3

In/visibilities in Data Studies: Methods, Tools, and Interventions

Miriam Fahimi,[1] Petter Falk, Jonathan W. Y. Gray,[2] Juliane Jarke,[3] Katharina Kinder-Kurlanda, Evan Light, Ellouise McGeachey, Itzelle Medina Perea, Nikolaus Poechhacker, Lindsay Poirier, Theo Röhle, Tamar Sharon, Marthe Stevens, Bernard van Gastel, Quinn White, and Irina Zakharova

Introduction

Research in data studies is rarely a straightforward pursuit and can lead one down existential pathways of both angst and joy. While the idea of 'data' may present itself as something simple enough to parse and probe, 'data' – their production, use, analysis, and manipulation – are complex to grapple with. The bounds of data are ambiguous; the meaning of data is polysemous, and the social impacts of data are diverse. Just as renderings of data focus attention on some things while eclipsing others, fixing any one critical analytical lens on data can provide insight on certain socio-political issues while inevitably precluding others from view. Rendering data visible in all its fine details thus necessarily instigates, provokes, and challenges power. For example, data are not only held but also defined by gatekeepers (for example, Big Tech corporations, governments) which makes access to and openness of data political. Diverging perspectives on which data count and the interventions to make them count or make the gatekeepers accountable provide another example.

These power imbalances underlie our explorations into methods, (conceptual) tools and interventions within data studies. This chapter examines data in/visibilities through a multi-perspectival lens – differentially rendering data with each turn of the kaleidoscope. The kaleidoscope is a tool that allows for multiple perspectives, and seemingly multiplies the reality of what we see.

Without it, the multiplicity of perspectives would not be possible. Bringing to the fore these different perspectives, we show how acts of rendering data visible – whether as data publishers, activists, or data studies scholars – quickly lead us to interrogating power dynamics and power imbalances. The sections meditate on a number of questions: How do we chart the contours of visibility for something as ephemeral and capricious as data? How do we evaluate what we can see – its truth, its accessibility, and its social value? How do we make sense of when and for whom data visibility is productive?

The sorts of data we probe in this chapter are diverse – ranging from government datasets to corporate databases to administrative algorithms. Sections of the chapter also fix their analytical lenses on diverse sites in a data lifecycle and interrogate the roles of a diverse set of gatekeepers in shaping data's meaning and perceptibility. The chapter is organized in three parts, each interrogating the intersections between in/visibility and power. In Part 1, authors discuss the challenges of (academic) knowledge production about and with data and explore ways to capture data's relativity, fluidity, and instability as well as trace and visualize the movement of data. For example, frameworks for studying frictions in the (often invisible) infrastructures of algorithms render visible actors, technologies, and their diverging values in handling data. In Part 2, authors offer empirical case studies of data renderings – analysing the consequences of data visibility while also reflecting on the methodological opportunities and challenges of foregrounding the values and norms encoded into data. In the final part, contributors outline tool-based interventions for bringing alternative data framings and narratives into visibility.

PART 1: CONCEPTUAL APPROACHES TO DEALING WITH DATA

Methodologies and performativities of data studies – Irina Zakharova[4]

The emerging interdisciplinary field of data studies is concerned with the role of digital data, algorithms, and automation in society. While research on digital data benefits from the variety of disciplinary, theoretical, and methodological approaches to digital data, scholars of datafied societies also deal with particularly complex empirical questions. Typically, any kind of research in social sciences needs to accommodate and explain complexity of a given empirical field, be it public administration, health, or research on everyday lived experiences of individual people. When studying datafied societies, however, researchers face several conceptual and practical challenges reflecting the characteristics and contingencies of digital data.

For example, one challenge emerges from an understanding of digital data as relational products of negotiations between various actors – technology

designers, policy makers, funders, and, sometimes, potential technology users or communities affected by technology design – rather than simple representations of social reality. Another conceptual challenge is in uncovering whose and which values and dispositions about the social reality certain data mean to represent. Furthermore, scholars and their study participants have to determine what data or algorithms are in a given research context. Digital data are not stable objects, the lifecycle of which can be followed through from their generation to their use and eventual decay, even the practitioners whose work we study struggle to provide clear definitions. In each step of their lives, data change forms, formats, ownership, and meanings in accordance with the individual, organizational, and political goals of the actors currently engaged with these data. Often the access to these data at different stages of their lives can be restricted through organizational and legal regulations, and by actors who own (access to) these data, while particularly public and nongovernmental organizations increasingly commit to opening their and others' data for further examination and use. To access and analyse such data and related technologies, researchers either rely on the resources and instruments provided by technological corporations or develop their own tools and methods.

The choice and selection of such instruments and methods is performative to what we know about the datafication of society. This performativity of research methods discussed by scholars of science and technology is crucial for the future paths of both academic and technology development (Law, 2004; Law et al, 2011; Ruppert et al, 2013). Usually, performativity is acknowledged as the potential of research activities to 'meddle' with the empirical reality, as methods textbooks and handbooks warn (early career) scholars about biases in research practice or advise on navigating research ethics and transparency. Instead, I argue here for taking performativity of research as a starting point in the methodological reflection on data studies. If we do not know the boundaries of our knowledge and do not reflect on how these boundaries come to be, it is quite difficult to imagine alternative paths for future development. Currently, various scholars engage with such a reflection given the pressing need for new, collective, and just futures for living with digital and automated technology (Dencik et al, 2022; Powell et al, 2022; Zakharova, 2022).

I argue that one way to reflect on the performativity of data studies is to combine methodological reflection with specific questions related to challenges in studying digital data. The performativity approach allows making visible which 'agential cuts' (Barad, 2007) scholars draw to conduct empirical studies. Drawing on John Law's methodological exploration of research practices in social sciences, there are four core interrelated elements of the research design (Law, 2004). These elements include, first, the researchers and how they position themselves within relevant academic

fields and empirical sites of practice. Second, how the researched persons and objects in their historical development constitute the object of study. Third, particular research processes, instruments, and procedures as well as guiding research politics and organizational contexts, such as research funding, comprise the bandwidth of practical factors shaping research design. Finally, the empirical site of practice where a particular research project is conducted provides empirical context of the methodological reflections.

To incorporate the challenges of data studies in the framework of the methods assemblages, the performativity perspective is crucial. Instead of searching for a unified definition of what digital data and related algorithmic systems are or should be, a performative approach to data studies embraces the multiplicity and uncertainty of societal datafication processes. The individual, situated, contextualized knowledge of what digital data stand for in particular research situations is what draws each method assemblage together and provides researchers with sensitivities to render their approaches visible and understandable in the interdisciplinary field that is data studies. The goal of such reflection is to develop new vocabularies that loosen the long-standing binaries of technology/society, nature/culture, research/practice.

In/visibilities of algorithmic infrastructure: frictions as methodological lens for algorithmic ethnography – Miriam Fahimi, Nikolaus Poechhacker, and Katharina Kinder-Kurlanda

Studying algorithms in development poses a methodological challenge for ethnographic research (Kitchin, 2017). Ethnographers need to deal with different moments of in/visibility of algorithmic systems, which are opaque black boxes (Burrell, 2016; Christin, 2020), diffuse and heterogeneous (Seaver, 2017), and even mysterious (Ziewitz, 2016) things for the ethnographer that undertakes to study them. Determining where algorithms begin and end, as well as the mechanisms that lead to the stabilization of values and norms, is therefore often challenging to observe. In such highly mediated environments, the modes of observation have to adapt, and ethnographers require specific expertise to be able to follow and understand practices with and by computers.

One possible approach to ethnographically grasping opaque and 'boring things' (Star, 1999) as algorithms is by thinking of them as infrastructure. Infrastructures sustain social structures (Bhaskar, 2008), they are a representation of the world, and they can encode inequalities (Star, 1999). They are especially effective, and tend to become unquestioned once established (Katzenbach, 2021). However, in the moment of their breakdown not only does infrastructure become visible again, but so do social norms and assumptions conveyed therein. Yet, it remains unclear in which situations an algorithmic infrastructure 'fades away' (Star, 1999).

