

INTRODUCTION: DATA POLITICS

Copenhagen, October 2018: “I would prefer a statistician to a doctor anytime.” These words are spoken quietly by a person sitting near me, a healthy-looking man in his early twenties. I notice the wearable devices on his wrist: he seems to be collecting data on his own health. I’m attending a meeting framed as “an opportunity to discuss the prospect of data-driven healthcare.” It is not the first time I hear someone at meetings, seminars, and conferences on this topic say that they would prefer a statistician to a doctor—about which I cannot help but feel puzzled, not least because of my profound respect for both doctors *and* statisticians. In my experience, practitioners of both professions tend to appreciate each other’s skills, and yet they rarely wish to replace each other. What kind of future healthcare is in the minds of people who prefer a statistician over a medical doctor? What do they think they can achieve with data?

It was around 2014 that I began regularly attending such events about data and the future of medicine. By now, I have become accustomed to people telling me that we are at the cusp, or even in the middle, of a revolution. Informally during breaks, or formally as part of silver-tongued presentations, they foretell an imminent state where ubiquitous computing means that practically any activity will give rise to data, it will be possible to gather enough data on individuals to predict the advent of most diseases, and such predictions will facilitate the prevention of disease. Instead of waiting for symptoms to manifest themselves, people will be warned of future disease by data-profiling devices. Personalized regimes will replace standard treatments. Big data will replace evidence-based medicine (EBM). AI—artificial

intelligence—will replace human interpretation. Apps will sideline doctors on many issues. As a consequence, the story goes, healthcare will be cheaper, more efficient, and people will live longer and healthier lives.¹ On top of this, old-fashioned administration and governance will be cast aside, to be replaced by smart algorithms designed to optimize performance. In these prophecies, data just need a bit of statistical help to do the job.

Many of the people who attend seminars on the prospects of data-driven healthcare have experienced firsthand how data-intensive technology can bring about significant changes in their own lives or their professional practice. For them, this feels like more than a prophecy. Now they want to see data used to optimize healthcare at large. They have observed how data-intensive companies have disrupted several business areas. Innovative technologies have changed how they search for information, and data-driven social platforms have made people connect socially in new ways (van Dijck, Poell, and De Wall 2018). Why should not healthcare be next? Companies like Google, Apple, and IBM are moving into health research. They are using big data methods to build decision support tools, and they are aiming for new ways of generating medical evidence (Prainsack 2017; Sharon 2016). Other big tech companies such as Amazon are experimenting with data-intensive ways of providing healthcare and dispensing drugs (Lewis 2016; Wakefield 2017; Shah 2020; Son 2021).

However, in the business area of health technology, big news quickly turns into old news. Promises fail to materialize, and business plans change. Apparently, there is something about healthcare in high-income countries that is remarkably resistant to change. Still, high-ranking civil servants, influential politicians, and chief executive officers (CEOs) continue to discuss data-driven healthcare as not only desirable, but inevitable. Data have moved to center stage of healthcare politics. What do these data promises entail? What is happening within healthcare? How does the pursuit of data-driven healthcare affect people on the ground: patients, clinicians, medical researchers, administrators, policymakers, and people in the pharmaceutical and medical device industries?

With this book, I explore what is driving the data surge and how it affects healthcare. The high-flying promises inspire concrete policymaking. Policy-making interacts with healthcare practices, giving rise to new experiences among patients and staff. How do these policies, practices, and experiences relate? Promises do not trickle through layers of practice and permeate

everyday experience in the way that water soaks through cloth. Promises interact with practices and experiences—and practices and experiences bounce back. In my venture to understand the drivers for, and implications of, intensified data sourcing in healthcare, I take Denmark as my primary case. It is because Denmark is in many ways extreme. It is a country with thoroughly digitized health services, pervasive data sourcing, highly integrated data infrastructures, and personal identity numbers that make it possible to track citizens across sectors and throughout and beyond their individual lifespans. It is a country eager to be at the forefront of the prophesized data revolution. Denmark is the perfect place to explore the politics of intensified data sourcing.

DATA PROPHECIES: GLOBAL PROMISES AND LOCAL PRACTICES

The buoyant promises of digital disruption that I have outlined thus far are part of a more general societal change. In the whirlwinds of the twenty-first century, data promises have come to pervade policy environments in all corners of the world—not just in healthcare, but also in education, social service, and transportation. Across these sectors, data are referred to as gold mines or the new oil. First concentrated in relatively small circles in Silicon Valley on the American West Coast and around places like the Massachusetts Institute of Technology (MIT) on the East Coast, ideas about how the world *could* be governed based on computational calculation are rapidly turning into ideals for how the world *should* be governed. These ideas, however, are not simply travelling to other parts of the world: they are reinvented, entrenched, and made powerful by interacting with local histories, opportunities, and desires.

