



Constraining context: Situating datafication in public administration

new media & society
2022, Vol. 24(4) 903–921
© The Author(s) 2022



Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/14614448221079029
journals.sagepub.com/home/nms



Lisa Reutter 

Norwegian University of Science and Technology, Norway

Abstract

The imaginary of data-driven public administration promises a more effective and knowing public sector. At the same time, corporate practices of datafication are often hidden behind closed doors. Critical algorithm studies, therefore, struggle to access and explore these practices, to produce situated accounts of datafication and possible entry points to reconfigure the emerging data-driven society. This article offers a unique empirical account of the inner workings of data-driven public administration, asking the overall question of how sociotechnical imaginaries of datafication are constrained in the context of public administration. Teams working on datafication in two Norwegian public sector entities have been followed and interviewed over the course of 2 years (2018–2020). While sociotechnical imaginaries thrive in organizational culture and policy discourse alike, the observed data teams struggle to realize data assemblages due to a variety of structural and institutional constraints.

Keywords

Constraints, critical algorithm studies, data assemblage, datafication, public administration, sociotechnical imaginaries

Introduction

Across European countries, interest in the sociotechnical imaginary of data-driven public administration is growing (Misuraca and van Noordt, 2020). According to the Organisation for Economic Co-operation and Development (OECD), data-driven technology promises

Corresponding author:

Lisa Reutter, Department of Sociology and Political Science, Norwegian University of Science and Technology, Postboks 8900, Torgarden, Trondheim 7491, Norway.

Email: lisa.m.reutter@ntnu.no

to deliver more effective state benefits to families in need, save the lives of patients, protect children from abuse, and predict possible terrorist attacks (van Ooijen et al., 2019)—in other words, improving both service delivery and the efficiency of the public sector. While sociotechnical imaginaries represent collectively held and institutionally stabilized future visions of society, they nevertheless can be multiplied and contested (Mager and Katzenbach, 2021). Furthermore, future visions of datafication are mediated and negotiated by the institutions, workers, and users set on the quest to materialize the grand ideas of datafication (Caplan et al., 2020). This article asks the overall question: how does the context of public administration constrain the sociotechnical imaginaries of datafication as these are realized in specific data assemblages? The article offers unique empirical insight into the inner workings of public administration datafication in Norway and contributes to critical data and algorithm studies by situating datafication beyond Internet platforms, providing possible entry points for the reconfiguration of the data-driven society.

Data-driven technology is far from benign. Social scientists have pointed out a variety of shortcomings tied to the ever-growing influence of datafication in society, such as its impenetrable opaqueness, reinforcement of discrimination, and facilitation of surveillance (boyd and Crawford, 2012; Pasquale, 2015; van Dijck, 2014; Zuboff, 2019). A growing body of case studies has begun to address the unintended consequences of data-driven systems in public administration (Allhutter et al., 2020; Eubanks, 2018). Ultimately, this critique addresses a shift in the power dynamics between governments and citizens, as the state's potential to understand, predict, and control the activities of citizens is vastly enhanced (Dencik, Hintz, et al., 2019). Data activist studies have, therefore, started to research bottom-up initiatives in response to unjust data practices (Beraldo and Milan, 2019). This article builds on this research and directs analytical attention toward the ways dominant imaginaries of datafication are mediated and responded to within various contexts in which they are employed, prior to their working in the world (Kazansky and Milan, 2021). By doing so, the article aims to produce an actionable critique rather than a general and oversimplified account of datafication. This critique takes into consideration the variety of conditions of production and implementation and how these might mediate the power and influence of data-driven technology. Recognizing the already existing mediations helps to extend the understanding of how to respond to and govern datafication, to mitigate data harms, and to facilitate discussion about the ways in which datafication is contingent and therefore might be otherwise.

Social science researchers must find a balance between problematizing general trends in society and particular applications (Dencik, Redden, et al., 2019). Internet platform datafication has received significant scholarly interest (Bucher, 2018; Mejias and Coudry, 2019). Corporate practices of datafication, however, are often hidden behind closed doors. Researchers, therefore, struggle to access and explore datafication practices and to establish a balance between myth and hype and actual applications of data-driven systems (Elish and boyd, 2018; Ziewitz, 2015). Research carries the risk of disregarding how imaginaries are negotiated by the institutions, workers, and users as they strive to realize these future visions, and researchers may overdetermine its power or fetishize data and algorithms in this regard (Caplan et al., 2020; Thomas et al., 2018). Earlier research on administrative reforms related to data-driven public administration has already shown that the public sector encounters a variety of

obstacles when negotiating such visions of the future (Andrews, 2019; Mergel et al., 2016; Redden, 2018). Observed obstacles are often tied to the very nature of public administration and its constituent institution. Public administration consists of thousands of human minds; the work of its organizational units is heavily governed by legal documents, organizational structures, and existing data infrastructures. Its subjects are not consumers, but citizens; its goals are decided by politicians rather than corporate managers. Its inner workings are not “black boxed” because of trade secrets, but instead open for scrutiny by law.

The case studies presented in this article were conducted in two Norwegian public sector entities and use machine learning projects as an entry point into newly established data assemblages. The Norwegian Tax Administration and the Norwegian Labor and Welfare Administration are the two largest public sector entities in Norway, owning and managing vast amounts of data on its citizens and entrusted with broad public mandates. Data teams have been followed and interviewed over the course of 2 years (2018–2020). Before presenting the analysis of public administration mediations, the article introduces the concepts of datafication, sociotechnical imaginaries, and data assemblages, along with its research design.

Approaching datafication: from sociotechnical imaginaries to data assemblages

Datafication is the practice of quantifying every aspect of life so that it can be analyzed (Mayer-Schönberger and Cukier, 2013). The datafication of public administration consists of two interwoven processes: the use of *more and different data* (big government data) and the deployment of *more advanced methods* to analyze these data and feed it back into existing work processes. These interwoven practices materialize in concrete projects of automated decision-making, sorting algorithms, and the development of decision support tools enabled by machine learning. Datafication represents a strategy of administrative reform primarily concerned with the reform of knowledge production practices and the introduction of algorithmic forms of ordering (Yeung, 2020). Data-driven public administration operates according to the logics of categorization, classification, scoring, and selection and presents an extension and profound change in public sector data practices (Dencik et al., 2019). The datafication of public administration is built on the ideological foundation of *dataism*, the belief that data can promote a better, more effective, and more ‘objective’ societal apparatus (van Dijk, 2014). It is this belief in data’s potential that is the core of the sociotechnical imaginary of data-driven public administration.