When we understand algorithmic infrastructure as relational (Star and Ruhleder, 1996), friction – the tiny moments of breakdown during continuous maintenance efforts – represent exciting moments for ethnographers. Drawing on our own positionality as ethnographic researchers and science and technology studies scholars, we experienced how social norms and values on gender in a credit scoring system became visible after paying specific ethnographic attention to their mediation by the infrastructure and multiple situations of friction: there, in a European credit agency, personal data were stored in a database, including people's name, address, date of birth, and a (binary) gender entry. While all personal data in this database were editable, such as name changes following marriage, address changes resulting from relocation, or corrections of typographical errors in date of birth, the gender entry could not be edited, even in situations where corrections were necessary. While these frictions had not been recognized as such by the computational practitioners for decades, this changed with public debate and the introduction by law of a third gender option. In our case, it created friction for our interlocutors, namely when a gender entry was entered incorrectly, or when a person legally changed their gender entry. As a result, the practitioners started several attempts at tinkering with the database, for example, by making the 'correct' gender entry visible through a commenting function that was added. This can be seen as an ongoing maintenance of the infrastructure with the aim to keep it up to date with changing normative (e)valuations (Graham and Thrift, 2007). And yet, these maintenance efforts did not translate over to the logic of the scoring algorithm. The algorithm could not process the comments, but only the original data entries, to describe the person's gender. Thus, the efforts to make misclassifications visible via tinkering and ex-post workarounds did not carry over to the actual computational model. This friction was eventually smoothed out by the introduction of a new IT system, which then allowed for changes of gender entries and the inclusion of a third gender option as a variable. We conceptualized this as a second more impactful moment of tinkering and maintenance: while the first moment was important to make the problem visible, the second one created (again) a seamless integration of the now changed normative account on gender.

Friction in practices related to algorithmic infrastructures creates both moments of crisis and tinkering. Subsequently, we consider posing the following questions as insightful for ethnographic research in data-dense environments:

- *Tiny moments of breakdown*: In which practices do algorithmic systems produce friction? Which (former invisible and taken-for-granted) norms, values, and works become visible?

- *Fixing algorithmic infrastructure*: How and by which relevant actors are such frictions addressed, solved, tinkered with, fixed? Which novel problems arise while and after 'fixing'?

To conclude, we highlight that ethnography is also a way to intervene in the field of computational predictions, as it sensitizes us to how norms and values fold and unfold in algorithmic systems. By raising awareness of the moments of friction we can also become active agents of change in the complex interplay of making issues and their solutions in/visible.

Challenges of tracing and visualizing data journeys – Itzelle Medina Perea

One approach that can be helpful to explore the circulation of data is 'data journeys' (Bates et al, 2016). This methodology offers a way to examine the movement of data across space and time, through different cultures and sites of practice from their initial generation through to reuse in diverse contexts.

Data journeys has proven to be useful in tracing data flows in different contexts (for example, Swist et al, 2019), and the insights gained through this exploration are useful to inform the design of accessible visualizations and descriptions which can contribute to efforts of making data flows more transparent (Bates et al, 2023). However, a number of challenges may arise when implementing this methodology. In this contribution, I focus on the challenges I experienced in tracing and visualizing health data journeys in my empirical exploration of journeys of health data produced in the UK healthcare sector and reused for research purposes conducted between 2018 and 2020.

For the exploration of UK health data journeys, my initial plan was to follow the journeys of patient data flowing to two types of sites: universities and pharmaceutical companies. Therefore, I intended to conduct interviews with key informants at these sites, as well as at data intermediary sites responsible for processing and providing access to patient data for both universities and pharmaceutical companies.

While I successfully recruited key informants at universities, significant challenges arose when attempting to engage with pharmaceutical companies and data intermediary sites. Despite multiple efforts, gaining access to data reuse sites controlled by private-sector actors proved unattainable, mainly due to transparency issues. As a result of these difficulties, the research design was refocused solely on the reuse of patient data for research purposes in universities. This challenge revealed an important feature of the data journeys methodology: its full potential can only be realized when stakeholders involved demonstrate willingness to participate and are transparent. Although access to sites controlled by private-sector actors was not obtained, the process of seeking access provided valuable insights into the characteristics of these

sites and showed interesting points of comparison between these sites and sites controlled by different actors (for example, universities).

Furthermore, this access challenge was useful for identifying black boxes of data practices that require the design and application of additional data collection strategies to grey them out or, in other words, to make them less opaque. This highlighted the importance of making adaptations when using this methodology in similar contexts. For instance, where black boxes are identified, it might be necessary to: (1) attend public events to capture the public discourses of people at these sites and key aspects of their culture; (2) invest significant time to try to build contacts with key informants at sites where black boxes are identified, to try to develop an understanding of what is happening at these sites over time; (3) draw on alternative sources to collect data, for example journalistic reports or freedom of information requests.

I also encountered a number of challenges when creating visual representations of data journeys. Through this process, I discovered that visuals are useful to tell the story about the path data follow as they travel between different sites of practice, or in other words, how they move from one point to another. In this sense, the visuals created in the context of this research allowed me to communicate some key insights gained through the exploration of health data journeys regarding how health data are processed and reused within diverse sites of practice.

However, they are less effective in depicting other things that were uncovered through the application of the data journeys methodology, such as how sociocultural values and material factors assemble to shape and justify practices of data reuse; and how in combination these sociocultural values and material factors contribute to the generation of the socio-material conditions that shape data flows. For example, despite multiple attempts, I could not effectively depict how the enthusiasm and excitement of researchers for conducting research with patient data combined with the provision of material resources by funders and other key stakeholders has helped data to flow to the hands of university-based researchers.

Creating effective visuals can be particularly challenging for someone that lacks training as a designer, as was my case. Furthermore, deciding what to include in the visualizations was also a complex task. Visuals can help with explaining tricky concepts and processes that take place within systems (Annan-Callcott, 2021) (for example, how data linkage, aggregation, and de-identification are conducted). However, incorporating representation of these processes into the visualizations may result in overwhelming or convoluted visuals, potentially causing confusion for those engaging with them. Hence, I recommend that researchers who aim to create engaging and effective visualizations of data journeys consider the inclusion of a professional designer in their team. This approach has been successfully

adopted by other research teams, as exemplified by the fruitful outcomes of Living with Data (Kennedy et al, 2022).

This contribution focused on the challenges of tracing and visualizing data journeys. These reflections aim to be helpful for others considering the deployment of the data journeys approach in similar contexts.

PART 2: EMPIRICAL CASE STUDIES OF DATA RENDERINGS

Investigating the historical contingencies of a dataset's form – Lindsay Poirier and Quinn White