Though apparently attuned to positivist ideas of “evidence,” the ideals of a data-driven society also exert a form of spiritual appeal. Informatics scholar Morgan Ames describes Silicon Valley ideals as exerting a *charismatic* authority of the type that Max Weber spoke about in relation to religious leaders (Ames 2019; Weber 2003a). Historian Yuval Harari literally designates the type of “dataism” that these ideals promote as a *religion* (Harari 2018). Social scientists such as Susi Geiger and Richard Tutton describe the expectations of data in big tech circles as a *mythology* operating at the intersection of capitalist extraction and promissory gospel (Geiger 2020; Tutton 2017). Some computer scientists genuinely discuss a messianic future state

of “singularity,” wherein human life is transposed into an eternal state of information and where machines and humans are one being (Shanahan 2015). For people believing in singularity, this concept implies a transcendence of death. If everything is data, the reasoning goes, then everything can live forever as just that: data. Influential voices in the movement promoting data have even founded a company called the Singularity University, where policymakers from all over the world take courses and engage with their thinking (Bernsen 2019).

Expectations as to data and digital technology are intertwined;² the one shapes the other. Elements of the buoyant belief in the power of data can be traced to a hopeful counterculture that emerged with the Internet in the 1980s. It was a culture where people experienced the informatization of social interaction as a liberation. The early users of the Internet spoke about access to data as empowering. In 1995, the computer scientist Nicholas Negroponte announced the coming of an era where bits would take over from atoms (Negroponte 1995). Humanity was to be freed from its material prison. What would be important would be information. In 1996, John Perry Barlow from the Electronic Frontier Foundation even presented a “Declaration of the Independence of Cyberspace,” in which he said about state power: “You have no moral right to rule us nor do you possess any methods of enforcement we have true reason to fear” (Barlow 1996: 2). For people like Barlow, the Internet represented a sphere beyond state power, despite the fact that the American military held a key role in building it.

In the 1990s, Negroponte had begun serving as a government advisor, along with other tech-savvy geeks. Those in government circles were attracted to their hopeful forecasts and saw in their vision something other than a reason to abandon state power: it was to be reinvented as a seamless, supportive force. Ideas about a digital makeover of public service began to sizzle in policy environments all around the globe. Today, people with very different levels of authority—from low-ranking information technology (IT) specialists to high-ranking civil servants, from well-paid consultants in companies like Gartner, McKinsey, or Deloitte to incredibly wealthy business owners such as Tesla founder Elon Musk—have become voices in a choir singing about a data-driven future. In this future, everything is information and information is everything. Each through their own path of discovery, people in very different positions have come to see data as a solution even when dealing

with very different problems. How, then, do data prophecies perform on the ground? How do these prophecies affect contemporary politics?

Data are high on the health policy agenda everywhere, not just in Denmark but also in the European Union, the United Kingdom, the United States, and beyond. China is developing an integrated data infrastructure in the form of a social credit system (Lengen 2017; Liang, Das, Kostyuk, and Hussain 2018), and India has launched digital smart cards connecting citizens to colossal biometric databases (Nair 2021; Dahdah and Mishra 2020). To facilitate data (re-) uses, most policymakers focus on investments in *data integration*. The money goes to infrastructure building. In its *Data Strategy* document from February 2020, the European Commission identifies an ongoing digital transformation of the European economy and society “affecting all sectors of activity and the daily lives of all Europeans. Data is at the centre of this transformation and more is to come” (European Commission 2020b: 1). In the same document, the Commission suggests investing in data infrastructures. Then, just after its publication, the COVID-19 pandemic struck, and the sense of crisis gave rise to a record-high stimulus plan emphasizing investments in what is green and digital. The ideas from the *Data Strategy* could now be backed with significant funding (European Commission 2020c): as much as 20 percent of the greater than 700 billion euros for this stimulus plan will be used to support digital integration. By November 2020, the Commission had proposed a Data Governance Act (Vestager 2020) setting up a legal framework for data sharing in eight sectors, including health. Specifically, the plans in the European Union of building a European Health Data Space, where all types of data (including medical records, genetic data, and patient-generated data) can be stored, exchanged, and reused for governmental planning, research, and cross-border care (European Commission 2020a). These ambitions reflect a form of geostrategic rivalry in which the European Union seeks to use its existing health data resources to gain territory in the global market for data analysis, which is currently dominated by American and Chinese companies.

The political attention to data meant massive investments even before the pandemic. Again, most investments have gone into digital infrastructures. The EU Data Strategy involves ensuring the “interoperability of health data through the application of the Electronic Health Record Exchange Format” (European Commission 2020b: 30)—an ambition that was repeated in the pandemic recovery plan and is to inform the European Health Data Space.

This is not the first time that a crisis has led to a call for integration of health data infrastructures. In the United States, President Barack Obama sought to combat the financial crisis of 2007–2008 by investing in the interoperability of electronic health records, among other things (Wachter 2017). The investments went into the development of what was described in the Act—known as the stimulus package—as a “Health information technology architecture that will support the nationwide electronic exchange and use of health information in a secure, private, and accurate manner” (US Congress 2009: 132). A budget of around 30 billion US dollars was allocated to health data integration. With its support of personalized medicine, the Obama administration also worked to ensure “secure access to the electronic healthcare information of more than 125 million patients” (US Food and Drug Administration 2013: 40).

