5

Practitioner Interventions in Data Power

Hadley Beresford,[1] Iris Muis,[2] Susan Oman,[3] Joanna Redden,[4] Elise Renkema,[5] Marlee Tichenor,[6] Caitlin Bentley,[7] Teresa Cerratto-Pargman,[8] and Jo Bates[9]

Introduction

Are we entering a new phase of urgency over crises in ethical data practices? Indeed, what is data work? At the time of writing, the media is filled with predictions and proposals that generative artificial intelligence (AI) will be responsible for the end of the world. More mundanely, there has been much concern that increasingly automated processes will displace both unskilled and skilled labour. This includes a variety of data work. This chapter reflects on the nature of data work: what it looks like in practice and how it is differentiated across various data practitioners. We do so from a critical data studies position and the hope that in reflecting on the growing variety of practitioners, we may contribute to the data work that intervenes in data power.

Data work involves an ecosystem of interdependent roles, professions, and actors, as Beresford illustrates below. They draw from across domains and sectors, explaining that while policy makers and others in power are increasingly called upon to reflect on what good or fair data practices may involve, data scientists are increasingly called upon to create solutions that fit whatever neat or narrow criteria may be set out. Tichenor's research with those who identify as part of a 'global statistical community'[10] found that, while this group would not identify as data scientists as a rule, their professional identity was dependent on ideas that they were doing good with data. Yet, they were resistant to the rise of data science, as something at odds with traditional statistics as a practice of governance that has used data for hundreds of years, and driven by an ethics of care for those found within the data. These participants were reflective about the negative impacts of their work on certain populations. In short, they felt accountable for their counting practices and acted with a generative tension with various framings of data work.

Research with council members in the Netherlands, by contrast, found that these people who work with data to make decisions that affect populations distanced themselves again. Renkema and Muis suggest that these people who work with data consider themselves 'laypersons', who are indeed not data professionals, thus becoming unaccountable for decisions made. When is someone data literate 'enough' to make decisions about data and algorithms? And, who decides what the demarcation of data literacy is in local governance? Ask Renkema and Muis.

Oman comments on the UK government's line that they were 'following the data' to deny accountability when making controversial policy decisions, particularly during the COVID-19 pandemic. The empirical research she discusses here, however, talks about those in the data ecosystem who have discomfort with their day-to-day data responsibilities, feeling that they do not have the data expertise to collect or process these data. In short, instead of worrying about data for good, or what a good data practice is, they worry about whether they are good enough to work with data.

The chapter concludes with a reflection on what data work is, and how training data scientists in being more reflective and aware of the impacts of the work of data may improve future data practices. Redden extends 'data work' to those important activities that activists, journalists, and members of affected communities have taken on in order to limit the negative effects of automated decision systems (ADS). Redden and her colleagues' work has shown that an often critical factor in the removal of a harmful ADS has been civil society critique and mobilization. In this way, we also assert that the work of critical data studies is data work in itself – as an attempt to capture what is hidden, nefarious, and/or negligently injurious in current practices with the hope that data futures can be better for all. Further, that critical data studies has much to gain by recognizing and learning with those engaged in political and power struggles about if and how data should be used in practice.

Data scientists and algorithmic bias mitigation in the public sector – Hadley Beresford

In recent years, the UK civil service, as well as the public sector more broadly, have moved towards further utilizing data-based approaches, specifically in the form of algorithmic technologies (Algorithm Watch, 2019; Big Brother Watch, 2021). The increased use of algorithmic technologies in the UK context has been at least somewhat influenced by the legacy of austerity policies – local councils and public-sector departments are having to 'do more with less' (Dencik et al, 2018; Oswald et al, 2018). Indeed, Eubanks (2018) notes that the increase of algorithmic technologies in public services has arrived alongside the rapidly rising economic insecurity of the last decade. As these technologies enter more mainstream use, cases of algorithmic bias have become more prevalent (Angwin et al, 2016; O'Neil, 2017; Eubanks, 2018). Algorithmic bias describes how, through a combination of social, technical, and probabilistic mechanisms, some people are penalized, or denied opportunities, due to their membership of a marginalized group.

Thus far, many of the solutions data scientists have produced to mitigate the risks of algorithmic bias have been technical or procedural in nature, including debiasing techniques which focus on ensuring datasets are either more representative of their target population, or relying on achieving statistical parity when comparing the outcomes of different groups based on protected characteristics (Galhotra et al, 2017). Other approaches have included finding new ways to operationalize the concept of fairness within a statistical framework, allowing data scientists to better perform statistical checks on their models (Bellamy et al, 2018).