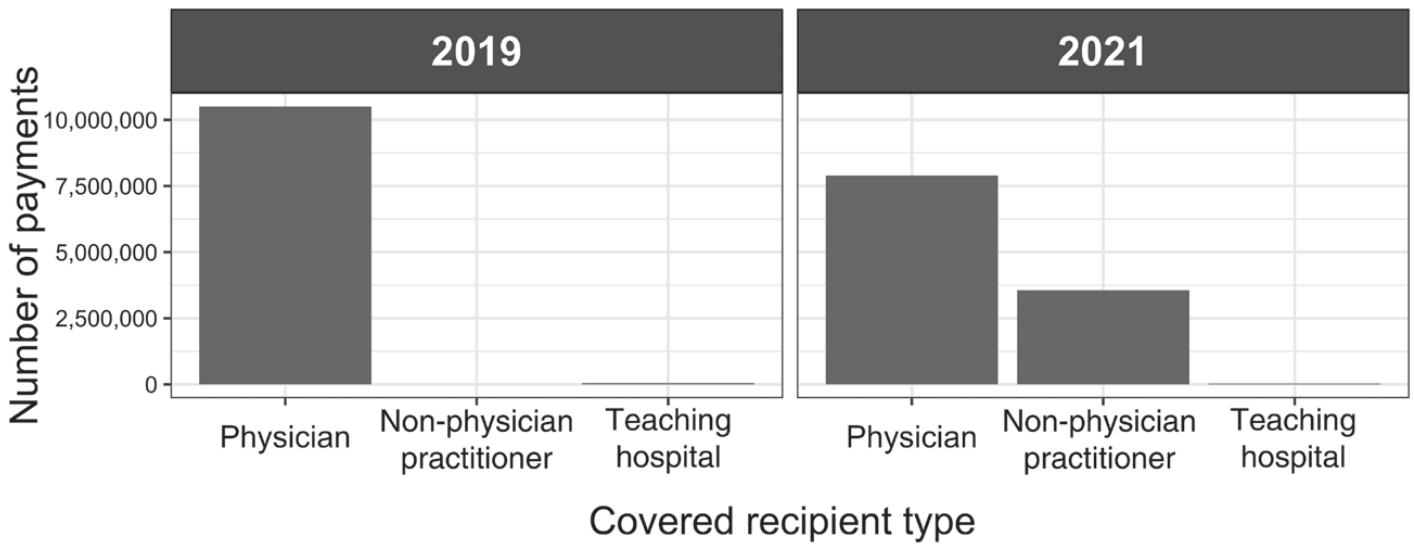
Datasets structure encoded information by dividing observations into rows, delimiting the variables by which data get described, and grouping like variables through systems of classification. These organizing infrastructures make it possible to sort, aggregate, filter, and plot data. How can data studies scholars make sense of combinatorial social forces through which a dataset's form materializes? How can we make sense of why datasets render certain accounts more visible than others?

As a case study, consider a United States public interest dataset known as Open Payments. In 2013 the US Senate passed the Physician Payments Sunshine Act – legislation that required pharmaceutical and medical device companies to report payments to physicians in this public-facing dataset, which is now published annually by the US Center for Medicare and Medicaid Services (CMS). The legislation passed at a time when 'transparency' was gaining traction as a corporate social responsibility framework and marked a widely held conviction that rendering financial relationships visible to the public could highlight medical conflicts of interests, allow patients to make more informed healthcare decisions, and 'disinfect' biases in prescribing. Today anyone can navigate to a CMS website and look-up how much money their US physicians have accepted from medical manufacturers, along with the nature of the payments. Tabular datasets can also be downloaded from CMS's website for analysis and visualization.

While Open Payments shines a light on financial transactions that have historically taken place behind the scenes, the history of political debates that have shaped its material configurations are less easily discernible. Culturally, datasets are often treated as 'given' – as instrumental tools for analysis, rather than as iterating cultural artefacts. As it turns out, deliberative social action and advocacy can be rendered visible in the dataset, but in order to perceive them an analyst must fix a hermeneutic lens on the dataset's structure and values. This involves curating and interpreting an archive of materials that document the sociocultural provenance of the definitions and standards guiding data reporting.

For example, check out the plot shown in Figure 3.1. It documents how many payments were made to each type of provider in the dataset across two

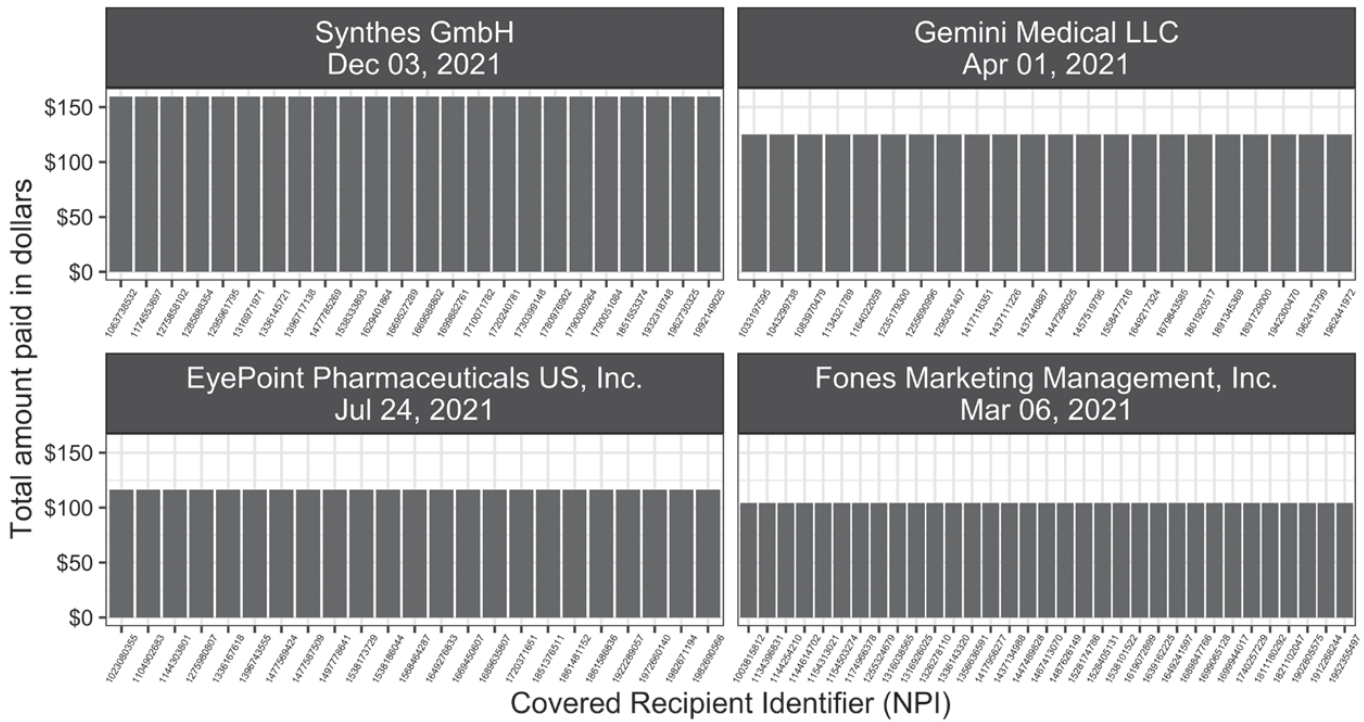
Figure 3.1: Barplot showing the number of payments to each type of covered recipient in Open Payments across two years



years. We see considerable year-to-year changes in the number of payments made to non-physician prescribers. Up until 2020, manufacturers were only required to report payments made to physicians and teaching hospitals, which notably excluded nurse practitioners, physician assistants, and other non-physician prescribers. We studied the provenance of this decision across several sources – interviewing political and agency staff involved in legislating and standardizing data collection, watching archived videos of senate round tables where the scope of reporting was debated, and reading through hundreds of public comments submitted prior to the formulating of the final standards. We learned that the exclusion of these recipients was contentious. Similar state laws required reporting for these categories of prescribers, and many transparency advocates publicly argued that, in excluding these categories, manufacturers might shift money to non-physician prescribers to evade reporting. In interviews, legislative staff noted that federally legislating reporting for physicians was just a first step, and at the time there were not as many concerns about conflicts of interest when it came to non-physician prescribers. Yet, in the mid-2010s, as the opioid crisis swept the US, there were heightened concerns regarding the role of non-physicians in prescribing opioids. Broad legislation aiming to tackle this epidemic included a provision to expand reporting to these categories of providers.

Political debates can also be seen in the individual values reported in the data. One of the most common lamentations we heard in interviews was how much time was spent while preparing the data standards in 'arguing over bagels'. Imagine this scenario. A pharmaceutical representative shows up to a medical practice and provides a spread of bagels. Throughout the day, physicians pop in and out, grabbing a bagel or two from the spread. Should the pharmaceutical company be required to track the names and consumption of every physician that participated? What if the spread was dropped off at an all-day conference where it may be difficult to oversee who ate what? And does a bagel spread even indicate a potential conflict of interest? In public commentary, industry representatives suggested that having to track this information would place undue regulatory burden on companies. While many activists agreed that the purpose of Open Payments was not to track something as mundane as bagel consumption, they also acknowledged that providing meals was a common way industries attempted to curry favour with physicians and thus needed to be reported accurately. When devising the regulations for the dataset, CMS attempted to find middle ground. When determining the value of a meal to assign to a physician, pharmaceutical and medical device companies would be required to report the 'per person' cost of the meal based on the number of individuals that actually partook, regardless of what they ate or drank. We can see the traces of these arguments and advocacy in visualizations of a company's food and beverage payments to providers on a single date (Figure 3.2).