In 2015, the Council of the European Union similarly adopted a strategy paper on personalized medicine according to which member-states should work to “support the standardisation and networking of biobanks to combine and share resources” and “promote the interoperability of electronic health records to facilitate their use for public health and research” (General Secretariat of the Council: Working Party on Public Health 2015: 7). In January 2019, the European Union invested an additional 21 million euros in the construction of an interoperable electronic health record through the Horizon 2020 program (<https://www.smart4health.eu>, see also Felt, Öchsner, and Rae 2020), and in its 2020 data strategy paper, “the development of national electronic health records (EHRs) and interoperability of health data through the application of the Electronic Health Record Exchange Format” was again a priority area receiving an additional two billion euros (European Commission 2020b: 30). As these examples illustrate, the integration of health data has been subject to investment for some time now. Yet apparently it is not as easy and smooth to ensure data flows as the data gospel suggests. In fact, it seems to be extremely difficult to make data do what the optimistic prophecies otherwise suggest is “inevitable.”

Nevertheless, low-income countries are following suit to ensure good data flows. A highly influential United Nations (UN) report, *A World That Counts*, states that “[i]mproving data is a development agenda in its own right” (The United Nations Secretary-General’s Independent Expert Advisory Group on a Data Revolution for Sustainable Development 2014: 3). The report even articulates a new vision for global public health: “Never

again should it be possible to say, ‘We didn’t know.’ No one should be invisible. This is the world we want—a world that counts” (3). Along with their multilateral investments, nongovernmental organizations (NGOs) such as the Bill and Melinda Gates Foundation donate significant amounts to create data infrastructures. Data almost seem to comprise a new form of, or an alternative to, development aid. Dreams of data ubiquity have captured the imagination of policymakers even in countries still fighting to fulfill basic health needs (Adams 2016b; Erikson 2016; Hoeyer, Bauer, and Pickersgill 2019). This is, in the most literal sense, a world that *counts*—a world that does its counting before it acts (Jensen and Winthereik 2013).

Policymakers clearly *aim* for data ubiquity all around the world. Therefore, there is a need to know more about what integrated data infrastructures can imply in practice. Where can you go if you wish to observe such an integrated data infrastructure? A typical choice for policymakers is a study trip to Denmark. This tiny EU country of just 5.8 million inhabitants receives numerous delegations from North America, Europe, Asia, and Australia. They visit the offices of a platform called *sundhed.dk*, which gives all citizens online access to their electronic health records; they visit an organization called MedCom, known for setting standards for data exchanges between individual health suppliers; or they go to the National Biobank, containing blood samples on most citizens—samples that can be linked to healthcare, social, and educational data. The biobank alone has in recent years received delegations from all corners of the world, including 156 delegations from Japan and 59 from the United States (Tupasela 2021b).

Why do they visit Denmark? At the turn of the millennium, Iceland, another Nordic country, was seen internationally as the epicenter for genomic research. A national biobank and a genealogical database were established, and a company was commissioned not only to run the two databases, but also to build a national electronic health record system and combine all three for research purposes. The company attracted investments on a scale affecting the overall gross domestic product of Iceland (GDP). However, it also created much controversy (Almarsdóttir, Traulsen, and Björnsdóttir 2004; Árnason and Simpson 2003; Fortun 2008; Pálsson and Rabinow 2001; Potts 2002; Sigurdsson 2001; Thorgeirsdóttir 2004). As a result, the company did not complete the electronic medical record system.

Today, Denmark is one of the most fiercely digitized countries with one of the most integrated health data infrastructures in the world (Aanestad

and Jensen 2011). For decades, Denmark has been competing with the other Nordic countries to be the most attractive place for medical research (Tupasela 2017b; Tupasela, Snell, and Tarkkala 2020; Tupasela 2021b). In Denmark, there are tissue samples, medical records, registries, and quality databases available on practically all citizens, and they all can be used for research (Bauer 2014). Most of them even with informed consent exemptions. When the Obama administration began investing in the digitization of healthcare records, Denmark had already long been using digital health records and established centralized national databases. In Denmark, all pharmacies receive prescriptions electronically, and drug data are stored on individual profiles that can be accessed by patients and health professionals regardless of institution or place. Similarly, referrals to specialized care are placed on central servers where hospital wards or specialists fetch them. Most Danes take this level of integration for granted and can hardly remember how it was to fiddle around with a paper prescription or referral. This is why delegations go to Denmark to see what integrated data infrastructures look like in practice. Denmark has become a key site for studying health data integration.

Data are used for ever more purposes, and not just in Denmark. Besides treatment and research, they are used for monitoring clinical quality, for achieving administrative objectives, and for facilitating remuneration. Each of these purposes draw upon increasingly complex algorithms depending on multiple data sources. In the past, an ostensibly simple activity, such as remuneration, operated in relatively uncomplicated data loops (combinations of diagnostic and treatment codes), but they are now increasingly qualified with data sources meant to represent treatment *outcomes*, such as data on employment, sick leave, educational status, patient satisfaction, patient-rated outcomes, and so on (Hogle 2019). The data landscape has become a site where multiple goals, all dependent on more data, converge. To reach these many goals, the need for speed has come to the fore: researchers, clinicians, administrators, and industry all want what they term “real-time” data. When prophecies hit the ground, they hit multiple grounds, as it were. Each data promise may sound feasible, but in practice, each one interacts with many other uses of data. The many purposes generate friction (Edwards 2010; Pellegrino and Mongili 2014; Edwards, Mayernik, Batcheller et al. 2011).