These types of technical checks fit nicely into the skillsets of data scientists, however, these methods have been criticized by academics and third-sector organizations for not addressing the social structures and processes that contribute to algorithmic bias. Focusing on biases which may exist within datasets and statistical methods, critics argue, means inadequate attention is paid to the ways these systems embed biases from the socio-economic system the algorithm is deployed within (Balayn and Gürses, 2021). In Hoffmann's paper 'Where fairness fails', she critiques these types of technical solutions as positioning the problem as one which is caused by 'bad algorithms' which can be fixed, distracting from consideration of the wider socio-economic influences and logics these algorithms are embedded within (Hoffmann, 2019). Hoffmann, instead, calls for a focus on justice, rather than fairness, to consider these wider socio-economic influences (Hoffmann, 2019). Similarly, Green calls for data scientists to become cognizant of the political assumptions embedded within these technologies and question them within their own working practices (Green, 2021).

While these academic perspectives help distinguish between different ways of operationalizing the 'social good', and draw attention to implications attached to different framings, there is still uncertainty around the types of strategies public-sector data practitioners can utilize in their day-to-day work. When I have spoken to public-sector data scientists, many genuinely want to find working practices which produce 'good' and non-discriminatory algorithms. However, the definition of 'social good' promoted in these cultures is often heavily structured by the political and legal structures these data practitioners work within. Often, these practitioners are creating digital 'products' designed to serve a wide-ranging spectrum of citizens – each of whom will have their own conception of the 'social good'. Additionally, data practitioners find the types of approaches available to them are strongly influenced by legal structures (Orr and Davis, 2020). Furthermore, data practitioners are, unfortunately, having to 'do more with less', due to increasing workloads with little extra funding from government bodies (Oswald et al, 2018). These combined issues mean socially motivated practitioners find themselves trying to work towards some form of 'social good' amenable to the context they work within, with the outcome of this process often uncertain.

Moreover, it is not only the framing of these problems which restricts data practitioners' ability to work towards making tangible improvements in this area. Algorithmic technologies are not the providence of data scientists alone – these technologies are created using multidisciplinary teams, including policy experts, ethics experts, data engineers, administrative staff, and front-line staff responsible for data collection. Depending on the service, department, or project in question, these actors can be incredibly varied. This means effort is required to create a shared understanding of these things across stakeholders with often disparate skillsets, expertise, knowledge-bases, values, and beliefs. There is a need for practitioners to create interdisciplinary processes which allow them to address contextual organizational and structural constraints in public-sector algorithmic bias mitigation.

Sustainable development data practitioners – Marlee Tichenor

In many ways, in our data-laden world, we are all data practitioners, and as such we are always negotiating the ways that constantly collected and processed data help constitute our understandings of phenomena, both global and intimate. Thus, perhaps it is unsurprising that there would be such tension around who defines themselves a 'data practitioner' in the context of these phenomena constituted by data. However, to give it some specificity, the data practitioners with whom I most recently worked were interested in representing social, economic, and environmental problems as global problems, with the goal of intervening upon them and monitoring progress on these interventions. They were trained in statistics, international development, or economics, and they called themselves members of the 'global statistical community', connected by the annual United Nations Statistical Commission (UNSC) and its various working groups, and currently shaped by the context of the Sustainable Development Goals (SDGs).

The SDGs were shaped and agreed upon by the United Nations General Assembly, whose deliberations were fed in by various consultancy processes. These processes included the UN Secretary General-run 'Post-2015

Development Agenda' – whose goal was to build a next-generation version of the Millennium Development Goals (MDGs), and which has been widely criticized as a donor-driven set of goals largely directed at low- and lower middle-income countries (for example, Amin, 2006; Saith, 2006) – and the Open Working Group that emerged from the Rio+20 Conference on the Environment and Development – which was driven by a country-led coalition and whose goal was to foment more structural change (Fukuda-Parr and McNeill, 2019). However, it was also driven by an increased attention to official statistics and development data in the global governance space – partly fuelled by the previously mentioned MDGs and the monitoring of their indicators towards success – a flurry of initiatives and declarations whose apex might be best captured in the UN's report, 'A world that counts: Mobilising the data revolution for sustainable development' (IEAG, 2014). In this way, at least rhetorically, the SDGs were to be shaped by the twin revolutions of data-driven development and of country-led coalitions from the Global South to reverse the usual top-down approach to development (Bandola-Gill et al, 2022).