Figure 3.2: Barplot showing examples where companies reporting to Open Payments paid an identical amount to many providers on a single date



While datasets can bring certain social issues into visibility, their austerity can veil the historical contingencies of the information encoded within them. Fixing a hermeneutic lens on a dataset can draw different data narratives into visibility – narratives that highlight the shifting discursive, regulatory, and sociocultural landscapes from which its form materializes.

Deobfuscating state surveillance in Canada – Evan Light and Ellouise McGeachey

In recent years, there have been calls to regulate state surveillance via strategies such as social movement building and moratoriums on certain technologies to buy policy makers time to make policy. We agree that the situation is urgent and demands an urgent response, yet to plough forward without precision would be a fool's errand. We have observed that it is general practice in Canada for state security agencies to use emerging technologies in their work before relevant legal frameworks have been developed. Today, Canadian state security agencies use a wide range of surveillance technologies with little or no oversight, legal guidance, and accountability. Through our research project, *Deobfuscating State Surveillance Capabilities in Canada*, we seek to document the use of surveillance technologies by state security agencies at every level of government. We do so with a combination of tools: access to information/freedom of information requests, proactive disclosure research, and legal research. Additionally, we maintain all of our research results in a public online archive. This section will map out these tools and how we put them to use.

Government procurement research

There are two general approaches to beginning research on government procurement, both of which we utilize. Each takes a lengthy amount of time, attention to detail, and the ability to manage large volumes of documents.

In the first, we are essentially *lurking* in public-sector procurement venues, documenting procurement in real time by creating accounts on the third-party websites used for government procurement. Sites we find popular with Canadian governments are SAP Ariba (2023),[5] MERX (2023),[6] and Biddingo (2023)[7] and this will vary greatly depending on your jurisdiction. In these spaces, government actors or agencies will issue public requests for goods and services, many of which include finely detailed information. For example, one call we encountered sought bidders to retrofit a Canadian federal prison with a new video surveillance system. The accompanying document included blueprints of the prison, precise notation of the areas to be surveilled and precise descriptions of the types of surveillance to be conducted. Ultimately, it is possible to follow this process into a second

wherein correspondence is exchanged between a company and a government actor, a contract is awarded, and work is conducted. Through this process as well as other research, we have developed a living 'Companies of Interest' list, entities whose contracts with governments we believe will help fulfil our goal.

The second method for gathering data is a combination of two processes and is facilitated by open government resources and access-to-information laws. The first (Government of Canada, 2023a) requires the federal government to proactively disclose all contracts of more than $10,000, while the second permits anybody to pay a small sum of money to ask for and receive copies of these contracts, as well as any other government information. For instance, we have uncovered documents showing the Royal Canadian Mounted Police and other federal agencies purchasing forensic management software and cellphone hacking equipment from multiple vendors. In this case, contracts show where equipment has been delivered and in the future we will file further requests to learn precisely how it is used and how its use is governed.

With paltry oversight of policing bodies in Canada, this information will provide us with the information necessary for regulatory and oversight proposals.

Access to information/freedom of information

In Canada, the Access to Information Act[8] facilitates public access to federal government information. One pays $5 to request any number of documents that would be otherwise unpublished. In addition, a live database of previous requests is available and these can be requested for free (Government of Canada, 2023b). Each province and territory in Canada has a Freedom of Information Act which governs how the public can access government information at both provincial and municipal levels. Provinces, territories, and municipalities charge substantial fees both to make requests and fulfil them. When possible, we make requests at the federal level that may provide us with information concerning other levels of government – for instance, email chains, meeting minutes, presentation decks, and briefing notes. Historically a journalistic research technique, practitioners have built their own personal systems of organizing information and their own collections of this information. In an effort to share information and to not duplicate labour, we have created an online archive of our work and will be inviting others to contribute their resources.[9]

Finally, we conduct regular scans of cases within all levels of the Canadian judicial system, seeking cases where surveillance technology has been used in policing. This research enables us to identify, for instance, whether technologies are being used lawfully and if laws and oversight mechanisms

exist to control their use. The use of most surveillance technologies in Canada is lightly regulated or entirely unregulated and it is our hope that our research can help lead to a future where policing at all levels is held to a higher level of respect for one's personal privacy.

PART 3: TOOL-BASED INTERVENTIONS

Data Ethics Decision Aid (DEDA) – Petter Falk and Theo Röhle

Data Ethics Decision Aid (DEDA) is an impact assessment workshop developed by the Utrecht Data School (Franzke et al, 2021). The tool functions as a form of deliberative framework specifically designed for use within public administration data projects and is composed of a worksheet with roughly 40 questions on data ethics, project management and organizational responsibilities. These questions are addressed in project teams and working groups in charge of implementing or facilitating data-centred projects. The process provides a critical perspective that outlines (1) the goals of the organization for data-driven tools, (2) the actual results they produce, and (3) the ethical and political aspects that exist between the two.

As a research method, DEDA has been presented as an applied approach for data studies, drawing on a tradition of action research (Schäfer et al, 2022). Such a perspective can contribute with novel or challenging approaches to empirical investigations (Breit et al, 2019; Loukissas, 2019) as well as potentially heading calls for data studies to move towards a more generative critique (Zakharova, 2021). In seeking to address the ethical dimensions of systems at the design stage, rather than at the implementation stage, it shares an agenda with approaches such as 'reflective design' (Sengers et al, 2005) and 'values in design' (Knobel and Bowker 2011). It also ties in with ongoing discussions in computer science that seek to address algorithmic fairness, accountability, and transparency in systems design (Laufer et al, 2022).

Public administration is an area where the relationship between automated and manual decision-making has special urgency. As Gansky and McDonald (2022: 1989) point out: '"Street-level bureaucrats" are often the primary institutionally-provided mechanism for ensuring the realization of normative goals, like fairness, accountability, and transparency – and their discretion can be critical for handling errors, edge cases, and contextual considerations for rule-based systems.' As a research method specifically geared towards this question, DEDA has the potential to highlight very specific real-life settings where the discretionary power of public officers is about to be replaced. By involving both technical and domain expertise, DEDA workshops open up a space for discussing the consequences of such replacements with regard to the values of the organization and the participants. The workshops also create a temporary halt in the developmental pace of the project – a

kind of artificially induced 'infrastructural inversion' (Bowker and Star, 2000: 34) – that renders aspects visible that otherwise tend to fade into the background of taken-for-granted 'enabling environments' (Peters, 2015: 3) of data infrastructures.

As such, the methodology has the potential to practically challenge notions of data objectivity and address power imbalances inherent in data practices and algorithmic systems, while the functionality of the system is not yet established. Furthermore, there is an opportunity to build on the growing body of empirical material exploring DEDA as an applied critical research methodology as the basis for a larger comparative outlook. Originating in the Netherlands, the framework has been used in over 100 workshop settings in Dutch municipalities and government agencies, starting in 2016. As of 2023, the DEDA method has been applied in a handful of research and development settings across Europe. Researchers at Karlstad University and Gothenburg University have adapted the method for the Swedish context, and research collaborations between the Netherlands, Germany, Sweden, Finland, and Estonia are under way. There are a number of challenges for a comparative project of this scale, but mapping similarities and differences between various local data settings holds a significant potential to gauge the importance of ethical questions in an inductive manner, rather than simply applying a given ethical framework.