In the chapters that follow, I show what happens when the data promises from American big tech and the circles around the Singularity University

interact with a local healthcare system, which is in some ways far more advanced in terms of digitalization and data integration than the fragmented American healthcare system. I believe that experiences from the Danish experiment in health data infrastructuring can illuminate some of the challenges faced by other healthcare systems. Still, all healthcare is local. Unique local circumstances find ways of bouncing back. My study of Danish data infrastructures is meant to help raise awareness of the importance of locality. My hope is that this awareness can help both policymakers and researchers to reconnect policy, practice, and experience. They need to explore the specificity of their own cases and exert judgment instead of importing standard answers.

THE POLITICS OF INTENSIFIED DATA SOURCING: A BOOK ABOUT PARADOXES

In this book, I will put aside assumptions about the disruptive effect of big data and instead explore who wants which data and what they use them for. I take as my point of departure the basic empirical observation that more actors want more data, of better quality and about more people—while the actors often disagree about who should be allowed to use those data, and for what purposes. I call this phenomenon *intensified data sourcing*. My interest revolves around the *data politics* characterizing this development. Again, my guiding curiosity is: What drives intensified data sourcing in healthcare, and what are the implications for the governance of healthcare, for health professionals and for patients?

I have used this term, “intensified data sourcing,” several times now, and I should explain why. It is, I admit, less seductive than its more popular cousin, “big data.” Big data has become, effectively, a buzzword (Vincent 2014). It is typically associated with high-tech solutions, AI, and organizational “disruption.” Sometimes it is defined as a methodological shift to work with velocity, variety, and volume (the three Vs), but in policy discourse, it remains vaguely defined (Boyd and Crawford 2012). The high-tech association turns the gaze toward future applications and potential impacts that have not necessarily materialized yet. I wish to capture the interplay between policy, practice, and experience and therefore do not wish to be constrained by the preconceptions bound up in the term “big data.” I thus focus on intensified data sourcing to explore both blinkering

high-tech solutions and the type of low-tech manual data collection that the push toward becoming data-driven also, in some areas even predominantly, has set in motion. Intensified data sourcing is a way of naming that which can be observed empirically in health services: a range of people (from clinicians and researchers to politicians and industry representatives) wanting more data, of better quality and on more people—while disagreeing on how these data should be used. I prefer “data sourcing” as an alternative to more established terms such as “data mining” because the mining metaphor is entirely inadequate to capture the *type* of data work involved. Data are not dug out of the ground—they are made. Intensified data sourcing includes the dynamic processes of creating, collecting, curating, and storing data, while simultaneously making them available for multiple purposes.

When I explore the interplay between policy, practice, and experience in intensified data sourcing, it is because it has important consequences for the options people have for pursuing health. Health is a basic condition for human beings in their pursuit of whatever matters to them. Illness causes suffering, pain, and sorrow. How societies respond to such agony is important. Even measures of prevention can become sources of grievance. It is important to understand how intensified data sourcing affect these very intimate aspects of people’s lives. I focus on the data sourced through health-care services because the relations people establish when seeking help from a health professional typically relate to something of outmost importance to them—something that they would not necessarily share with just anybody. Healthcare data matter to people.

As I have become interested in the stories we tell—and those that we ought to tell—about the consequences of intensified data sourcing in healthcare, I have found that the drivers are manifold and the implications *paradoxical*. I have noticed that patients, clinicians, and administrators sometimes use similar data to tell almost contrasting stories. Even more important to my interests in the drivers for and implications of intensified data sourcing is the fact that my interlocutors—the wide range of people I have met and interviewed—often tell almost opposing stories about why data are needed and the effects they have. As a consequence of this experience, this book has come to revolve around paradoxes. For some reason, I had not anticipated this. The opposing stories were causing me many a headache: I could not make the stories align. Gradually, I realized that this was a key point.

Paradoxes are classic figures of Western thought. Philosophy has grappled with paradoxes for thousands of years, such as pondering the truth value of statements such as “I am a liar” (Schad 2016). More recently, technological developments, and digitalization in particular, have been seen by some scholars as generating paradoxes. The philosopher Ezio Di Nucci talks about “the control paradox,” where AI and other technologies simultaneously provide enhanced control and give rise to loss of control (Di Nucci 2021). The STS scholar Judy Wajcman makes similar points about digital time management and observes further how these technologies save time—or are supposed to save time—yet also generate new sources of time pressure (Wajcman 2015, 2019).

The Greek etymology of the term is *para*, meaning “contrary to,” and *doxa*, meaning “opinion.” According to Merriam-Webster, the term “paradox” can mean any of several things, including “one (such as a person, situation, or action) having seemingly contradictory qualities or phases.” Contradicting stories can often run counter to *doxa*, common opinion. Drawing on this understanding, I refer to paradoxes in this book as situations where ostensibly contradictory stories are both (partly) true. Or, at least, they appear equally true to different stakeholders. In short, I am interested in the societal effects of people subscribing to different, but coexisting, truths.