The sustainable development data practitioners driving that first engine made up a heterogeneous group, and most of them would not consider themselves data scientists. In fact, in my interviews with them, it was frequently mentioned that there was palpable tension and discomfort in the community about the rise of data science in the last 15 years, in this space that has been shaped by statistical thinking since the first annual UNSC in 1947 – what one interlocutor called "the traditional official statistical universe" (UN Statistician, 3).[11] As my colleagues and I have shown elsewhere, these data practitioners and the work that they do are shaped by three kinds of reflexivity: epistemic, care-ful, and instrumental (Bandola-Gill et al, 2023). Many of our interlocutors were quite aware of the performative nature of the decisions they made about the global measurement of social, economic, and environmental problems (Callon, 2006) – they acknowledged the epistemic limits of measurement and the effects of these decisions on 'distorting' policy on the domestic level down the line. We found that, for many of them, their work was also framed by a 'duty of care' to those whose lives they represented in numbers, precisely to attempt to minimize the harm produced by the limits of numbers – they were motivated by caring for these populations through number production. Finally, we found that these two forms of reflexivity became resources for these actors to use to negotiate the politics around promoting particular agendas – they instrumentalized reflexivity as a way to build trust with stakeholders and to build consensus.

While engaging in these reflexive data practices, tensions arose between the intention of caring for these populations through enumeration and data disaggregation, and the outcomes of these practices of enumeration and data disaggregation. This includes the frequent conceptual collapsing of the practice of enumeration with addressing social harms, a framing that is often used to advocate and seek funding for the labour and materials needed for constant monitoring and enumeration. As one interlocutor acknowledged, with the case of building and supporting civil registration and vital statistics (CRVS) systems globally – systems which are less robust in many parts of the world – there is this perception produced by "those [CRVS] guys" that "if one got to universal civil registration and timely production of vital statistics, magically … it'll just have a flow on effect and so many other development problems and health problems will be resolved" (UN Statistician, 13). He noted that this illusion of collapsing measurement with solving development problems was a "huge Achilles' heel" of efforts to promote effective statistical capacity around the SDGs and global agendas in general. In this way, these sustainable development data practitioners conducted their work through tensions in their identities as such, as well as in the gaps between their actions' efforts and their felt effects.

Democratic control over data and AI projects in the local public sector – Elise Renkema and Iris Muis

In this chapter, we share our initial findings and experiences with interventions in the field of datafication and algorithmization of policy making and the role of elected representatives within this trend. We argue for an empirically driven and socially engaged practice of critical data studies, covering local contexts of data practices. This provides in-depth insight into the discourses of data power, facilitating effective knowledge transfer and social engagement for building a fair and open digital society.

During our work with municipalities in the Netherlands, we focused specifically on elected representatives who were seated in city councils. We interacted with them through teaching masterclasses around themes of datafication, algorithmization, and democratic control. An example of this is the masterclass series for city council members of the city of Utrecht, which consisted of four two-hour meetings. An array of topics was discussed, from the impact of sharing platforms on public space to formulating a political stance on prediction algorithms.

Most research in the field of critical data studies in government focuses on policy officers (civil servants), but not a lot of work has been done on researching the role of elected representatives in the datafied society. There has only been one significant research report on the role of city council members (Das et al, 2020), its main conclusion being that they lacked sufficient knowledge and expertise to fulfil their role as elected representatives when it comes to processes of digital technology. This is a conclusion we also made during our research.

We have identified several challenges of council members to fulfil their representing and controlling role. One main issue is how council members perceive their relationship to data and algorithms. Council members often say they are fulfilling their role in the council as 'laypersons', someone who is not a professional in the given field. We have noticed that council members are often either unsure of their ability to make policy or control decisions about data and algorithms or do not see that it is their responsibility. This brings us to the question: when is someone data literate 'enough' to make decisions about data and algorithms? And, who decides what the demarcation of data literacy is in local governance? What complicates this matter is that the political arena is often not considered a 'safe space' to learn or make mistakes.

A second issue is that not all council members see that data and algorithms belong in the political arena, due to it being seen as a matter of execution and a non-political issue and therefore not their responsibility. "Leave the IT questions to the IT experts, I do not even know how to properly operate my own laptop", is something we often hear. We have also noticed that some council members perceive that data can be seen as a neutral tool. As a result, they have not developed an ideological vision on this matter. Moreover, the issue might not reach the political agenda of the council because it stays in the administrative part of local government.

Overall, this can be seen as a question of identification. Do elected representatives see themselves as competent enough to provide democratic control over data projects and algorithms? And, are such projects identified as part of the political arena by elected representatives? We argue for more research on how the domain of digitalization and algorithmization can be embedded in existing democratic institutional processes. Such research should be embedded in the local contexts of data practices in order to understand how elected representatives, government officials and public servants perceive their responsibility in achieving sound accountability and control in the domain of digitalization and algorithmization.