Sphere Transgression Watch: reflections on translating a philosophical theory to a digital tool – Marthe Stevens, Tamar Sharon, and Bernard van Gastel

In the past decade, the large tech corporations which we know so well from the world of computational software, hardware, and the Internet, have expanded well beyond their original sphere of activity into new societal domains, or 'spheres'. For example, major tech companies have been developing software and wearables for remote clinical studies (Apple, 2015), making their mark with e-learning platforms which seek to personalize learning and facilitate remote teaching (Google, 2014; Apple, 2018), and assisting farmers with sustainable and regenerative agriculture practices (X.company, 2020).

In order to study this Big Tech expansionism, we developed the Sphere Transgressions theoretical lens, drawing on Michael Walzer's theory of justice (1983). In short, Walzer argues that social life is made up of different spheres that are organized around different conceptions of justice. In a just society, advantages in one sphere – such as wealth or political power – should not translate into advantages in another. Such translations consist in 'sphere transgressions' and ultimately unjust and tyrannical societies. We argue that we are currently witnessing a series of sphere transgressions by tech companies, whereby the (legitimate) advantages they have accrued in the sphere of digital goods, namely digital expertise and know-how, are

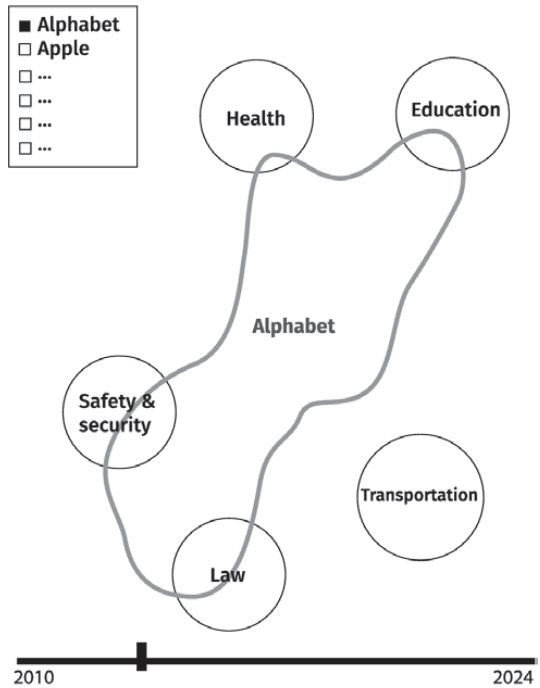
translated into (illegitimate) advantages in other societal spheres (Sharon, 2021a, 2021b; Stevens et al, 2024).

While studying this growing phenomenon, we learned that Big Tech corporations were expanding quickly and in societal spheres well beyond our main area of expertise (health and medicine). Also, we noticed that there was little awareness about this phenomenon in non-academic settings. Therefore, we decided to translate the Sphere Transgressions theoretical lens into visual form. We gathered a team of philosophers, social scientists, computer scientists, and interaction designers all working at Radboud University's interdisciplinary Hub for Digitalization and Society (iHub, the Netherlands). Collectively, we were able to launch the Sphere Transgression Watch (STW), a digital tool that tracks the growing involvement of eight large tech corporations in various societal sectors, in April 2022 (Stevens et al, 2022). It is a public-facing, open data tool, which seeks to both visualize and render more tangible the threat of Big Tech expansionism through society for a broad audience. Users of the tool can scroll through the timeline and see the influence of tech corporations in particular societal spheres growing. They can also click on a particular sphere and gain access to a rich archive of newspaper articles, blogs, and information from company websites that can be used for research purposes.

Translating a theoretical lens into visual form was not straightforward and we had to navigate many questions related to in/visibility and power. One of the first questions that we had to answer was how to chart the contours of something that is 'invisible'. We brainstormed about ways to collect relevant data for the tool as there was no public archive where all Big Tech's initiatives are registered. We realized that we should collect the data ourselves and that the best sources were newspaper articles, blogs, and information from company websites that were used to present, promote, or reflect on Big Tech's innovations. But simply listing these – often very positive – messages could easily be seen as a celebration of Big Tech's innovative capabilities, which was certainly not our intention.

In order to render more tangible the threat of Big Tech expansionism through society, we had to ask ourselves how we should visualize sphere transgressions and what aspects to foreground in the visualization. Figures 3.3 and 3.4 show drawings made during our brainstorming sessions. Figure 3.3 focuses on companies and how they expand to various societal domains. In contrast, Figure 3.4 foregrounds societal spheres and how they are being infiltrated by large tech corporations. In the end, we decided to focus on societal spheres and how the influence in the spheres grows over time (Figure 3.5) as we wanted to highlight the extent of this phenomenon in various societal spheres at the same time. These decisions made us realize how seemingly small design choices had an important impact on the visualization and how it would be perceived by others. For example, in selecting the colours for the website, we opted for

Figure 3.3: Drawing that shows how Big Tech expands to various societal spheres

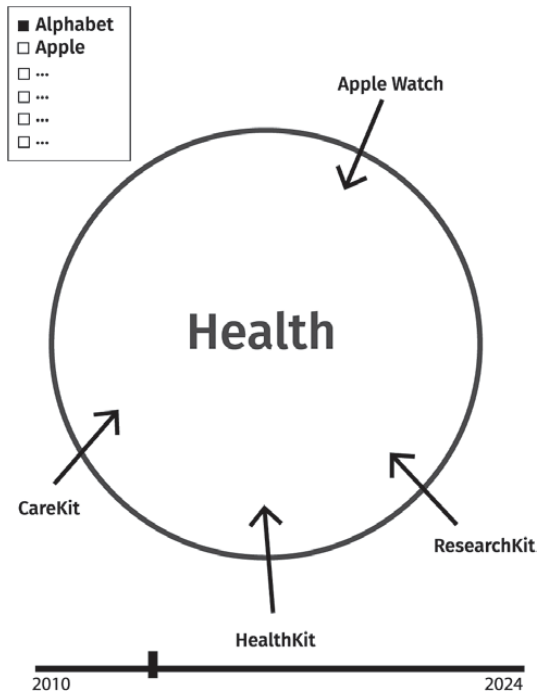


a colour scheme which suggests risks and harms: red, black, and white, rather than more reassuring colours such as green or blue.

There is value in translating a philosophical theory to a digital tool. We noticed that the development of visualizations can be an important way to refine a theoretical lens. Theoretical concepts needed to be made concrete and so needed to be defined early on. For example, the technical experts needed to know how many spheres should be included on the website. Or, could an initiative fall into multiple spheres at the same time? Are certain initiatives more significant than others? Such questions were difficult for the philosophers, who preferred not to see spheres as well-defined entities, but as a heuristic that can be defined on various levels depending on their use. Eventually the philosophers were able to pin-down some concepts for the website, but also argued for a highly flexible tool to which they could make alterations later on. For example, they wanted the possibility to change the names of the spheres, if one turned out to be problematic later on. Looking back, there was value in defining such concepts early on, as this forced the team to think through the consequences of theoretical choices earlier.