Like philosophy, organizational psychology has also held a keen interest in paradoxes, albeit not for quite as long. Rothenberg (1996) analyzed the creative processes of outstanding thinkers like Niels Bohr and Albert Einstein and suggested that an ability to think in opposites in tandem lay at the root of their creativity. Bohr, for example, tried to resolve how energy acted both like waves *and* particles. Organizational psychologists today talk about “paradoxical frames” as a form of mindset that enhances thriving and creativity in complex organizational environments full of tensions between competing demands (Miron-Spektor, Ingram, Keller et al. 2018; Liu, Xu, and Zhang 2020). Paradoxical frames are understood as “mental templates that individuals use to embrace seemingly contradictory statements or dimensions of a task or situation” (Miron-Spektor, Gino, and Argote 2011: 229). Personally, I am not suggesting any particular psychological benefits of a paradoxical mindset (nor that such a mindset exists as a measurable disposition), but I do believe that paradoxes are part and parcel of contemporary healthcare organizations.

Paradoxical thinking has also found its way into management theory as a response to “competing demands that cannot be resolved by making trade-offs” (Henriksen, Nielsen, Vikkelsø et al. 2021: 1; see also Luscher and Lewis 2008). Poole and van de Ven noted that when “organizational theories attempt to capture a multifaceted reality with a finite, internally consistent statement, they are essentially incomplete” (Poole and van de Ven 1989: 562). They therefore encouraged organizational analysts to build better theories by embracing paradoxical claims. According to organizational paradox theory, managers often need to do both A and B and adhere to multiple versions of truth to find solutions that work for all organizational actors (Lewis and Smith 2014).

Paradoxical thinking implies embracing ambiguities. Ambiguity contrasts with contemporary policy ambitions around data-driven healthcare systems that promulgate ideas about simple and unambiguous answers. The policy ambition is typically to resolve uncertainty with data and give an answer about effect (yes/no) and efficiency (low/high). Instead of promoting answers of that type (“either/or”), I encourage a willingness to think in terms of “both/and.” It is stimulating to consider paradoxes. Schad and colleagues even quote the philosopher Søren Kierkegaard as an incentive for those inclined to retreat from paradoxes: “The thinker without paradox is like a lover without feeling: a paltry mediocrity” (Schad et al. 2016: 4).

Throughout this book, I thereby point to drivers for and implications of intensified data sourcing that sound like contradictions, but which nevertheless coexist. I will discuss, for example, how data intensification has created both less work and more work; how data both empower and disempower staff and patients; how data both uncover patient concerns and cover up patient concerns; and how data intensity both tightens organizational control and generates new forms of organizational disintegration. In formulating each of these paradoxes, I use the term “data” as shorthand for “intensified data sourcing” because the stories my interlocutors tell about data often set them up as agents in their own right. Indeed, data can exert agential powers, but what they do depend on the types of sourcing and use that they inform. Sometimes a paradox is solved simply by realizing that the same data initiative produces almost opposite effects in different places, or for different people, or during different periods of time. On some occasions, it is a case of the old adage: one person’s meat is another’s poison. In other instances, the paradox becomes a

reason for rethinking the phenomenon at stake—for example, what counts as “patient concerns” if these concerns are both uncovered and covered up? More generally, however, I use the figure of the paradox to ponder on the productivity of apparently exclusive stories. If we accept several propositions as all conferring something partly true, what do their coexistence allow or produce in the given setting? Toward the end of each chapter, I will reflect on this question. The curious reader can peep at the full list of paradoxes that I discuss throughout the book by skipping ahead to table 7.1 in the conclusion, but otherwise I develop them chapter by chapter.

For people inhabiting data-intensive environments, there is also the more experiential dimension of the paradoxes relating to opportunity and risk. The ambiguity that I have described as a matter of contradictory stories about data also reflects something more general in our contemporary engagements with data and digitalization. If the Internet was a liberating space for the competent few in the 1990s, it has now permeated most aspects of everyday life, including that of people who are not the least tech-savvy. And the Internet is clearly no longer a space of freedom, but rather a zone of state power and big business.³ In his memoirs, the whistleblower Edward Snowden, who became known for revealing the data-collection practices of the US National Security Agency, described coming to the uncanny realization that Barlow’s Declaration of Independency had given sway to pervasive surveillance (Snowden 2019a). Today, he claims, anyone with an uncovered camera might have an intelligence officer looking at her or his face when typing. Shoshana Zuboff and others have described how big tech’s accumulation of data assets have become instrumental in the global economic system in ways that also encroach on freedom and privacy (e.g., Zuboff 2019). Any Internet search now provides an opportunity to collect data not just on the object of the search, but on the citizen looking for information. Should I look up a disease online, my search history constitutes a data point about me (or, rather, my Internet Protocol address). Such data can be offered for sale. The infrastructures in place have generated an explosion of opportunity and a wide range of new risks. Living in data-intensive environments is living with the paradox of being simultaneously enriched and exploited, empowered and disempowered.

For the remainder of this introduction, I reflect on the vocabulary that I will use in this book to understand the paradoxes involved in the politics of

intensified data sourcing—“data,” “infrastructures,” and “data politics”—and how this book situates itself in the literature on these topics. I then present some additional reflections on what it implies to see paradoxes as the analytical product; and then I end by introducing the individual chapters.

DATA: EVERYBODY SEEMS TO WANT THEM . . . BUT WHAT DOES THE WORD MEAN?