The concept of sociotechnical imaginaries addresses the normative function of “collectively held, institutionally stabilized, and publicly performed visions of desirable futures, animated by shared understandings of forms of social life and social order attainable through, and supportive of, advances in science and technology” (Jasanoff and Kim, 2015: 4). This analytical tool addresses the co-production of technology, society, and politics and provides researchers with a framework within which to investigate how practitioners and institutions make sense of technological change. Sociotechnical imaginaries provide visions of the kinds of society that sociotechnical

change could bring into being; they shape what is thinkable and the practices through which actors perform their roles (Ruppert, 2018). Connected to social and technological order, they are, therefore, often associated with state power and the selection of priorities, allocation of funds, and investment in material infrastructures (Jasanoff and Kim, 2009). Datafication lies at the heart of the future imaginary of the welfare state and guides and informs public administration reforms (Dencik and Kaun, 2020). The data-driven imaginary is built on the idea of fast and cheap data analysis, the possibility of gleaning data from entire populations, data speaking for itself, and agnostic algorithms that guarantee an impartial view from nowhere (Rieder and Simon, 2016). Such sociotechnical imaginaries include notions of proactive, rather than reactive, modes of governance. Datafication provides the public sector with a sense of being able to do more, better, faster, and more cheaply and is therefore perceived as a solution to the growing complexity of society and administration and as a tool to reduce uncertainty (Maciejewski, 2017; Strauß, 2015). The vast amounts of data in Nordic registries were into this overall sociotechnical imaginary of the public sector as a data gold mine (Tupasela et al., 2020).

The content and performativity of sociotechnical imaginaries of datafication has been discussed and assessed in a variety of ways in recent years (see, for example, Lehtiniemi and Ruckenstein, 2019; Tupasela et al., 2020; Williamson, 2018). Although studying sociotechnical imaginaries themselves presents interesting insights into the wider origin and legitimation of data-driven technology in public administration, it provides little insight into the ways they reconfigure public administration and are themselves altered in concrete settings. Jasanoff and Kim (2015) therefore emphasize the need to investigate how sociotechnical imaginaries enter, are translated, and negotiated into assemblages of materiality, which this article intends to do.

Kitchin and Lauriault's (2014) framework of the data assemblage holds the potential to advance this understanding further. This framework is intended to illuminate the contextual and situational forces that influence the way data systems are shaped, to gain a better appreciation of the work they do in the world (Kitchin, 2014: 24). Data assemblages consist of a variety of apparatuses (entwined and contingent parts of the larger system) and elements (see Table 1).

These elements work together to produce a data assemblage not only discursively but also materially. Deconstructing the assemblage of data-driven technology allows researchers to show that these technologies are value-laden, socially constructed entities influenced by a variety of technological, political, social, and economic apparatuses (Kitchin, 2017). In their analysis of datafied child welfare services, Redden et al. (2020) apply this framework and show how systems of thought, ownership structures, policy agendas, organizational practices, and legal frameworks influence the datafication of public administration. They use these influencing factors to point out the sociotechnical aspects of datafication and the value-ladenness of systems. The following analysis is an attempt to reframe this argument: rather than regarding the apparatuses presented by Kitchin and Lauriault (2014) only as value-laden influencing factors, this analysis attempts to regard them as possible mediators that constrain sociotechnical imaginaries and allow for reconfigurations as these become specific data assemblages. This builds on the idea advocated by Tanweer et al. (2016) and Redden and Grant (2020) to study points

Table 1. The apparatus and elements of a data assemblage (Kitchin, 2014).

Apparatus	Elements
Systems of thought	Modes of thinking, philosophies, theories, models, ideologies, rationalities, etc..
Forms of knowledge	Research texts, manuals, magazines, websites, experiences, word of mouth, chat forums, etc..
Finance	Business models, investment, venture capital, grants, philanthropy, profit, etc..
Political economy	Policy, tax regimes, public and political opinion, ethical considerations, etc..
Governmentalities and legalities	Data standards, file formats, system requirements, protocols, regulations, laws, licensing, intellectual property regimes, etc..
Materialities and infrastructures	Paper/pens, computers, digital devices, sensors, scanners, databases, networks, servers, etc..
Practices	Techniques, ways of doing, learned behavior, scientific conventions, etc..
Organizations and institutions	Archives, corporations, consultants, retailers, government agencies, universities, conferences, clubs and societies, committees and boards, communities of practice, etc..
Subjectivities and communities	Of data producers, curators, managers, analysts, scientists, politicians, users, citizens, etc..
Places	Laboratories, offices, field sites, data centers, etc., and their agglomerations
Marketplace	For data, its derivatives, analysis, analytical software, interpretations, etc..

of breakdown, where progress is stopped due to material and discursive limitations; these provide valuable insights and allow new imaginations and configurations of data assemblages to be developed. Bridging the concepts of sociotechnical imaginaries and data assemblages, this article shows how collectively held future visions are translated and negotiated into assemblages of materiality. This also creates the potential to find an analytical balance between problematizing general trends in society and emphasizing particular applications (Dencik et al., 2019). The sociotechnical imaginary of the data-driven society can be observed across both private and public institutions, but the apparatuses of data assemblages are highly context-dependent.

Researching public administration datafication

Machine learning, where algorithms are fed with big government data, is only one of the technologies associated with the datafication of public administration. This article uses this to illustrate the realization of sociotechnical imaginaries of data-driven public administration into specific data assemblages. Machine learning projects present an entry point into the data assemblage of data-driven technology; this article offers a concrete example of how such administrative reforms are (or are not) implemented.