In addition, we noticed that the visualizations allowed our theoretical ideas to travel further and reach new audiences. For example, the website is

Figure 3.4: Drawing that shows how societal spheres are infiltrated by Big Tech initiatives

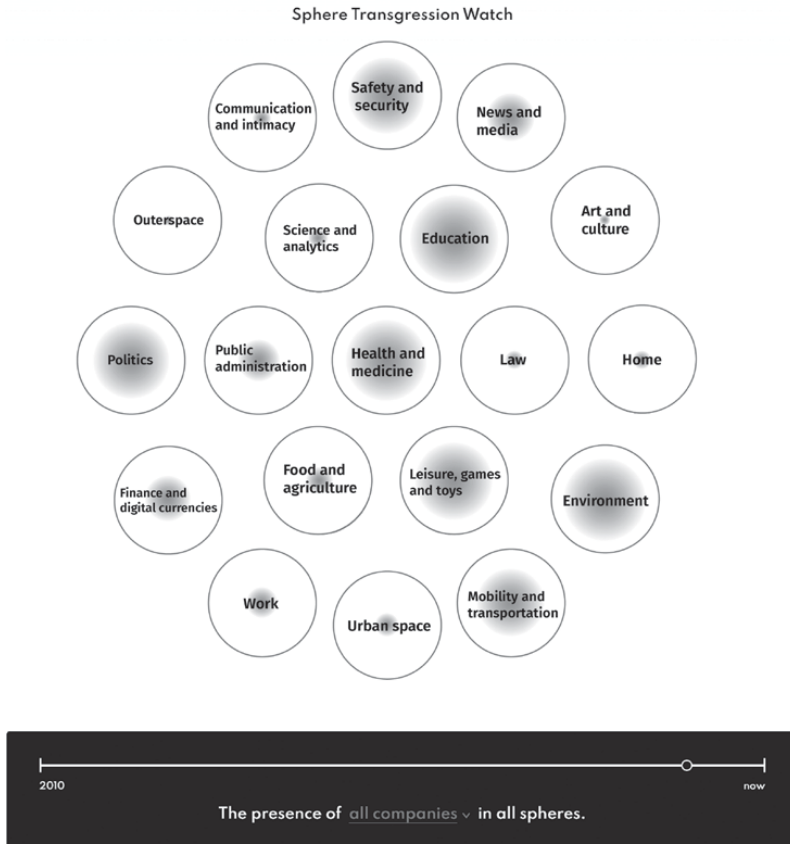


currently being used for research purposes and to bring researchers together for a special issue (Stevens et al, 2024). It is also used in presentations that members of the research group give in non-academic settings, such as policy circles (Dutch Ministry of the Interior) and professional conferences (for example, on AI in health for medical professionals). The tool has been highlighted in Dutch and international media (for example, BNR Radio, De Volkskrant, and the Swiss newspaper *Le Temps*) thereby stimulating timely discussions about Big Tech expansionism.

Conclusion

The contributions of this chapter highlight how the themes of visibility and invisibility offer a central moment of reflection on how we can and ought to study and understand digital data in our research. They highlight the challenges and issues, but also the opportunities that come with questions of positionality, access, transparency, as well as questions of infrastructure and bureaucracy. This enables us to see – and as such shows us – that the kaleidoscope does more than provide different perspectives. While each

Figure 3.5: The final design of the digital tool



of the contributions highlights different aspects of power and different subjects of study in data studies, the contributions are bound together by the kaleidoscope itself.

The kaleidoscope enables multiplicity in a way that enables us to relate these perspectives *to each other*. Similarly, the contributions of this chapter show different perspectives on (methodological) challenges in data studies but are bound together in a common understanding of data studies that does not only accept multiplicity, but ties it together in a shared world that is related and multiple at the same time – or as John Law (2002) argues: fractal. The question then is not only what is being made visible in the different perspectives, but how can these be brought together without denying that they are fractal enactments? In this sense, our choice of the kaleidoscope as metaphor addresses the dilemma that Vogl (2007: 22) ascribes to the historical development of the telescope from instrument to medium: 'every visibility is surrounded by an ocean

of invisibility. … With every deepening of clarity comes a new depth of the unclarifiable.'

Earlier in the chapter, Zakharova argues that what we know about processes of datafication of society relies to a large degree on the methods and approaches that we are using. Thus, what is being made visible (in an always already structured world) relies in a first moment on the modes of engagement that we choose. In an interdisciplinary field like data studies, there is the necessity to become open and transparent to others but also to oneself about what part of the kaleidoscope we are using to look at the complex and multiple phenomenon of datafication.

And this points towards the need to also make ourselves visible in our encounters and our research. This choice in itself is already political, as it results from a theoretical contingency that needs to be translated into a practical singular perspective. What do we observe? What do we measure? And what is being observed and made in/visible in the datasets that we work with? Poirier and White show how a political act for more transparency in Open Payment initiatives is not only qua existence political, but also in the nitty-gritty details of its design. What about the bagels? Should we track how the industry provides food for US physicians? This example demonstrates that activists themselves are not always clear how to respond to that question. The contribution of the kaleidoscope of data studies here is to show that data as such are not sufficient to address the political dimension of datafication, but that we also need to be critical about the nuanced design questions of digital infrastructures that construct, transport, and translate data. This does not only include questions of databases, servers, or network switches, but also refers to the fact that these infrastructures are embedded and part of an (always already existing) organizational structure that runs them, works with them, and makes sense of them.

As Light and McGeachey show, organizational (presentations, meeting notes) and legally mandated practices (public tenders) create traces that can enrich our perspective on the datafication of society. On the same note, these infrastructures and the data practices they are co-shaping can become visible upon breakdown. As Fahimi et al argue, infrastructural breakdowns are important moments of revealing how these organizations work and how data are a part of their operational logic. Moments of breakdown also offer moments of reflection and opportunities for intervention. Falk and Röhle introduce with DEDA a tool-based intervention that provokes moments of infrastructural inversion (Bowker and Star, 2000) – and thus create calculated moments of possible change – and with that adding a facet to the kaleidoscope where not only the perspectives become multiple, but also the very phenomenon that we study. This, of course, also means that we as scholars and activists have even more responsibility: involvement in

translations is also political engagement. Politics here means that it could always be otherwise[10] (Mol, 2002; Latour, 2004). This is reflected in the question that Stevens et al were facing: how do we visualize sphere transgressions to make the threat more tangible and how do we avoid being a proponent of Big Tech? Or in other words: how do we translate our inquiry in a way that makes aspects visible that are abstract and based in a theoretical potentiality? It is also reflected in the challenges experienced by Medina Perea, for example, how to trace and visualize the movement of data when gaining access to data reuse sites controlled by private-sector actors is unattainable, mainly due to transparency issues.

Coming back to the starting metaphor of the kaleidoscope. In our opinion, the field of data studies is the kaleidoscope that enables multiple perspectives, addresses different forms of in/visibilities and thematizes the political dimension of these methodological inquiries. Considering distinct perspectives, as highlighted in this chapter, allows us to gain deeper insights into the social consequences of data. It is thereby clear that the field, as interdisciplinary as it is, by itself constitutes a multiplicity, and that the multiple engagements, translations, and interventions multiply this even more. In these engagements, we add new perspectives, lose others, and thus change the kaleidoscope as such, allowing at any given moment for contingency and fluidity. However, what holds the kaleidoscope together is the insight that our methodological approaches and the in/visibilities they create and tackle at the same time are born out of an understanding that goes beyond a modernist narrative of data as description of a detached reality (see also Bowker, 2006; Gitelman, 2013). As such, the metaphor of the kaleidoscope is the always present reminder that our methodological engagements are situated, reflexive, and political.

Discussant – Jonathan W. Y. Gray

How to study what data makes (in)visible and with what consequences? The sections in this chapter explore methods, approaches, tools, and interventions for accounting for the social lives of data in society – from tracing frictions and fixes, to following journeys and histories, to making freedom of information requests, maps, and visualizations. In being invited to serve as a discussant for this chapter and to surface connections with my own work, I ruminate on resonances with shared research practices for studying data – the details of what is done, what is asked, what is noticed – as well as on how different ways of accounting for data can make a difference.

As alluded to throughout the chapter, feminist science and technology studies scholars emphasize situated and relational approaches – with particular attentiveness to who and what is missing, who and what might

be affected by but absent from the shaping of technoscience. Susan Leigh Star wrote memorably of her dedication to studying relations between 'lived experiences', 'technologies', and 'silences' (Star, 2007). Donna Haraway encourages 'staying with the trouble', 'nurturing capacities to respond', and 'cultivating ways to render each other capable' (Haraway, 2016). Feminist data scholars also emphasize ethics of care (Taylor, 2020; Fotopoulou, 2021), the possibilities and politics of refusing data (Barabas, 2022; Garcia et al, 2022), and affirming epistemic plurality, counter-narratives, and marginalized perspectives (D'Ignazio and Klein, 2020).