“Data” is a peculiar word. It is derived from the Latin *dare*, meaning “to give.” The American pragmatist John Dewey once noted that it would have been more appropriate to talk about what is *taken* (Dewey 1929). Data are not given but “*selected* from this total original subject-matter which gives the impetus to knowing; they are discriminated for a purpose—that, namely, of affording signs or evidence to define and locate a problem, and thus give a clue to its resolution” (Dewey 1929: 178, emphasis original). Dewey’s point was epistemic, but it has gained new political relevance as big tech companies now thrive on data accumulation and ways of transforming these traces of our everyday lives into capital accumulation (van Dijck et al. 2018; van Dijck 2013; Sadowski 2019; Zuboff 2019). As financial assets, most data are taken rather than given (Fourcade and Kluttz 2020).

To dwell a little longer on Dewey’s epistemic point, “data” can refer to anything used to inform an understanding. The geographer Rob Kitchin, in his acclaimed book *The Data Revolution*, refers to data as “numbers, characters, symbols, images, sounds, electromagnetic waves, bits—that constitute the building blocks from which information and knowledge are created” (Kitchin 2014: 1). With a similar emphasis on knowledge making, the information scholar Christine Borgman (2015) writes: “Data are representations of observations, objects, or other entities used as evidence of phenomena for the purpose of research or scholarship” (28). Data, it seems, are “building blocks” and “representations” that can take practically any form—so long as they can be made subject to computation. I use the plural form of “data” in recognition of this “building block” view, which in turn characterizes the quest for “more data” to be accumulated as assets (Pinel 2020, 2021; Birch 2017; Birch, Cochrane, and Ward 2021). In healthcare, however, data are no longer just used for “research or scholarship,” as Borgman’s and Kitchin’s definitions suggest. As I will argue throughout the book, this changes how analysts must think about data.

Data always have histories (Loukissas 2019). They are made, not found (Leonelli 2016); they are cooked, never raw (Biruk 2018; Gitelman and Jackson 2013). Data are “inscribed,” as Latour would say (Latour 2014), through processes of computation that make them at the same time exchangeable as stable objects and flexible in the sense of being open for multiple interpretations. They come into being through datafication. One of the most widely circulated books about big data defines “datafication” in this way: “To datify a phenomenon is to put it into a quantified format so it can be tabulated and analyzed” (Mayer-Schönberger and Cukier 2013: 78). Anthropologists similarly discuss “datafication” as a “conversion of qualitative aspects of life into quantified data” (Ruckenstein and Schüll 2017: 261). I, conversely, think we need to distinguish between datafication and quantification. To explain why, I will have to say a bit more about how the term “data” is used in practice.

I had worked with intensified data sourcing for some time and thought a lot about these academic definitions before realizing that I had forgotten to ask what the word “data” meant to the people I was interviewing. When I began asking this question, “What does the word ‘data’ mean when you use it here in your organization?” my interviewees often looked puzzled. As Rosenberg notes, the word “data” is used to discuss other concepts, but not in itself seen as a concept worth exploring (Rosenberg 2013). I now wish to use some of the responses I received to my question to illustrate the evolving meanings of the concept of data. In *Philosophical Investigations*, Ludwig Wittgenstein (2001) argued that words acquire their meaning through familiarity with situations already known to the user. No word carries any inert meaning; the meaning of a word emerges in its *use*. To understand what words mean, Wittgenstein therefore suggested studying actual language-games—the ways in which language is used—as a meaning-producing practice.

In policy papers, at conferences, in interviews, and when directly asked, people mostly refer to data as something digitally recorded, often providing registry data as examples. Anne, from a regional quality assurance organization, first said, “I actually think we have a fairly broad conception of what data are.” She then added that “we talk about both data and metadata.” As I remained quiet, waiting for her to continue, she resumed: “We call it ‘data’ as soon as it can make up a data set with some rows and columns.” For her, data were something to be registered in a structured manner, ready for computation. The definition was not so broad after all—it fitted well with the common focus on quantification in the literature.

Clinicians, conversely, often talk about how “data cannot capture everything.” One general practitioner, Frederik, felt slightly overwhelmed by my invitation to define “data”: “That’s a really big question! When does something in the world turn into data? Well, it involves some kind of registration, I guess. To make it documentable. It is an attempt to turn real-world experiences into something transferable.” Again, data relate to the *transfer* of information, but first and foremost they are about documentation. They need not constitute numbers presented in rows and columns. Clinicians are very aware of the legal aspects of documenting clinical practice, and those data need not be subjected to calculation in order to matter.

At the governmental level, the civil servant Ninna first said that “data” meant “a technical piece of information,” but then spoke about a historical development leading to an expanded understanding of data, where data is more akin to an amorphous mass of information (which is why in this quote, I translate “data” in the singular):

It need not be just a numerical value anymore. It used to be, when people spoke about data, but not anymore. It can be free text and everything. All of this we do see as data today, whereas earlier it needed to fit into a spreadsheet. Today, it is all types of information (. . .) on a continuum to what is really close to something like knowledge.

Here, we see not only an expanded notion of data, but also a slippage from data into knowledge. I asked: “When did this happen?” And Ninna responded:

I’m not sure. (. . .) Traditionally, we made a distinction between pieces of information out in the health services, in the physical health record, health information, also legally. And then, when it was turned into something electronic, for example through entry into our registries, then it became data. But the world isn’t like that anymore. There are no paper records from which selected pieces of information are turned into data. Today, everything is information that circulates.