Kitchin (2017) and Dencik (2019) encourage research into specific datafication projects and their contexts. This article therefore employs an empirical research design consisting of fieldwork, interviews, meeting observations, and document study. By law, the Norwegian public sector is required to provide access for researchers, which provides unique insight into its inner workings.¹ Two data teams in two different public entities have been followed over 2 years, the Norwegian Labour and Welfare Administration (NAV) and the Norwegian Tax administration (SKATT). Neither entity relies on private sector competence, which further eases access to the field itself. Contact with the NAV team was established in 2017 and with SKATT at an AI fagforum meeting in 2018. As the progression of these projects is slow, the NAV team was followed from January 2018 to May 2019 and the SKATT team from June 2019 to November 2019. This included five visits to each of these public entities, with each visit lasting 1 week or less. In total, the assembled data include 37 days of field notes, 17 interviews with key organizational actors and interviews with the data team, 16 meeting observations, 12 PowerPoint presentations, and approximately 60 pages of strategic documents or internal documentation. Several check-in meetings were conducted after the initial field work to follow-up on the projects. To understand the wider foundations of the work more thoroughly, I also attended and observed several industry and public sector conferences in 2018 and 2019. At conferences, notes were taken both at presentations and reflecting informal discussions among data team members. Key policy documents were analyzed: these included the digitalization strategy for the public sector, the Norwegian AI strategy, the concept selection study on sharing of data within the public sector, and mandate letters for the two organizations (2017–2019). An abductive analysis of the data material was conducted (as described by Tavory and Timmermans, 2014), based on in-depth reading of literature, immersion in fieldwork, and subsequent coding in NVivo. The element of surprise—the fact that most observed datafication efforts seem to meet a variety of obstacles along the way despite strong sociotechnical imaginaries—guided this research project significantly.

Research site: the NAV data team

NAV is the largest Norwegian public agency and is at the forefront of an ongoing nationwide digital transformation. It is a public welfare agency that delivers over 60 benefits and services, such as unemployment benefits and pensions. Operating under the Ministry of Labour, NAV has approximately 19,000 employees and manages approximately one-third of the overall state budget. NAV established a “big data”-laboratory in 2016, collecting a variety of actors from across the organization and exploring possibilities for big government data recirculation. The NAV data team resulted from this initiative, as part of the newly established division for Data and Insight. This section is mandated to collect all environments that develop and manage data products in the organization and is placed within the IT department. The NAV data team has now approximately 10 data workers from a variety of backgrounds, including physics, political science, and data engineering. At the time of writing, NAV has no machine learning model in production (November 2020).

One of the major projects followed in the fieldwork was the “sickness benefit” project. This project was established in 2018 and aims to have a first pilot in production by

December 2020. The project intends to predict the sickness absence of employees and give an indication for the necessity of an assessment meeting between a NAV case worker, the employee on sick leave, and the employer. The project has two full-time data team members and trains its models with sick leave certificates, using a boosting tree model. At this time, the final product is intended to be a decision support tool in the expert system for the case workers.

Research site: the SKATT data team

SKATT operates under the Ministry of Finance. Its main objective is to secure the funding of the welfare state through tax collection. It employs approximately 6500 people and has a variety of control and tax functions. The SKATT data group has a well-established team of data analysts, having existed since the early 2000s. Its main tasks include both traditional data analysis and report production, and project-based recirculation of data. This team has 27 employees based in two locations, and most of these employees still work on traditional analysis. SKATT has seven models in production that can be classified as machine learning.

One of the main projects followed in the field is the value-added tax project (PSU project). This project was established around 2013, aiming to systemize and automate the selection of control cases in value-added tax reporting by companies. Initially a rule-based engine (using if-then selection criteria) was employed in this process, but the organization's control unit wished to further optimize this process. Machine learning was not within the initial mandate of the data team's work but was explored along the way. The developed PSU model, a supervised machine learning model, is now one out of the seven working machine learning models in production within the agency. Based on historical data, it risk-scores and sorts all incoming reports, which are then controlled by street-level bureaucrats. Controlled and resolved cases are then fed back into the machine learning model.

Constraining context: toward a data-driven Norwegian public administration

Although the Norwegian public sector is one of the most digitized in Europe and socio-technical imaginaries are enacted widely, a recent general mapping shows that there are few data-driven technologies in production (Broomfield and Reutter, 2019). Both data teams observed in this study have struggled to develop and implement new data assemblages within their organizations. It therefore appears that the context of Norwegian public administration is one where sociotechnical imaginaries flourish within the organizational discourse but where new data assemblages struggle to stabilize, despite the vast amounts of registry and case-work data on citizens. The following analysis reflects an attempt to deconstruct data-driven technology production and implementation within Norwegian public administration. It follows the sociotechnical imaginary from collectively held visions of the future as they do or do not stabilize into concrete data assemblages. This analysis begins by presenting the policy and the strategy guiding the work of data teams; then it moves on to the organizational and institutional constraints that

provide the framework within which the data teams work, the legal and regulatory environment for data utilization, the data infrastructure, and the practical aspects of implementing machine learning. The apparatuses presented in the analysis do not offer an exhaustive list of obstacles that mediate data-driven technologies in the process of their production. They serve, however, as a starting point to engage with these administrative reforms, to better understand how future visions are negotiated and translated into public administration practices.

Policy and strategy

Sociotechnical imaginaries are often associated with the ways nation states and governmental actors envision technological development and unfold their power to imagine, govern, and program technologies (Jasanoff and Kim, 2009). This is strongly evident in the field of public administration. Systems of thought and sociotechnical imaginaries about data-driven public administration are interwoven, manifested, circulated, and enacted in both national and multinational policy and strategy documents, as, for example, shown in the introduction of the national AI strategy:

We know that Norway will be affected by an aging population, climate change and increased globalization, and that we must work smarter and more efficiently to maintain competitiveness and the level of welfare in the years to come. Digitalisation and new technology are the key to achieve this—and artificial intelligence will be central. (Ministry of Local Government and Modernisation, 2020)

The sociotechnical imaginary of public sector datafication in Norway promises to solve the problem of ineffective public administration. As do other countries, Norway supposedly faces a variety of future challenges, any of which might decrease the level of national welfare. These challenges include the downsizing of the oil sector, growing immigration, climate change, and globalization. Norwegian value creation in both public and private entities is deemed to be dependent on a thriving data-driven economy and government, according to policy documents. Datafication is thus regarded a common sense solution and part of the evolution of public administration. Public administration is concerned with the implementation of government policy and the administration and execution of public law. Policy documents, therefore, reduce sociotechnical imaginaries to actionable plans and focus areas.