These approaches may provide methodological inspiration and orientation in studying data. In a recent project studying open data portals, which may be considered devices for making data public, data request mechanisms provide an indication of datasets which were sought but not found (Gray, 2023).[11] Just as Mimi Ọnụọha's Library of Missing Datasets provides a 'repository of those things that have been excluded in a society where so much is collected',[12] so gathering unsuccessful requests from portals around the world may provide an indication of data considered missing, users left disappointed, questions unanswered, and how data practices, policies, and portals have failed to live up to aspirations and expectations. As well as disclosing details about data infrastructures, data portal interfaces may disclose the failure of arrangements meant to inform. Attending to and articulating trouble has become part of digital culture and society (Meunier et al, 2021), as well as a way to situate and reflect on taken-for-granted digital practices (van Geenen et al, 2023).

How might data studies modify data practices? How can methodological enactments for understanding and situating data (whether through tracing data journeys and frictions, maps, or information requests) make a difference to its role in culture, society? As part of a collaboration between those who do and study data journalism, we identified themes and challenges for 'critical data practices' (Bounegru and Gray, 2021). For example, drawing on a chapter by Helen Verran (2021) on narrating numbers and staying with the trouble of value, we considered how data journalists might tell stories both with and about data – that is, both using data as a medium for exploring and narrating issues, as well as reporting on the making of data.

Reporting on data (in)visibilities may be relevant for both those using and those studying data in society – as illustrated by a 2021 *Washington Post* piece exploring what carbon emissions data suggests about both the emission of carbon as well as the making of data.[13] Such approaches to following data in society may suggest a shift from looking at datasets as resources, to exploring how data infrastructures embody relations between actors, organizations, processes, and methods – and how other kinds of datasets may be made and other arrangements may be composed (Gray et al, 2018).

The more-than-academic relevance of investigating data infrastructures and data (in)visibilities may also surface through collaborations which could be characterized as inventive or experimental (Lury and Wakeford, 2012; Lezaun et al, 2016). As well as critically reconsidering conventional infrastructures and outputs of scholarly communication – from societies to monographs to pay-walled peer reviewed journal articles (Eve and Gray, 2020) – the modification and hybridization of formats may provide generative entrypoints for exploring the broader salience of data studies (de Mourat et al, 2020). Careful attention to how questions are formulated, problems are articulated, methods are enacted, and communities are assembled may enrich collaborative investigations with and about data (Gray et al, 2022). Thinking along with feminist science and technology studies, the kinds of methods, tools, and approaches outlined in this chapter may contribute not only to the study of data in society (including scrutinizing its role in the social production of invisibilities, absence, and marginalization) but also to its unscripting, taking-apart, respecification, repurposing, and recomposition.

Notes

[1] Alphabetical author order
[2] Discussant
[3] Facilitator
[4] Research presented in this section draws on Irina's doctoral thesis which she received from the University of Bremen.
[5] Available at: https://service.ariba.com
[6] Available at: www.merx.com/
[7] Available at: www.biddingo.com
[8] Available at: https://laws-lois.justice.gc.ca/eng/acts/A-1/
[9] Available at: https://surveillance.glendon.yorku.ca/
[10] A great and very fitting variation of this theme – namely: 'IT could be otherwise' – has been developed by the Technologies in Practice Group (TiP) at the IT University Copenhagen.
[11] https://datanotfound.jwyg.org/
[12] https://mimionuoha.com/the-library-of-missing-datasets
[13] https://www.washingtonpost.com/climate-environment/interactive/2021/greenhouse-gas-emissions-pledges-data/

References

Annan-Callcott, G. (2021, 2 September) *What we know about communicating how health data is used | Understanding patient data*. Understanding Patient Data. Available at: http://understandingpatientdata.org.uk/

Apple (2015, 9 March) 'Apple introduces ResearchKit, giving medical researchers the tools to revolutionize medical studies'. *Apple Newsroom*. Available at: www.apple.com/newsroom/2015/03/09Apple-Introduces-ResearchKit-Giving-Medical-Researchers-the-Tools-to-Revolutionize-Medical-Studies/

Apple (2018, 26 June) 'Apple's free Schoolwork app now available for teachers'. *Apple Newsroom*. Available at: www.apple.com/newsroom/2018/06/apples-free-schoolwork-app-now-available-for-teachers/

Barabas, C. (2022) Refusal in data ethics: Re-imagining the code beneath the code of computation in the carceral state. *Engaging Science, Technology, and Society*, 8: 35–57.

Barad, K. (2007) *Meeting the Universe Halfway: Quantum physics and the entanglement of matter and meaning*. Duke University Press.

Bates, J., Kennedy, H., Medina Perea, I., Oman, S., and Pinney, L. (2023) Socially meaningful transparency in data-based systems: Reflections and proposals from practice. *Journal of Documentation* (ahead of print). Available at: https://doi.org/10.1108/JD-01-2023-0006

Bates, J., Lin, Y.-W., and Goodale, P. (2016) Data journeys: Capturing the socio-material constitution of data objects and flows. *Big Data & Society*, 3(2). Available at: https://doi.org/10.1177/2053951716654502

Bhaskar, R. (2008) *A Realist Theory of Science*. Routledge.

Bounegru, L. and Gray, J. (eds). (2021) *The Data Journalism Handbook: Towards a critical data practice*. Amsterdam University Press.

Bowker, G. C. (2006) *Memory Practices in the Sciences*. MIT Press.

Bowker, G. C. and Star, S. L. (2000) *Sorting Things Out: Classification and its consequences*. MIT Press.

Breit, E., Egeland, C., and Løberg, I. B. (2019) 'Cyborg Bureaucracy: Frontline Work in Digitalized Labor and Welfare Services'. In: J. S. Pedersen and A. Wilkinson (eds) *Big Data*. Edward Elgar Publishing, pp 149–169.

Burrell, J. (2016) How the machine 'thinks': Understanding opacity in machine learning algorithms. *Big Data & Society*, 3(1). Available at: https://doi.org/10.1177/2053951715622512

Christin, A. (2020) The ethnographer and the algorithm: Beyond the black box. *Theory and Society*, 49(5): 897–918. Available at: https://doi.org/10.1007/s11186-020-09411-3

D'Ignazio, C. and Klein, L. (2020) *Data Feminism*. MIT Press.

de Mourat, R., Ricci, D., and Latour, B. (2020) 'How Does a Format Make a Public?' In M. P. Eve and J. Gray (eds) *Reassembling Scholarly Communications: Histories, infrastructures, and global politics of open access*. MIT Press.

Dencik, L., Hintz, A., Redden, J., and Treré, E. (2022) *Data Justice*. Sage.

Eve, M. P. and Gray, J. (eds) (2020) *Reassembling Scholarly Communications: Histories, infrastructures, and global politics of open access*. MIT Press.

Fotopoulou, A. (2021) Conceptualising critical data literacies for civil society organisations: Agency, care, and social responsibility. *Information, Communication & Society*, 24: 1640–1657.

Franzke, A. S., Muis, I., and Schäfer, M. T. (2021) Data Ethics Decision Aid (DEDA): A dialogical framework for ethical inquiry of AI and data projects in the Netherlands. *Ethics and Information Technology*, 23(3): 551–567.

Gansky, B. and McDonald, S. (2022) 'CounterFAccTual: How FAccT Undermines Its Organizing Principles', *2022 ACM Conference on Fairness, Accountability, and Transparency*, pp 1982–1992.

Garcia, P., Sutherland, T., Salehi, N., Cifor, M., and Singh, A. (2022) No! Re-imagining data practices through the lens of critical refusal. *Proceedings of the ACM on Human–Computer Interaction*, 6: 315:1–315:20.

Gitelman, L. (2013) *"Raw Data" Is an Oxymoron*. MIT Press.

Google (2014, 6 May) 'Previewing a new Classroom'. *Google blog*. Available at: https://blog.google/outreach-initiatives/education/previewing-new-classroom/

Government of Canada (2023a) *Proactive Disclosure*. Available at: https://open.canada.ca/en/proactive-disclosure

Government of Canada (2023b) *Completed Access to Information Requests*. Available at: https://open.canada.ca/en/search/ati

Graham, S. and Thrift, N. (2007) Out of order: Understanding repair and maintenance. *Theory, Culture & Society*, 24(3): 1–25. Available at: https://doi.org/10.1177/0263276407075954

Gray, J. W. Y. (2023) What do data portals do? Tracing the politics of online devices for making data public. *Data & Policy*, 5, e10.