Reflecting on the categories of data work in practice – Susan Oman

As a Lecturer in Data, AI and Society at the UK's University of Sheffield, I teach our future data scientists (broadly defined) the ethical, practical, environmental, and social limitations and ramifications of data practices. The process of our discussions in the co-authoring of this chapter has led me to reflect on how teaching is not only a 'practitioner intervention in data power', but should be considered a data practice, in and of itself. It is a pedagogical intervention in future data work (or, data work futures), through the shaping of reflective data practitioners. This work is fuelled by the aspiration that future data practitioners will not only be able to reflect on the implications of their own work on people and society, but how it connects to the wider, complex ecosystem of data practices. Teaching responsible data sciences is, therefore, a category of data labour and data work.

Teaching through contextualizing data practices by 'following them around' (Oman, 2021c) the wider data ecosystem, makes them less abstract to students, and crucially, almost impossible to conceive of data without them being connected to people and practices. Who does what with whose data, and who is affected, become urgent questions for those who had perhaps not considered these questions before. The aspiration is that cohorts of future data practitioners will broaden in perspective, life experience, knowledge, and understanding – while developing the core skills expected of a data science programme. Thus, potentially changing what a future data practitioner is as the affective capacities of data and their dynamics become clearer through their learning journeys.

My research looks at what I call 'data contexts' (Oman, 2021a): how data work in different contexts and how different kinds of data work for and against people, differently. I have often focused on administrative data that are routinely collected for social good: to understand well-being or inequality, for example. Much of my research aims to understand different aspects of the public sector and its uses of data, and their implications (whether known or overlooked), and I am currently on an AHRC-funded policy placement in the UK government's department responsible for data and AI.

There are, therefore, many 'layers' of data work (Oman, 2021b) that extend far beyond the traditional idea of a data scientist. In my research, I talk to all kinds of people who work with data across these 'layers' of expertise and practice to understand the performativities of data, and how they 're-perform' (Oman, 2021b). One layer might be construed as the expert–policy nexus, ranging from senior people in national statistics offices, civil servants, and experts who guide or advise on data policy. Another layer is the expert in domain contexts, such as those working in publicly funded organizations who collect and analyse data (including those who consider themselves non-expert). I also talk to people who do not work with data, or at least would not identify as doing so, and so consider this layer as the 'everyday understanding' of data. But, of course, we all work with data. Whether in our jobs, or completing a form to receive welfare or registering with a doctor. We are all doing the emotional labour of the dynamic data infrastructure that permeates all aspects of our lives. Crucially, we also all feel the relations between knowledge and power, even if we may not foreground this in this way when talking about what matters most to us.

We decided that this chapter's focus would be on the tensions of identity, context, and knowledge in relation to data power. As such, it speaks to what I have found across my research projects and professional roles that aim to intervene in bad data practices. This leads me to reflect on what a good data practice even is. Those in national statistics offices believe the administrative data that they work with are ostensibly used to inform knowledge on how to improve society. Those in public data roles often do not have the chance to reflect on the ways evidence and knowledge generated from their data practices are used. There are no reflections on the limits to an improved society. Yet these data about population well-being, or inequality, are pivotal in decisions that affect our life chances, livelihoods, and quality of life.

'Following the data' was a familiar phrase in COVID-19 policy communications for the UK government. If you follow data back centuries, you find it has long been used to track the health and wealth of society. Even when the assumptions that have underpinned over 200 years of social science, statistical and policy work are called into question at the practitioner level, are they questioned enough? Finding ways to follow data practices in their contexts is one way to reveal how data uses actually operate and what they enact. This critical data practice calls for, and, we hope, enables data practices more widely to become more reflective. It is, therefore, a category of data work that incorporates overlooked data labour practices in order to intervene in data power.

Data work by activists, journalists, and community – Joanna Redden

Throughout this chapter we have been drawing on our research and teaching to argue that data practitioners work in a range of roles that extend beyond data science. Our goal is to expand how we think about data work and who is doing that work. The previous sections of this chapter have focused on the data work being done by civil servants, statisticians, economists, international development workers, teachers, and elected representatives. As noted by Oman, there are 'many layers of data work' across a range of contexts. As suggested by Tichenor, in a world so mediated by data practices we are all becoming data practitioners as we confront the ubiquitous collection of our data as datafied ways of knowing influence the ways we are scored, provided or denied opportunities, targeted, and are able to access needed services.

As detailed by Hadley, data practitioners are influenced and often limited by the contextual forces they work within. Renkema and Muis raise important questions about power imbalances, our abilities to know and influence data practices, as well as the extent to which our democratic institutions are at present equipped to ensure accountability, effective oversight, and prevent harm while protecting human rights.