Policy documents stress the tremendous unused potential in Norway's data sources. Much emphasis is placed on the sharing and re-use of data in public administration and between public and private organizations. The sheer availability of data is expected to lead to new solutions and improved services. This vision of the future is increasingly influenced by the private sector propagating notions of technological progress and benefit, as evident in a public spheres elsewhere (Hockenfull and Cohn, 2021). This could be observed in relation to the stakeholders consulted in policy production and the conferences visited in the field work. The private sector regards data as a key value-creation opportunity. According to national policy, users of public services also experience the services provided by the public sector as bothersome. Data sharing is expected to enable new data practices and more effective and seamless public service.

Policy documents are an important apparatus within the data assemblage, as they frame what and how data are used and recirculated (Kitchin, 2014). They provide public sector leadership with legitimation and a planning framework for the implementation of administrative reforms. Mandate letters, given to the public entities observed, shape the actual day-to-day work in the public entities. These are produced every year by the governing ministry and determine what the public entity is to work on through a 1-year period. They provide fiscal frameworks, priorities, expected results, and reporting requirements. The areas of special interest within these documents often define where the data teams will work and what kind of projects are initiated. Sickness benefit follow-up, for example, has been key to several mandate letters.

Organizational scope

Organizations and institutions are key apparatuses of the data assemblage (Kitchin, 2014). In addition to mandate letters, each of the organizations has detailed strategies which lay out long-term plans for the organization. These offer descriptions of how the organization is expected to operate in the future, echoing the sociotechnical imaginaries outlined at policy level and emphasizing the tremendous potential for effective change and the fear of missing out on a technological opportunity.

Technological developments can make the agency more efficient and free up resources for other tasks. [. . .] It is likely that several of the agency's tasks over time, in whole or in part, will be replaced by smart ICT solutions. The Tax Administration must avoid ending up in a situation where the agency's systems and solutions become outdated and incompatible with the rapidly changing outside world. (Skatteetaten 2025, 2017)

The institutional framework of public administration, therefore, is an important factor shaping the data assemblage. Data teams act as executive agents of sociotechnical imaginaries within their organizations and play a key role in maintaining and circulating these imaginaries. Within a public sector organization, however, the data team's influence is mediated by its assigned scope. Both of the data teams observed are relatively small compared to the size of their organizations. The teams are part of a larger effort to recirculate data with stakeholders across the whole organization. Meanwhile, the two teams are placed in different sections in their organizations. The SKATT data team is relatively well-established within the analysis department, having its origins in the data warehouse section established in the early 2000s. The NAV data team, however, was created recently and is part of the IT department. The placement of a data team and its size, in addition to its organizational constraints, limits what it can do. Tight connections to the IT department ease the integration of systems into already existing digital infrastructure, as observed in NAV. At the same time, extensive experience curating, accessing, and analyzing data provide the SKATT data team with competence in the field of data analysis and existing data infrastructures.

Most datafication efforts and projects are led by the central core of public administration and are intended to be spread to lower levels. The public entities examined in this study differ in size, organization, and administrative scope. The NAV data team has a

wider administrative scope and a more fragmented and distributed organizational structure than the SKATT data team. This has complicated data access, due to differences in legal mandates between the municipal entities and the central level at NAV, according to member of the data team.

The financial apparatus of Norwegian public administration places great emphasis on project-based practices. Projects are mostly owned and ordered by other entities within the organization. Data teams are expected to deliver projects to these entities, which manage and decide how much to spend. As projects are time-limited, and data-driven efforts are often costly and regarded as long-term strategies by these organizations, several projects were “*put on ice*” during the period of observation because they either lacked resources or were not prioritized by the project owner. The data teams do not initiate projects by themselves, but instead are expected to act on “needs” identified by others within the organization. Surprisingly, both teams experienced little interest in data-driven technologies in their wider organizational environment. The SKATT team has struggled to identify possible future projects and, on multiple occasions, had to initiate an internal effort to find projects. The NAV data team spent its first months in existence presenting what machine learning is, and how it can be used, to a variety of actors within the organization. Both data teams highlighted that the so-called “need” for data-driven technology often needs to be produced by the data team together with the domain knowledge experts. Presentations about what data-driven technology is, and what it can help to accomplish, seemed to be a key activity for the circulation and enactment of sociotechnical imaginaries. The data teams use significant resources to distribute and circulate these visions of desirable futures for data-driven public administration, building authenticity through pilot projects or success stories. This has been especially difficult for the NAV data team, as they do not yet have a success story of their own. As Hockenfull and Cohn (2021: 16) point out in their study on corporate sociotechnical imaginaries of datafication, these imaginaries work in complex ways, being both vacuous and productive. These authors use the notion of “hot air” to point out the way sociotechnical imaginaries can reinforce common sense notions of progress, using buzz words to create a rhetorical ethos and building authority through exemplars.

Overall frameworks for projects are decided on by the domain experts as project owners. The framework of the sickness benefit project and its goals were assigned to the data team members, with the specific goal being to predict the necessity of assessment meetings. What the data team members actually attempted to predict is whether a worker would be on sick leave for longer than 16 weeks, which according to subject matter experts indicates a need for follow-up. What can be done, and how, is therefore often quite restricted and rule-based and is not steered by data as such. As one of the data team members points out: “*The 16 weeks are maybe not an ideal cut. But we make the best out of it*” (NAV data team member).

The project-based organization of work also leads to a high degree of dependence on other actors and sections within the organization. The data teams’ progress, therefore, also depends on their ability to cooperate across the organization and to include key actors in their work: this includes subject matter experts, data managers, and legal staff. A key apparatus of the data assemblage is the form of knowledge itself. The public mandate of NAV is often connected to relatively vulnerable situations in people’s lives, such

as unemployment, pensions, and illness; the organization therefore regards social work as a key form of knowledge. SKATT, on the contrary, regards accounting as its key form of knowledge, which is often connected to the practice of control.