Today, every piece of clinical documentation is digital right from the start, and thereby it already is an object of potential transfer. It need no longer be “in rows and columns.” When I asked: “Is there anything which isn’t or can’t become data then?” Ninna laughed, then paused, and after a while, said: “Of course, there must be something which isn’t turned into data. I just can’t think of it. (. . .) What matters is traceability. Everything traceable is data.”

Two elements of Ninna’s reflections are particularly worth noticing: the emphasis on traceability/exchangeability and the conflation of data and knowledge. To have data about someone (or something) is seen as *knowing*

something about them. Interestingly, when I interviewed people from patient organizations, they were very aware of this move toward ubiquitous transferability. Hanne, who chaired one of several cancer patient organizations in Denmark, said that “data are something you gather somewhere so that others can go and look at them.” Lone, who suffered from diabetes and who had worked for years in promoting patient rights and giving a voice to patients with chronic conditions, similarly emphasized the exchangeability of data, but she added that it could feel like surrendering parts of her life: “It’s actually my entire life history they ask for. It’s there somewhere [in the digital archives], and they can access all of it.” She thereby conveyed how she, as a patient in a fully digital and integrated system, felt little control over who gets to know what about her.

The pharmaceutical industry and the medical device businesses are also among the actors currently wanting more data. I interviewed an experienced lobbyist, Bent, one afternoon in his organization’s beautiful offices in the capital region of Denmark. He had worked for years to promote industry access to public health data. When I asked which “data” he and his colleagues were lobbying to access, he said:

everything that’s in the registries, the admittance registry, cancer registry, cause-of-death registry and so on (. . .). And then it is everything that the Regions [responsible for running the hospitals] have got. The quality databases (. . .) on a continuum to health records, which are not that structured, but contain information all the same (. . .). And then there are the biobanks. Wet data. (. . .). What I hear industry say, is that they want the knowledge in all of those data.

Bent talks about data as “tools” for capturing “everything,” or else as a “totality” akin to what Zuboff (2019) sees as typical for the big US data companies. This involves conflating many different types of information, in many different media, created for many different purposes, and repurposing them in line with business strategies. Like Ninna, he conflates data and knowledge. He does not talk about numbers: *everything*, in his view, can be turned into information, decontextualized, and transferred to new users who will find a way to make value out of it.

Taken together, these musings on the meaning of “data” show—in line with Wittgenstein’s notion of meaning emerging through language-games—how people associate the word “data” with what they do in the course of their daily work. Running through most of these quotes, these “language-games” in Wittgenstein’s sense, is a dominant sense of something traceable

that can be subjected to computation, decontextualized, and transferred to other users. In this sense, datafication and digitalization are twin developments: digitalization turns all types of records into data. In short, *data are all the many types of information that are traceable, decontextualizable, transferable, and reusable*. Still, in everyday conversation, the sense of data as being what is coded and countable often takes precedence.

Note how people, in these musings on data, often speak about the same data (same in the sense of being a particular entry of, for example, a diagnostic code into an electronic health record), but they want to use them for different purposes (clinical memory, quality assurance and other more administrative goals, research, or to generate economic growth). Infrastructural integration turns the same data entries into elements of very different projects directed toward different objectives (Winthereik 2010). Data are thereby not just different representations in the epistemic sense suggested by Borgman and Kitchin; rather, they are ontologically *multiple* in the sense suggested by Mol (2002): aspects of several things at once. In short, with data intensification, data are not only multiplying; they are becoming multiple.⁴ The people I have interviewed do not recognize this multiplicity. I think they should. In the chapters that follow, I explain why.

Another thing that people rarely acknowledge when talking about data is the materiality of data. The notion of data as “pure information,” which can be found in the thinking of, for example, Negroponte, permeates both policy circles and practitioners. Still, I insist on thinking about how we encounter data in material forms. The media scholar Kate O’Riordan (2017) points out that data always necessitate a medium, however fluid it might be: “Information, like humanity, cannot exist apart from the embodiment that brings it into being as a material entity in the world; and embodiment is always instantiated, local, and specific” (127). Similarly, the science and technology studies (STS) scholar Paul Edwards (2010) notes that “data always have a material aspect. Data are *things* (. . .) with dimensionality, weight, and texture” (84, italics in original). This observation points in the direction of the next analytical concept: infrastructure.

INFRASTRUCTURES: MATERIAL, POLITICAL, AND MORAL

I could have outlined how my interlocutors discuss and make sense of the concept of infrastructure—their language-games—in the same way as I just

did with data. I will not do that. Among the people building health data infrastructures in Denmark, there are anthropologists, information scholars, and people with training in STS. In conversations with me, many of them talk about infrastructures as sociotechnical accomplishments: an interplay of people and technology. Still, these insights tend to be forgotten when investments are made. Policy papers mostly describe infrastructures as “things” that someone will “build.” I will therefore elaborate on how I approach infrastructures analytically to explicate both the thinking that informs my own analysis and a mode of thinking that circulates among some people working in the field I study, but is sidestepped in many investment plans and strategy papers.