Gray, J., Gerlitz, C., and Bounegru, L. (2018) Data infrastructure literacy. *Big Data & Society*, 5: 1–13.

Gray, J., Bounegru, L., Rogers, R., Venturini, T., Ricci, D., Meunier, A., et al (2022) Engaged research-led teaching: Composing collective inquiry with digital methods and data. *Digital Culture & Education*, 14: 3.

Haraway, D. J. (2016) *Staying with the Trouble: Making kin in the Chthulucene*. Duke University Press Books.

Katzenbach, C. (2021) 'AI will fix this': The technical, discursive, and political turn to AI in governing communication. *Big Data & Society*, 8(2). Available at: https://doi.org/10.1177/20539517211046182

Kennedy, H., Oman, S., Ditchfield, H., Taylor, M., Bates, J., Medina Perea, I., et al (2022) *Data Matters Are Human Matters: Final Living With Data report on public perceptions of public sector data uses*. The University of Sheffield. Available at: https://livingwithdata.org/project/wp-content/uploads/2022/10/LivingWithData-end-of-project-report-24Oct2022.pdf

Kitchin, R. (2017) Thinking critically about and researching algorithms. *Information, Communication & Society*, 20(1): 14–29. Available at: https://doi.org/10.1080/1369118X.2016.1154087

Knobel, C. and Bowker, G. C. (2011) Values in design. *Communications of the ACM*, 54(7): 26–28. Available at: https://doi.org/10.1145/1965724.1965735

Latour, B. (2004) Whose cosmos, which cosmopolitics? Comments on the peace terms of Ulrich Beck. *Common Knowledge*, 10(3): 450–462.

Laufer, B., Jain, S., Cooper, A. F., Kleinberg, J., and Heidari, H. (2022) 'Four Years of FAccT: A Reflexive, Mixed-Methods Analysis of Research Contributions, Shortcomings, and Future Prospects', *2022 ACM Conference on Fairness, Accountability, and Transparency*, pp 401–426.

Law, J. (2002) *Aircraft Stories: Decentering the object in technoscience* (1st edn). Duke University Press.

Law, J. (2004) *After Method: Mess in social science research*. Routledge.

Law, J., Ruppert, E., and Savage, M. (2011) *The double social life of methods*. Centre for Research on Socio-Cultural Change (CRESC), Working Paper No 95.

Lezaun, J., Marres, N., and Tironi, M. (2016) 'Experiments in Participation'. In U. Felt, R. Fouche, C. A. Miller, and L. Smitt-Doer (eds) *Handbook of Science and Technology Studies* (4th edn). MIT Press, pp 195–222.

Loukissas, Y. A. (2019) *All Data Are Local: Thinking critically in a data-driven society*. MIT Press.

Lury, C. and Wakeford, N. (eds). (2012) *Inventive Methods: The happening of the social*. Routledge.

Meunier, A., Gray, J., and Ricci, D. (2021) 'A New AI Lexicon: Algorithm Trouble: Troublesome Encounters with Algorithms that Go Beyond Computational Processes'. In *A New AI Lexicon*. AI Now Institute.

Mol, A. (2002) *The Body Multiple: Ontology in medical practice*. Duke University Press.

Peters, J. D. (2015) *The Marvelous Clouds: Toward a philosophy of elemental media*. University of Chicago Press.

Powell, A. B., Ustek-Spilda, F., Lehuedé, S., and Shklovski, I. (2022) Addressing ethical gaps in 'Technology for Good': Foregrounding care and capabilities. *Big Data & Society*, 9(2): 20539517221113776. Available at: https://doi.org/10.1177/20539517221113774

Ruppert, E., Law, J., and Savage, M. (2013) Reassembling social science methods: The challenge of digital devices. *Theory, Culture & Society*, 30(4): 22–46. Available at: https://doi.org/10.1177/0263276413484941

Schäfer, M., Falk, P., Muis, I., and Röhle, T. (2022) 'Analysing Data Power through Participatory Observation: A Call to Action', Data Power 2022 Conference.

Seaver, N. (2017) Algorithms as culture: Some tactics for the ethnography of algorithmic systems. *Big Data & Society*, 4(2): 2053951717738104. Available at: https://doi.org/10.1177/2053951717738104

Sengers, P., Boehner, K., David, S., and Kaye, J. (2005) 'Reflective Design', *Proceedings of the 4th Decennial Conference on Critical Computing: Between Sense and Sensibility*, pp 49–58.

Sharon, T. (2021a) Blind-sided by privacy? Digital contact tracing, the Apple/Google API and big tech's newfound role as global health policy makers. *Ethics and Information Technology*, 23: 45–57. Available at: https://doi.org/10.1007/s10676-020-09547-x.

Sharon, T. (2021b) From hostile worlds to multiple spheres: Towards a normative pragmatics of justice for the Googlization of health. *Medicine, Health Care & Philosophy*, 24: 315–327. Available at: https://doi.org/10.1007/s11019-021-10006-7.

Star, S. L. (1999) The ethnography of infrastructure. *American Behavioral Scientist*, 43(3): 377–391. Available at: https://doi.org/10.1177/00027649921955326

Star, S. L. (2007) 'Five Questions'. In J. K. B. Olsen and E. Selinger (eds) *Philosophy of Technology*. Automatic Press, pp 227–233.

Star, S. L. and Ruhleder, K. (1996) Steps toward an ecology of infrastructure: Design and access for large information spaces. *Information Systems Research*, 7(1): 111–134. Available at: https://doi.org/10.1287/isre.7.1.111

Stevens, M., Kraaijveld, S. R., and Sharon, T. (2024) Sphere transgressions: Reflecting on the risks of big tech expansionism. *Information Communication & Society*, 1–14. Available at: https://doi.org/10.1080/1369118X.2024.2353782

Stevens, M., Sharon, T., van Gastel, B., Hoffman, A. S., Kraaijeveld, S. R., and Siffels, L. (2022) *Sphere Transgression Watch*. Distributed by iHub, Available at: www.sphere-transgression-watch.org

Swist, T., Collin, P., and Third, A. (2019) Children's data journeys beyond the 'supply chain': Co-curating flows and frictions. *Media International Australia*. https://doi.org/10.1177/1329878X19828390

Taylor, L. (2020) The price of certainty: How the politics of pandemic data demand an ethics of care. *Big Data & Society*, 7: 2053951720942539.

van Geenen, D., van Es, K., and Gray, J. W. Y. (2023) Pluralising critical technical practice. *Convergence*, 30(1): 7–28.

Verran, H. (2021) 'Narrating a Number and Staying With the Trouble of Value'. In L. Bounegru and J. Gray (eds) *The Data Journalism Handbook*. Amsterdam University Press, pp 55–64.

Vogl, J. (2007) Becoming-media: Galileo's telescope. *Grey Room*, 29: 14–25.

Walzer, M. (1983) *Spheres of Justice: A defense of pluralism and equality*. Basic Books.

X.company (2020, 12 October) 'Mineral: Bringing the era of computational agriculture to life'. *X.company blog*. Available at: https://blog.x.company/mineral-bringing-the-era-of-computational-agriculture-to-life-427bca6bd56a

Zakharova, I. (2021) Data inquiry: Methodological considerations on datafication in social research. *AoIR Selected Papers of Internet Research*. https:// doi.org/10.5210/spir.v2021i0.12082

Zakharova, I. (2022) *Understanding data studies: A conceptual and methodological inquiry into research on datafication*. PhD thesis, Universität Bremen. Available at: https://doi.org/10.26092/elib/1675

Ziewitz, M. (2016) Governing algorithms: Myth, mess, and methods. *Science, Technology, & Human Values*, 41(1): 3–16. Available at: https://doi.org/10.1177/0162243915608948