Legal mandates

Legal frameworks are regarded as inhibiting the progress of datafication by data team members and have therefore received significant attention in recent policy documents. The translation and negotiation of sociotechnical imaginaries into material assemblages depends on an apparatus of governmentalities and legalities (Kitchin, 2014). Public administration represents the executive sphere of politics in modern democracies, and its realm of action is tied to some legal frameworks and mandates. In the Norwegian context, the public administration act, the context specific NAV law, and the tax administration act are each central to how specific public services are regulated, encompassing everything from what services are provided to how they should be performed. The analysis arising from this study clearly shows that mediation through law heavily influences new data assemblages in both data teams; this apparatus also manages to stop or delay most projects in these organizations.

“Legal assessments are a real show-stopper” (Project lead, SKATT). These data teams experience two main areas as especially challenging: (1) legal mandates that prohibit the recirculation of data collected in one area of the public entity into other areas of practice, and (2) the lack of legal frameworks to regulate the linking of different data sources. This concerns both existing legal frameworks and the absence of frameworks. Where legal frameworks do not yet exist juridical decisions can be restrictive, stopping projects before they start or along the way. Legal staff in both organizations struggle with the concept of government data recirculation. For example, the SKATT data team repeatedly noted that the assessment of projects can be frustrating and unpredictable and, in some cases, is experienced as essentially random. Existing legal frameworks significantly constrain projects in both data teams, for example, by excluding specific variables or datasets from use or analysis.

All data projects must be tied to a legal mandate, meaning that data teams and subject matter experts spend a significant amount of time identifying sections in the public administration act that can be connected to their projects, to legitimate the use of data-driven technology. In the case of the NAV sickness benefit project, the *folketrygdloven*, Chapter 8, § 8–7 states that *NAV must hold a meeting together with the sick employee, the employer, before the illness benefit reaches 26 weeks. This is unless a meeting is considered to be clearly unnecessary*. Here, the sentence “considered to be clearly unnecessary” has been used to legitimize the whole project. The machine learning model is supposed to show what is “clearly unnecessary” by predicting cases that do not require follow-up. This decision was previously taken by the case worker responsible for sickness benefit follow-up, independent of any data analysis. It is now intended to be taken by case workers informed by the decision support tool the team has developed. As both the SKATT and NAV data teams point out, the legal mandates tied to control issues are often wider. It is therefore easier to get projects tied to control approved.

Most machine learning work at SKATT begins with simple logistic regressions, as nearly all projects are built on already existing rule-based automated sorting. The Norwegian Public Administration Act clearly states that decisions must be explainable and transparent (Public Administration Act, 2021). The data teams therefore perceive that the range of methods allowed is small. Most of the models developed and discussed at the point of observation were decision tree models or advanced logistic regressions. Supervised machine learning is often chosen rather than complex deep learning or neural networks: the latter are not explainable and therefore cannot be used, even if they score higher for accuracy in the test data. The NAV data team, for example, has developed a neural network to analyze the text in the certificates but decided to stop as this would require further legal assessments. It is often not the coding work itself that is negotiated, but its outcomes and foundations. Both teams spent a significant amount of time assessing the question of fairness in their machine learning models, asking questions, such as what is fairness in machine learning models in general? What kind of fairness measures should be used in which situation, and who can decide on this? The NAV team in particular has worked both methodologically and practically on this issue.

While the SKATT data team has obtained approval for several projects, the NAV data team has often been blocked in their work. Again, this may be connected to the nature of the work in NAV, which is often connected to vulnerable situations in the lives of citizens and is therefore regarded as more complex than taxation. The NAV data team has, therefore, initiated a project to standardize the overall process of impact assessment and juridical work. This data impact assessment contains a detailed description of the work processes and different phases of projects, in addition to risk descriptions tied to different variables. It also attempts to assess the potential societal consequences of the system as a whole.

The most interesting aspect is the perception by both data teams of gaps in the legal frameworks governing their work. This was also connected to the fact that, although there is a significant body of regulations dealing with privacy issues, machine learning regulation has often been left to vaguer ethical guidelines. These guidelines, however, are not the core domain of legal staff in public administration in these organizations. Making decisions concerning what are good enough reasons to use data, where data can be combined, and what decisions can be influenced by machine learning is left to individual organizations and data team members. They, however, are not comfortable making these decisions, while questions of fairness or transparency are not resolved by clear guidelines.

Data and infrastructure

Data both limit and enable the work of the data teams observed in this study. Policymakers and public sector leaders see the availability of data as a key enabler and force within the sociotechnical imaginary of public administration. According to these data teams, how data can be used depend on access, quality, and the context within which the data are collected. Both teams have expressed their requirement for resources to make data machine learnable. Data thus are mutually constituted and bound together in a variety of practices (Kitchin, 2014).

All policy and strategy documents examined for this study highlight the tremendous potential within the public sector's gold mine of data. However, existing materialities and infrastructures have not been constructed with this in mind. The data infrastructure of both organizations is, therefore, currently being upgraded. Although their efforts differ in size and original impetus, both public entities envision a new data platform to ease data analysis and the re-circulation and use of existing data. A visionary in the NAV IT department used the metaphor of a self-serve buffet to describe their project, where data could easily be picked and combined by a variety of actors. These infrastructures are envisioned to replace traditional data warehouses in both organizations. Both envision data platforms as strategic goals for the future. The NAV data team was required to build their own data infrastructure to start their work, to realize a sociotechnical imaginary on their own. As projects often do not start with data, but are required to be connected to a need or problem and a legal mandate in the organization, data enter the scene after a variety of apparatuses are already arranged. What data are available, and in what quality, is often unclear to the data teams. How to access previously unused data or acquire sufficient information on datasets is regarded by both data teams as difficult. Data are often not organized and curated for data analysis, in either organization. According to data team members, the data available in warehouses are often in aggregate form inadequate for machine learning purposes.

Mining public sector data is seldom a straightforward process. How much data can be put together depend on a variety of practical constraints. The sickness benefit project has spent nearly 2 years collecting and analyzing one data source: sick leave certificates. These have been produced by doctors around the country and submitted to NAV in various formats. In addition, hospitals have still not digitized all sick leave certificates. Although this project intended to use other data sources, it had to halt such efforts: "*This project is pragmatically driven. We do not have time to dig into more data right now. There are so many owners and sources*" (NAV data team member). Aligning data sources and cleaning the data so they can be analyzed requires significant work. The data team also pointed out that they have received a lot of feedback from case workers on the data sources they currently use in their assessment of cases. These sources, however, cannot be accessed by the team, as they lie outside of their systems.