The work I have been doing with my Data Justice Lab colleagues adds to who we might think of as data practitioners by demonstrating the key data work being done by activists, civil society organizations, members of affected communities, and journalists. After previous work mapping and analysing the uses of ADS across public services as well as discussing data harms and resistance, our recent research involved investigating where and why government agencies had decided to cancel their use of ADS in Australia, Canada, Europe, New Zealand, the United Kingdom, and the

United States (Redden et al, 2022). We identified 61 systems cancelled and that a range of factors had led to cancellation. One of our central findings is that nearly half of the systems cancelled had been subject to civil society critique and mobilization as well as critical media coverage. In many cases activist mobilizations and media reporting made the existence of systems in use visible as well as the impacts such systems were having on people. We also found that nearly one third of cancelled systems were stopped as a result of legal action.

This work demonstrates that contrary to the view of some that ADS are neutral and apolitical, their use raises significant concerns related to rights and impacts. Through case study investigations involving interviews and document analysis we stress that there are a range of factors leading to cancellation, often working together. In addition to the factors listed earlier we also found the concerns of civil servants about systems not being as effective as promised as significant, as well as concerns about potential discrimination and bias. In combination, our study of cancelled systems reinforces previous work that stresses that technologies are sites of struggle and power, and that there are competing values, politics, and visions informing decisions about how technologies are employed (Eubanks; 2018; Benjamin, 2019).

Some examples of the key role played by activists, affected communities, civil society organizations, and journalists include work done in Los Angeles in response to police uses of predictive policing. The Stop LAPD Spying Coalition have been credited with researching the impact of police uses of these ADS systems. Their research demonstrated how systems called PredPol and Laser disproportionately negatively affected Black, Latinx, and other people of colour. Their research was featured in media coverage. Stop LAPD Spying Coalition community mobilizing work included calling on the Office of the Inspector General to review the systems. The Inspector General's report pointed to a range of problems. Laser was suspended in 2019 and PredPol in 2020. Similar community mobilizations, critical reporting, and litigation in Australia in response to its automated fraud detection system dubbed Robodebt, led to that system being cancelled in 2019; a judge ruled the system unlawful and a class action lawsuit was settled for $1.2 billion. These are only two examples of community mobilization discussed in our report. They demonstrate the kind of work being done to challenge the distancing and reductive effect of ADS. They also demonstrate the importance of care and community and present lessons in how data literacy and mobilization can be done together.

Conclusion

In combination, our research points to the importance of critical interventions into how we think of data practices and who is a data practitioner. Given the power and ubiquity of new and emerging data practices, we stress the importance of recognizing the data work being done across our societies, and in particular the tensions and struggles connected to datafication. While we focus on different areas of data practice, we each see ongoing concerns around governance and what good practice is and should be. Data work, as detailed throughout this chapter, involves contending with the ways contemporary uses of data are built upon longer histories of unjust, capitalist and colonialist ways of knowing and controlling. Tensions emerging around issues of care and connection versus harm and distancing are linked to context-specific limitations but also competing political visions. A politics of care approach stresses the importance of our connections and interdependence with each other and the natural world (Chassmen and Cohen, 2020; Woodly et al, 2021). In terms of data work, this would involve ensuring decision-making about data practices that is historically and contextually informed while centring social, economic, and human rights. In the context of automation, critical refusal has been argued as necessary for decision-making (Cifor et al, 2019; Gangadharan, 2021; Hoffmann, 2021). As our quotidian social, political, and economic practices continue to coalesce with our data practices – a merging that seems particularly rapid with the current exponential development of large language models – it is important for critical data scholars to take stock of the new forms of responsibility and care that arise, while helping construct our potential data futures. This is the critical data work we aim to present as our own practitioner interventions in data power.

DISCUSSANT RESPONSES

Rethinking data practice – Teresa Cerratto Pargman[12]

'Practitioner Interventions in Data Power' written by Hadley Beresford, Iris Muis, Susan Oman, Joanna Redden, Elise Renkema, and Marlee Tichenor is a compelling, timely, and critical piece of collective work on data power (Hepp et al, 2022). It contributes accounts of what data work entails in practice viewed from different practitioners' roles, mandates, and perspectives. Following the data work conducted by civil servants, local elected representatives, statisticians, economists, international development workers, teachers, and activists, readers will find in this chapter a call to question underlying assumptions regarding who a data practitioner is, how a data practitioner becomes one, and what data work entails in practice.

The chapter invites us to rethink what we understand by data literacy and responsibility as data are not generated, shared, analysed, and used only by data scientists (professionals) but also by the citizens in the different roles they occupy in an increasingly datafied society. This particular understanding that 'we all work with data', there are 'many layers' of data work, and 'we are all becoming data practitioners as we confront the ubiquitous collection of our data' is at the core of this collective work that seeks to generate debate about the following questions: What is a good data practice and what should a good practice be? When is someone data literate enough to make decisions about data and algorithms? How can the domain of digitalization and algorithmization be embedded in existing democratic institutional processes? How do data practitioners perceive their capacity to work with data? How do they see or define their responsibilities?

