https://doi.org/10.1145/3379177.3388909.
Olsson, H.H., & Bosch, J. (2014). The HYPEX Model: from opinions to data-driven software development. In J. Bosch (Ed.), Continuous Software Engineering (pp. 155–164). Springer International Publishing. https://doi.org/10.1007/978-3-319-11283-1_13.
Pan, S.L., Tan, B., 2011. Demystifying case research: a structured–pragmatic–situational (SPS) approach to conducting case studies. Info. Org. 21 (3), 161–176. https://doi.org/10.1016/j.infoandorg.2011.07.001.
Parker, R., Bradley, L., 2000. Organisational culture in the public sector: evidence from six organisations. Int. J. Public Sector Manage. 13 (2), 125–141. https://doi.org/10.1108/09513550010338773.
Raghupathi, W., Raghupathi, V., 2021. Contemporary business analytics: an overview. Data 6 (8), 8. https://doi.org/10.3390/data6080086. Article.
Redman, T.C. (1992). Data Quality: Management and Technology. Bantam Books, Inc., United States.
Reutter, L., 2022. Constraining context: situating datafication in public administration. New Media Soc. 24 (4), 903–921. https://doi.org/10.1177/14614448221079029.
Runeson, P., Höst, M., 2009. Guidelines for conducting and reporting case study research in software engineering. Empir. Softw. Eng. 14 (2), 131–164. https://doi.org/10.1007/s10664-008-9102-8.
Schmidt, D.H., van Dierendonck, D., Weber, U., 2023. The data-driven leader: developing a big data analytics leadership competency framework. J. Manage. Develop. 42 (4), 297–326. https://doi.org/10.1108/JMD-12-2022-0306.
Storm, M., Borgman, H., 2020. Understanding challenges and success factors in creating a data-driven culture. In: Proceedings of the 53rd Hawaii International Conference on System Sciences. Maui, Hawaii, 7-10 January.
Thusoo, A., & Sarma, J.S. (n.d.). Creating a Data-Driven Enterprise with DataOps.
Trieu, V.-H., Cockcroft, S., & Perdana, A. (2019). Decision-Making Performance in Big Data Era: the Role of Actual Business Intelligence Systems Use and Affecting External Constraints.
Vestues, K., Hanssen, G.K., Mikalsen, M., Buan, T.A., & Conboy, K. (2022). Agile data management in NAV: a case study. In V. Stray, K.J. Stol, M. Paasivaara, & P. Kruchten (Eds.), Agile Processes in Software Engineering and Extreme Programming (pp. 220–235). Springer International Publishing. https://doi.org/10.1007/978-3-031-08169-9_14.
Yu, W., Wong, C.Y., Chavez, R., Jacobs, M.A., 2021. Integrating big data analytics into supply chain finance: the roles of information processing and data-driven culture. Int. J. Prod. Econ. 236, 108135 https://doi.org/10.1016/j.ijpe.2021.108135.

Astri Moksnes Barbala received her PhD in Sociology in 2022. Her research interests are linked to how digitization affects organizations and interpersonal relationships. Astri is currently researching topics such as responsible digital transformation and the use of AI in the public sector. She is a researcher at SINTEF and pursuing a Marie Skłodowska-Curie-funded post-doctoral project at Galway University.

Geir Kjetil Hanssen is leading the Digital Process Innovation research group at SINTEF. He is educated at NTNU with a doctorate from 2010. His main research interests are agile software-intensive organizations as well as development and application of software in safety-critical domains.

Tor Sporsem does research in software development at SINTEF and is a PhD candidate at NTNU. Tor is part of a research group in Sintef that has researched agile methods for over 20 years. He focuses on requirements engineering and user involvement.