The SKATT and NAV data teams spend a significant amount of time "getting to know" the data before starting their analysis. This activity is connected to understanding how data have been collected and why, and includes an active dialogue between subject matter experts and the data team. Discovering the limitations of the data is regarded as highly important by the data team members. This is clearly in direct contrast to the expectations of fast and cheap data analysis, and whole-population data, which feed the sociotechnical imaginary of datafied public services (Rieder and Simon, 2016).

Doing machine learning?

The data teams spend little time actual coding machine learning algorithms or training and testing models, as most of their time is spent determining organizational requirements, holding project meetings, acquiring data, managing data quality, performing legal

assessments, and discussing which model to use in both teams. “*Building the actual machine learning models, is what takes the least time here (laughs)*” (SKATT data team member). It is only at this point that the sociotechnical imaginary is finally realized into a specific working data assemblage, able to do work in the world (Kitchin and Lauriault, 2014). To date, few machine learning models have been deployed in these organizations. Most new data assemblages remain within the data teams and have not yet managed to do work in the world. At the time of observation, SKATT had been able to deploy seven models, while NAV still has none.

The actual training and testing of models has not been observed in this project. Nevertheless, the outcome and content of choosing, training, and testing was discussed with data team members in both public entities. Training and testing machine learning models is often experienced as highly interesting for team members: “*They love tinkering with the models*” (SKATT, team lead). The required accuracy level of machine learning models is decided on by the data team together with subject matter experts and legal staff. The sickness benefit project, for example, has decided that an 82% accuracy rate will be sufficient for the decision support tool. Again, a variety of subjectivities and communities is involved in these negotiations.

The day-to-day reality that a lot of time is not spent on any form of direct data work could be experienced as bothersome by practitioners. The NAV data team in particular expressed severe frustration at times. The content of the practice of machine learning is itself contested; several members of the data teams argued that all of the other work also is part of their responsibility and cannot be disregarded. This quote shows exemplifies the variety of tasks of a data team member:

There were many processes to deal with and manual transfers of data and information. [. . .] The entire project consisted of many actors and it was our task to make this work. Client / management, tax assessment team, coordination team (management for the inspectors), project managers / analysts and controllers. In addition, one had to have an overview of the data flow in data warehouses and ask them to change their work routines. In other words, there was a lot of practical planning work. In addition, we were on tour to inform about the model and reassure the employees of what is going to happen. (SKATT, data team member)

Circulating and maintaining the sociotechnical imaginaries of data-driven public administration is regarded as equally important to actually building the data assemblage, at this point of time because the data teams depend on the acceptance of this future vision within their organizations. An analysis of presentations also shows how the data teams translate the dominant sociotechnical imaginaries developed by corporate and state actors into the context of public administration.

Interestingly, most data team members point out that machine learning or other predictive technologies may not be an appropriate solution to the problems they are assigned. The practice of machine learning, therefore, also includes the practice of understanding when machine learning is not the solution. In other words, data team members engage in negotiations concerning the extent and the boundaries of these sociotechnical imaginaries in their day-to-day work.

Discussion and conclusion: situating datafication

The datafication of public administration has been encouraged by both multinational and national policy and strategy documents for several years; the sociotechnical imaginaries related to this administrative reform have been circulated and been enacted in a variety of arenas. However, they largely represent visions of the future rather than any accomplished result. The context of Norwegian public administration prohibits a variety of datafication efforts. This situation is experienced as bothersome and frustrating by practitioners, who have set out to realize the promise of data-driven public administration. This observed friction, however, offers valuable insights into the variety of apparatuses that frame the nature, operation, and work of this administrative reform (Kitchin and Lauriault, 2014). Such insights contribute to our understanding of how data-driven technologies can be governed, controlled, and held to account (Dencik, Hintz, et al., 2019). A growing body of literature testifies to the societal consequences of datafication and shows how data-driven technologies are socially constructed rather than neutral (Beer, 2017; Kitchin, 2017; Redden et al., 2020). This article builds on the developing literature, reframing the apparatuses of the data assemblage by investigating how existing public administration apparatuses manage to constrain powerful future visions when actors attempt to translate and negotiate them into assemblages of materiality. This analysis demonstrates the rich observations made possible through situated, empirical studies on datafication in public administration in specific contexts, rather than attempting to make more abstract, general observations.

Sociotechnical imaginaries are contingent, multiple, and contested (Mager and Katzenbach, 2021). They often do not materialize in full, and not without interruption (Caplan et al., 2020). The data projects followed in this study are all intended to alter public administration but at the same time are themselves constrained by a variety of mediations. Structural and institutional constraints provide obstacles to the fulfillment of future visions of a data-driven society. While sociotechnical imaginaries clearly thrive in corporate discourse and policy, and within organizational discourse, the new data assemblages constructed by the data teams observed in this study remain highly unstable and demand immense resources (Hagendorff and Wezel, 2019). Practitioners are forced to constantly negotiate between existing apparatuses and broader future visions. The promise of a more effective and informed public administration is restricted by the technical, legal, organizational, methodological, and political constraints observed in these cases. Some of these constraints seem mundane at first but contribute to the mediation of sociotechnical imaginaries into small, often unstable, data assemblages. Datafication is always embedded into specific contexts, with structures, values, and conditions that are reflected in practice and thus already manage to mediate the power of data and algorithms.