These questions echo some of the enquiries driving our own research work in the project 'Ethical and Legal Challenges in Relationship with Emerging AI-driven Practices in Higher Education', funded by the Wallenberg Foundations in Sweden. In this project, we engage with issues regarding responsibility for and accountability vis-à-vis the student data that are collected, shared, analysed, and stored in the institutional management systems, online invigilator systems, or automated grading systems, fundamentally configuring the communication between students and teachers/institutions. As the authors in the chapter, we also view the need to conduct empirical studies to contribute to critical studies on data that inform how data practices unfold in situ and reflect on the importance of contextualizing data practices, making them less abstract, as well as connecting them to real consequences for different groups of people in society. Such empirical studies on situated data practices are most needed today to strengthen critical arguments contending the 'structural power differentials in society and the work necessary toward dismantling them' (D'Ignazio and Klein, 2019). Drawing on data feminism (D'Ignazio and Klein, 2019), we also call for a critical understanding regarding where AI systems, hungry for student data, should (and should not) be deployed, for whom, and why (Cerratto Pargman et al, 2023).

The chapter also reminds us that far from being 'objective', 'neat', and 'flawless', data are a site of struggle and power, reflecting competing values, especially when data are in the hands of governments and private global corporations. In this sense, the chapter emphasizes the relation between data as power and algorithmic bias by making clear that algorithmic bias mitigation in the public sector is not only a problem of the current limitations of the different types of technical checks available but also of the broader sociocultural, politico-economic context (people's practices) in which these algorithms operate. As such, the chapter convincingly argues for acknowledging data practice as involving not only data scientists but also a large variety of social and political actors with often disparate skills, knowledge, values, and interests. This renewed and fresh understanding of data practices also points to the heterogeneous composition of data practitioners and, thus, to the tensions and conflicting interests emerging among those dealing with, speaking about, and making daily decisions regarding people's data.

In summary, this is an original chapter of interest to anyone studying and working with data, algorithms, and people. It shares food for thought based on concrete, real experiences gained from the field that compels us to revise taken-for-granted ideas about data practices, data practitioners, and the inherent 'tensions of identity, context, and knowledge in relation to data power'.

Questioning data practitioner tropes and the need for diverse practitioner roles and responsibilities in data work – Caitlin Bentley[13]

The chapter, which examines the landscape of data work and its ethical dimensions, is an enlightening contribution to critical data studies. The authors shed light on the complex layers of what constitutes data work and how this is perceived differently by various actors – be it policy makers, data scientists, or those who see themselves as laypersons in the context of data work. The authors make a compelling case for the importance of critical data studies as a form of data work itself, capturing hidden, harmful, or negligent practices with the aim of making the future of data more equitable and ethical. It pushes the reader to not only question what data work is but also what it ought to be, compelling us to consider ethical considerations and power struggles that are usually glossed over.

The chapter opens with a timely question about the urgency of ethical considerations in data practices, which is especially relevant given current debates surrounding AI, automation, and labour transformation. But who is responsible for making data work more responsible and ethical in reality? The authors adopt a critical data studies lens to examine the wide array of practitioners involved in data work. What stands out is the chapter's incorporation of diverse perspectives – often highlighting how those that do data work, may not even realize it or identify as data workers. Indeed, the chapter does not shy away from highlighting the varying degrees of accountability – or lack thereof – among different actors in multiple data ecosystems. These accountability tensions come to bear, the authors argue, when we visibilize data work, who it is done by, and how.

In my own research, as an academic at a UK university, we have been investigating the requisite skill set that is needed to make data work more responsible and reflective. When examining frameworks to build on, one approach that has been quite common is categorizing skills based on roles or professions. This delineation, albeit clear, presented itself with a plethora of limitations, and we opted for a more flexible general skills framework encompassing a variety of technical, professional, and strategic skills that could be fit for purpose. Our resolution to adopt a flexible approach stemmed from our position that data and AI ecosystems are dynamic and rapidly changing. The roles and professions that are relevant today may undergo transformations, branching out into specialties or amalgamating into broader roles. In this flux, we focused on how a rigid role-based structure might foster obsolescence rather than agility. However, the chapter also shows that data ecosystems are vibrant tapestries, often blurring traditional professional boundaries, becoming sites where roles that might not conventionally be identified as data-centric play a pivotal role in shaping data narratives. Beresford rightly points out that algorithmic technologies are created by multidisciplinary teams.

The chapter thus aptly emphasizes that data work is not confined to those who are fully immersed in data roles; it stretches to encompass individuals from different realms who are gradually becoming a part of the discourse. This necessitates capacity to develop self-awareness around data work, listening to others, melding essential data work skills from various professions, thus nurturing a fertile ground for responsible innovation and ethical data practice.