The concept of the sociotechnical imaginary introduced in this research project offers an analytical tool to understand the nature and power of collectively held future visions, while the concept of the data assemblage directs our analytical attention toward the relational and contextual discursive and material practices shaping data-driven technologies. Currently, the constraints presented in this case study have prohibited the data teams from utilizing the vastly enhanced possibilities afforded by emerging technologies to understand, predict, and control the activities of citizens. At the same time, they also reduce the richness of

sociotechnical imaginaries, as can be observed in the narrow focus on control in projects. SKATT has a clear advantage when producing data-driven technology, as one of its main tasks is controlling, and public mandate is more clearly defined. Most constraints face contrary pressure from policymakers and other actors who work to maintain the sociotechnical imaginary and therefore aim to get rid of these obstacles. This has led, for example, to the idea of digitalization-friendly law, which is intended to enhance datafication (Plesner and Justesen, 2021). There is little reflection in the field on why these obstacles have been established in the first place. A deeper understanding of the interaction, interrelation, and negotiation between mediations is crucial to further develop the field of critical data and algorithm studies. To avoid oversimplified accounts of data-driven technology which are of little influence in this field of practice, research must take the variety of conditions of datafication into account (Moats and Seaver, 2019; O'Neil, 2016).

Data assemblages are not neutral and are influenced by a variety of apparatuses (Kitchin and Lauriault, 2014). Datafication may often be problematic, producing unintended consequences in its entanglement with power and politics (Beer, 2017; Bucher, 2018; Dencik and Kaun, 2020; Dencik, Hintz, et al., 2019). To understand datafication as more than a general, albeit substantial, shift in the organization of society requires us to situate it beyond Internet platforms. Data-driven technologies produced at the Google headquarters in Silicon Valley may differ substantially from monitoring systems in the Norwegian Tax administration. The critique of private sector datafication often seems to place society in passive or powerless positions, leaving little space for intervention and reification except after systems have already caused harm (Caplan et al., 2020). This study's approach avoids data-driven determinism, showing how policy, organizational structures, legal frameworks, subject matter experts, and existing data infrastructures are able to mediate datafication in significant ways. These constraints act as counterforces against dominant sociotechnical imaginaries strongly dominated by the private sector (Bareis and Katzenbach, 2021). The constraints laid out above provide us with an interesting point of departure to engage reflexively and critically in datafication practices. In many cases, data-driven public administration still remains a future vision rather than being realized in stable data assemblages; this might provide us, as citizens, users, or researchers, with opportunities to alter these imaginaries prior to or even during their translation into new and effective data assemblages. An actionable critique of the phenomenon might ask which limitations to interrogate and how these already existing constraints can and should be altered, and by whom. Further investigation will also require us to reflect more deeply on who is able to mediate them and who is not. The obstacles presented here can, therefore, act as points of entry to reconfigure the sociotechnical imaginary of data-driven public administration.

Author's note

The article is not currently being considered for publication by any other print or electronic journal.


Acknowledgements

The author thanks Heidrun Åm, Heather Broomfield, and Emil Røyrvik for reading and commenting on earlier drafts of the article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The research presented in this paper was funded by the Norwegian University of Science and Technology's (NTNU) Digital Transformation initiative, project: 'Digital infrastructures and citizen empowerment (DICE).'

ORCID iD

Lisa Reutter  <https://orcid.org/0000-0002-4952-0416>

Note

1. The research project had to constantly renegotiate access to the public entities and was heavily inspired by Nick Seaver's (2017) work on the ethnography of algorithmic systems. Here, Seaver argues to treat "access as a kind of texture, a resistance to knowledge that is omnipresent and not always the same." The topics of ethics and unintended social consequences have become quite prominent in the public sector: this was used to legitimate access to work practices in this study. While this article is theoretically driven and intended to contribute to the field of critical data and algorithm studies, other parts of the research project included more practically oriented research outputs and popular science presentations to practitioners.

References

- Allhutter D, Cech F, Fischer F, et al. (2020) Algorithmic profiling of job seekers in Austria: how austerity politics are made effective. *Frontiers in Big Data* 3: 5.
- Andrews L (2019) Public administration, public leadership and the construction of public value in the age of the algorithm and "big data." *Public Administration* 97(2): 296–310.
- Bareis J and Katzenbach C (2021) Talking AI into being: the narratives and imaginaries of national AI strategies and their performative politics. *Science, Technology & Human Values*: 1–27.
- Beer D (2017) The social power of algorithms. *Information, Communication & Society* 20(1): 1–13.
- Beraldo D and Milan S (2019) From data politics to the contentious politics of data. *Big Data & Society* 6(2): 1–11.
- boyd D and Crawford K (2012) Critical questions for big data: provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication & Society* 15(5): 662–679.
- Broomfield H and Reutter L (2019) *Kunstig Intelligens/data Science: En Kartlegging Av Status, Utfordringer Og Behov I Norsk Offentlig Sektor—Første Resultater*. Trondheim: Universitet i Oslo, NTNU, Difi. Available at: <https://ntnuopen.ntnu.no/ntnu-xmlui/handle/11250/2634733>
- Bucher T (2018) *If. . . Then : Algorithmic Power and Politics*. Oxford Studies in Digital Politics. New York: Oxford University Press.
- Caplan R, Clark M and Partin W (2020) *Against Platform Determinism—A Critical Orientation*. Available at: <https://points.datasociety.net/against-platform-determinism-899acd88a3d>
- Dencik L (2019) Situating practices in datafication—from above and below. In: Stephansen HC and Trere E (eds) *Citizen Media and Practices*. London and New York: Routledge.
- Dencik L and Kaun A (2020) Datafication and the welfare state. *Global Perspectives* 1(1): 12912.
- Dencik L, Hintz A and Wahl-Jorgensen K (2019) *Digital Citizenship in a Datafied Society*. Cambridge: Polity Press.
- Dencik L, Redden J, Hintz A, et al. (2019) The 'golden view': data-driven governance in the scoring society. *Internet Policy Review* 8(2): 1–24.

- Elish MC and boyd D (2018) Situating methods in the magic of big data and AI. *Communication Monographs* 85(1): 57–80.
- Eubanks V (2018) *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*. New York: St. Martin's Press.
- Hagendorff T and Wezel K (2019) 15 challenges for AI: or what AI (currently) can't do. *AI & Society* 35: 355–365.
- Hockenhuil M and Cohn ML (2021) Hot air and corporate sociotechnical imaginaries: performing and translating digital futures in the Danish tech scene. *New Media & Society* 23(2): 302–321.
- Jasanoff S and Kim S-H (2009) Containing the atom: sociotechnical imaginaries and nuclear power in the United States and South Korea. *Minerva* 47(2): 119.
- Jasanoff S and Kim S-H (2015) *Dreamscapes of Modernity: Sociotechnical Imaginaries and the Fabrication of Power*. London: University of Chicago Press.
- Kazansky B and Milan S (2021) “Bodies not templates”: contesting dominant algorithmic imaginaries. *New Media & Society* 23(2): 363–381.
- Kitchin R (2014) *The Data Revolution: Big Data, Open Data, Data Infrastructures & Their Consequences*. London: SAGE.
- Kitchin R (2017) Thinking critically about and researching algorithms. *Information, Communication & Society* 20(1): 14–29.
- Kitchin R and Laurialt T (2014) Towards critical data studies: charting and unpacking data assemblages and their work. The Programmable City Working Paper 2. Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2474112
- Lehtiniemi T and Ruckenstein M (2019) The social imaginaries of data activism. *Big Data & Society* 6: 1–12.
- Maciejewski M (2017) To do more, better, faster and more cheaply: using big data in public administration. *International Review of Administrative Sciences* 83(1_suppl): 120–135.
- Mager A and Katzenbach C (2021) Future imaginaries in the making and governing of digital technology: multiple, contested, commodified. *New Media & Society* 23(2): 223–236.
- Mayer-Schönberger V and Cukier K (2013) *Big Data: A Revolution That Will Transform How We Live, Work and Think*. London: John Murray.
- Mejias UA and Couldry N (2019) Datafication. *Internet Policy Review* 8(4): 1–10.
- Mergel I, Rethemeyer RK and Isett K (2016) Big data in public affairs. *Public Administration Review* 76(6): 928–937.
- Ministry of Local Government and Modernisation (2020) *Nasjonal Strategi for Kunstig Intelligens*. Available at: <https://www.regjeringen.no/no/dokumenter/nasjonal-strategi-for-kunstig-intelligens/id2685594/>
- Misuraca g and van Noordt C (2020) *Artificial intelligence in public services—overview of the use and impact of AI in public services in the EU*. EUR 30255 EN. Luxembourg: Publications Office of the European Union.
- Moats D and Seaver N (2019) “You social scientists love mind games”: experimenting in the “divide” between data science and critical algorithm studies. *Big Data & Society* 6(1): 1–11.
- O’Neil C (2016) *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. New York: Crown Publishing Group.
- Pasquale F (2015) *The Black Box Society: The Secret Algorithms That Control Money and Information*. Cambridge: Harvard University Press.
- Plesner U and Justesen L (2021) The double darkness of digitalization: shaping digital-ready legislation to reshape the conditions for public-sector digitalization. *Science, Technology, & Human Values* 47(1): 146–173.
- Public Administration Act (2021) Act relating to procedure in cases concerning the public administration (LOV-1967-02-10). Available at: https://lovdata.no/dokument/NLE/lov/1967-02-10#KAPITTEL_3

- Redden J (2018) Democratic governance in an age of datafication: lessons from mapping government discourses and practices. *Big Data & Society* 5(2): 1–13.
- Redden J and Grant A (2020) Learning from cancelled systems. Available at: <https://www.themj.co.uk/Learning-from-cancelled-systems/218519>
- Redden J, Dencik L and Warne H (2020) Datafied child welfare services: unpacking politics, economics and power. *Policy Studies* 41(5): 507–526.
- Rieder G and Simon J (2016) Datatrust: or, the political quest for numerical evidence and the epistemologies of Big Data. *Big Data & Society* 3(1): 1–6.
- Ruppert E (2018) Sociotechnical imaginaries of different data futures. 3e Van Doornlezing. Available at: <https://www.eur.nl/sites/corporate/files/2018-06/3e%20van%20doornlezing%20evelyn%20ruppert.pdf>
- Seaver N (2017) Algorithms as culture: some tactics for the ethnography of algorithmic systems. *Big Data & Society* 4(2): 1–12.
- Skatteetaten (2025) (2017) Skatteetaten 2025: virksomhetsstrategi for strategiperioden 2018–2021. Available at: <https://www.skatteetaten.no/globalassets/om-skatteetaten/om-oss/strategi/virk-somhetsstrategi-2018-2021.pdf>
- Strauß S (2015) Datafication and the seductive power of uncertainty—a critical exploration of Big Data enthusiasm. *Information* 6(4): 836–847.
- Tanweer A, Fiore-Gartland B and Aragon C (2016) Impediment to insight to innovation: understanding data assemblages through the breakdown—repair process. *Information, Communication & Society* 19(6): 736–752.
- Tavory I and Timmermans S (2014) *Abductive Analysis – Theorizing Qualitative Research*. London: The University of Chicago Press.
- Thomas SL, Nafus D and Sherman J (2018) Algorithms as fetish: faith and possibility in algorithmic work. *Big Data & Society* 5(1): 1–11.
- Tupasela A, Snell K and Tarkkala H (2020) The Nordic data imaginary. *Big Data & Society* 7(1): 1–13.
- van Dijck J (2014) Datafication, dataism and dataveillance: big data between scientific paradigm and ideology. *Surveillance & Society* 12(2): 197–208.
- van Ooijen C, Ubaldi B and Welby B (2019) *A data-driven public sector: enabling the strategic use of data for productive, inclusive and trustworthy governance*. OECD Working Papers on Public Governance. Paris: OECD Publishing. Available at: OECD Working Papers on Public Governance.
- Williamson B (2018) Silicon startup schools: technocracy, algorithmic imaginaries and venture philanthropy in corporate education reform. *Critical Studies in Education* 59(2): 218–236.
- Yeung K (2020) From new public management to new public analytics. ADM Workshop 1: Conceptualizing Automation in the Welfare Sector.
- Ziewitz M (2015) Governing algorithms: myth, mess, and methods. *Science, Technology, & Human Values* 41(1): 3–16.
- Zuboff S (2019) *The Age of Surveillance Capitalism: The Fight for the Future at the New Frontier of Power*. London: Profile Books.

Author biography

Lisa Reutter is a PhD candidate at the Norwegian University of Science and Technology, researching the (re)construction of algorithmic governance in the Norwegian public sector. She is part of the interdisciplinary research project Digital Infrastructures and Citizen Empowerment (DICE). Her research is located at the intersection between the fields of public administration research, sociology, and science and technology studies.