As my interest also relates to this arena in which multidisciplinary conversations happen, I was really hoping that the chapter could have delved deeper into specific methods, processes, or questions that data workers use in their reflections and debates. Tichenor's presentation of three types of reflexivity that sustainable development data practitioners use (epistemic, care-ful, and instrumental), and Oman's 'data contexts' and notion around 'following the data' point to innovative reflection techniques that could be implemented throughout data work processes, whereas Renkema and Muis and Redden's contributions suggest punctual activities that could precede or follow data work. Renkema and Muis, in their work with municipal actors in the Netherlands, question how much data literacy these actors need to be able to reflect and make appropriate decisions, suggesting there may be preceding factors before a reflection process. In contrast, Redden examined the use of automated decision-making in public-sector services for data harms and resistance, presumably using qualitative research methods post hoc. I hope another book can be written detailing all of these methods, describing and laying the roadmap for how critical reflection can be done by whom, when, and how.

Overall, this chapter is a must-read for anyone engaged in data work, be it academically or professionally. It offers rich insights into the multiple facets of data work and poses critical questions that challenge existing norms and practices, all while advocating for more ethical and accountable data ecosystems.

Notes

1 University of Sheffield
2 Utrecht University
3 University of Sheffield

[4] University of Western Ontario
[5] Utrecht University
[6] University of Edinburgh
[7] Discussant
[8] Discussant
[9] Facilitator
[10] A community connected by the annual United Nations Statistical Commission (UNSC) and its various working groups, and currently shaped by the context of the Sustainable Development Goals (SDGs).
[11] This is from an interview I conducted with a statistician tied to the United Nations, and this is how this person was coded in our project documentation and other publications on the topic.
[12] Stockholm University
[13] Kings College London

References

Algorithm Watch (2019) *Automating Society: Taking Stock of Automated Decision-Making in the EU*. Algorithm Watch. Available at: https://algorithmwatch.org/en/automating-society-introduction/

Amin, S. (2006) The Millennium Development Goals: A critique from the South. *Monthly Review*, 57(10): 1–15. Available at: https://monthlyreview.org/2006/03/01/the-millennium-development-goals-a-critique-from-the-south

Angwin, J., Larson, J., Mattu, S., and Kirchner, L. (2016) *Machine Bias*. ProPublica. Available at: www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

Balayn, A. and Gürses, S. (2021) *Beyond Debiasing: Regulating AI and its Inequalities*. European Digital Rights (EDRi). Available at: https://edri.org/wp-content/uploads/2021/09/EDRi_Beyond-Debiasing-Report_Online.pdf

Bandola-Gill, J., Grek, S., and Tichenor, M. (2022) *Governing the Sustainable Development Goals: Quantification in Global Public Policy*. Sustainable Development Goals Series. Cham: Springer International Publishing. Available at: https://doi.org/10.1007/978-3-031-03938-6

Bandola-Gill, J., Grek, S., and Tichenor, M. (2023) The rise of the reflexive expert? Epistemic, care-ful and instrumental reflexivity in global public policy. *Global Social Policy*, 0(0): 1–18. https://doi.org/10.1177/14680181221145382

Bellamy, R. K. E., Dey, K., Hind, M., Hoffman, S. C., Houde, S., Kannan, K., et al (2018) AI Fairness 360: An extensible toolkit for detecting, understanding, and mitigating unwanted algorithmic bias. arXiv [cs.AI]. Available at: http://arxiv.org/abs/1810.01943

Benjamin, R. (2019) *Race after Technology: Abolitionist Tools for the New Jim Code*. Cambridge, MA: Polity Books.

Big Brother Watch (2021) *Poverty Panopticon: The hidden algorithms shaping Britain's welfare state*. Big Brother Watch. Available at: https://bigbrotherwatch.org.uk/wp-content/uploads/2021/07/Poverty-Panopticon.pdf

Callon, M. (2006) What Does It Mean to Say Economics Is Performative? *CSI Working Papers Series*, 5, pp 1–58. École des Mines de Paris. Available at: http://halshs.archives-ouvertes.fr/docs/00/09/15/96/PDF/WP_CSI_005.pdf

Cerratto Pargman, T. C., McGrath, C., Viberg, O., and Knight, S. (2023) New vistas on responsible learning analytics: A data feminist perspective. *Journal of Learning Analytics*, 10(1): 133–148.

Chassmen, D. and Cohen, J. (2020) *The Politics of Care: From Covid 19 to Black Lives Matter*. New York: Verso.

Cifor, M., Garcia, P., Cowan, T. L., Rault, J., Sutherland, T., Chan, A., et al (2019) Feminist Data Manifest-No. Available at: www.manifestno.com/

D'Ignazio, C. and Klein, L. F. (2019) *Data Feminism*. Cambridge, MA: MIT Press.

Das, D., Faasse, P., Karstens, B., and Diederen, P. (2020) Raad weten met digitalisering – Hoe de gemeenteraad kan sturen op de maatschappelijke impact van digitale technologie (Rathenau Instituut).

Dencik, L., Hintz, A., Redden, J., and Warne, H., et al (2018) *Data Scores as Governance: Investigating uses of citizen scoring in public services*. Research Report. Cardiff University.

Eubanks, V. (2018) *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*. New York: St Martin's Press.

Fukuda-Parr, S. and McNeill, D. (2019) Knowledge and politics in setting and measuring the SDGs: Introduction to Special Issue. *Global Policy*, 10: 5–15. https://doi.org/10.1111/1758-5899.12604

Galhotra, S., Brun, Y., and Meliou, A. (2017) 'Fairness testing: Testing software for discrimination'. *Proceedings of the 2017 11th Joint Meeting on Foundations of Software Engineering*. ACM (ESEC/FSE'17). https://doi.org/10.48550/arXiv.1709.03221

Gangadharan, S. P. (2021) 'Digital Exclusion: A Politics of Refusal'. In: L. Bernholz, H. Landemore, and R. Reich (eds) *Digital Technology and Democratic Theory* (pp 113–140). Chicago: University of Chicago Press. https://doi.org/10.7208/9780226748603-005

Green, B. (2021) Data science as political action: Grounding data science in a politics of justice. *Journal of Social Computing*, 2(3): 249–265. https://doi.org/10.23919/jsc.2021.0029

Hepp, A., Jarke, J., and Kramp, L. (2022) 'New Perspectives in Critical Data Studies: The Ambivalences of Data Power—An Introduction'. In: A. Hepp, J. Jarke, and L. Kramp (eds) *New Perspectives in Critical Data Studies: The Ambivalences of Data Power* (pp 1–23). Cham: Springer International Publishing.

Hoffmann, A. L. (2019) Where fairness fails: Data, algorithms, and the limits of antidiscrimination discourse. *Information Communication and Society*, 22(7): 900–915. https://doi.org/10.1080/1369118X.2019.1573912

Hoffmann, A. L. (2021) Even when you are a solution you are a problem: An uncomfortable reflection on feminist data ethics. *Global Perspectives*, 2(1): 21335. https://doi.org/10.1525/gp.2021.21335

Independent Expert Advisory Group on a Data Revolution for Sustainable Development (IEAG) (2014) *A World that Counts: Mobilising the Data Revolution for Sustainable Development*. United Nations. Available at: www.undatarevolution.org/wp-content/uploads/2014/11/A-World-That-Counts.pdf

O'Neil, C. (2017) *Weapons of Math Destruction*. London: Penguin.

Oman, S. (2021a) *Understanding Well-being Data: Improving Social and Cultural Policy, Practice and Research*. London: Palgrave Macmillan. Available Open Access at: https://link.springer.com/book/10.1007/978-3-030-72937-0

Oman, S. (2021b) Re-performance: A critical and reparative methodology for everyday expertise and data practice in policy knowledge. *International Review of Public Policy*, 3(3). Available Open Access at: https://journals.openedition.org/irpp/1833

Oman, S. (2021c) *Following the Data*. [podcast]. 17 September 2021. Available at: https://well-beingdata.com/films-on-player-sheffield-ac-uk/

Orr, W. and Davis, J. L. (2020) Attributions of ethical responsibility by Artificial Intelligence practitioners. *Information Communication and Society*, 23(5): 719–735. https://doi.org/10.1080/1369118X.2020.1713842

Oswald, M., Grace, J., Urwin, S., and Barnes, G. C. (2018) 'Algorithmic risk assessment policing models: Lessons from the Durham HART model and 'Experimental' proportionality. *Information & Communications Technology Law*, 27(2): 223–250. https://doi.org/10.1080/13600834.2018.1458455

Redden, J., Brand, J., Sander, I., and Warne, H. (2022) *Automating Public Services: Learning from Cancelled Systems*. Carnegie UK Trust and Data Justice Lab, October. Available at: www.carnegieuktrust.org.uk/publications/automating-public-services-learning-from-cancelled-systems/

Saith, A. (2006) From universal values to Millennium Development Goals: Lost in translation. *Development and Change*, 37(6): 1167–1199. https://doi.org/10.1111/j.1467-7660.2006.00518.x

Woodly, D., Brown, R. H., Marin, M., Threadcraft, S., Harris, C. P., Syedullah, J. and Ticktin, M. (2021) The Politics of Care. *Contemporary Political Theory*, 20: 890–925. https://doi.org/10.1057/s41296-021-00515-8