Inspired by Star and Ruhleder’s seminal paper (1996) on research infrastructures, I think of infrastructures as *activities* rather than things. They are activities in the sense that they are not stable. What they are depends on what they do, or what people do with them. Everyone and everything connected through an infrastructure can engage in different relations; hence there is not one infrastructure, but many interrelated infrastructures that come into being through different social practices. Unlike large-scale material infrastructures encountered in the public space, such as roads, electricity, and water supplies (Larkin 2013), information infrastructures are difficult to observe empirically. Perhaps this is why Bowker and Star (1999) note how infrastructures are typically invisible until they cease to work. Breakdowns make them stand out and demand attention. Just like roads, data infrastructures need repair. They are never finished (Gupta 2018; Howe, Lockrem, Appel et al. 2016). Infrastructures are more than wires and software: they are what people do with them and because of them.

In her influential work on biological research infrastructures, the philosopher Sabina Leonelli (2014) suggests exploring infrastructures by way of observing the work practices and standards that make up a “data journey” in steps of de-contextualization, re-contextualization and re-use for new purposes. Leonelli focuses on research uses of data, whereas I follow data journeys in multiple directions that reflect the many different uses of data, including research but also administration, clinical use, and industrial profit making. It is a classic insight that any form of coordination demands some kind of work (Strauss, Fagerhaugh, Suczek, and Wiener 1997), and it remains a special task for STS to create awareness of all the tacit work (French 2014; Jensen 2022a), as well as the values and priorities (Carusi and De Grandis

2012), that shape data flows. Infrastructures in healthcare are important. They interact in a most literal sense with bodies, they affect life-and-death decisions, and they can inspire hope, shame, and sorrow (Johansen and Andrews 2016; Langstrup 2013; Petersson 2019). They are means of politics.

DATA POLITICS: THE INTERPLAY OF POWER, KNOWLEDGE, AND TECHNOLOGY

Although I could make people discuss the meaning of the words “data” and “infrastructure,” “data politics” is not a term used by policymakers and practitioners. Some of my informants even juxtapose data and politics, as if they belong to opposite domains. When I say that intensified data sourcing has *political* implications, it stems from my analytical interests—it is not how the people I study frame it. So what do I mean by “data politics”?

Intensified data sourcing is political because data sourcing intervenes in people’s possibilities for achieving their aims. The political scientist David Easton (1953, 1965) famously defined politics as “the authoritative allocation of values for the society.” I aim for something less confined to the processes of state power, something more distributed. Still, I can use Easton’s classic definition to explain what I am thinking of with the term “politics” because intensified data sourcing potentially intervenes in each element of his definition—“authority,” “allocation,” “values,” and “society”—along with affecting what these terms even mean. *Authority* in data-intensive organizations asserts itself through the creation of particular claims to knowledge and by seeking to control access to this knowledge (Lyon 2019). Data are increasingly used to *allocate* resources through performance measurement regimes and algorithmic automation of decision making (Lury and Day 2019). Data also establish and certify *values* as well as communicate them (Prainsack 2019; Sharon and Zandbergen 2016). Finally, it is through data that populations are constructed in ways that make *societies* emerge as governable entities (Grommé and Ruppert 2020; Ruppert 2012; Desrosières 1998; Didier 2009; Tupasela, Snell, and Cañada 2015).

I am not the first to use the concept of data politics. Annelise Riles (2013) refers to data politics as a mode of power where there is ever more data, but with no increase in certainty. Evelyn Ruppert and colleagues discuss data politics as materially mediated interventions in the lives and rights of citizens (Ruppert, Isin, and Bigo 2017). They emphasize the performative

power of data—the ability of data to conjure the objects and relations that they are said to portray. I am fully in line with this approach. The media scholar Tim Jordan (2005) defines a related concept, “information politics,” as “a complex antagonism that is driven by a social and cultural relationship, understood as a dynamic of forces of care and capture, in which some benefit is gained by extracting some kind of value that is lost, and hence impoverishes others” (11). Like Jordan, I want to explore forces of “care and capture” and their uneven effects. Unlike Jordan, however, I do not see data politics as focused on resources alone. Politics, for me, is not a zero-sum game.

Intensified data sourcing affects intimate aspects of people’s identity, emotional struggles, and relations to others. Data politics operates at the knowledge/power nexus. The interconnection between power and knowledge was at the heart of Foucault’s interest in biopower (Foucault 1973, 1991, 2002), and along with the emergence of forms of governing at a distance, this interconnection has been explored in great detail and with incisive elegance in governmentality studies (Dean 2010; Miller and Rose 2008; Rose 1999, 2007). Data serve this ambition of governing at a distance.

Data politics sweeps across many social arenas, not just healthcare. It is changing societies. Datafication almost serves as a label for our current moment in time (Maguire, Langstrup, Danholt, and Gad 2020). In recent years, I have attended many conferences called things like the “Data Moment,” “Data Times,” “The Age of Data,” and so on. The academic literature is inundated with social diagnostic labels trying to capture our time and age: “metric society” (Mau 2019), “audit society” (Power 1997), “dossier society” (Laudon 1986), “evaluation society” (Dahler-Larsen 2012), “risk society” (Beck 1999), “network society” (Castells 2010), “platform society” (van Dijck et al. 2018), “algorithmic society” (Peeters and Schuilenburg 2021), and of course, the popular terms used in politics and media such as “information society” and “knowledge society” (Webster 2002).

Each of these diagnostic labels has been used to describe social developments that are intimately connected with data practices, as well as with wider power struggles revolving around forms of knowing. These labels have been proposed by scholars of astute vision who have contributed with incisive insights into the forces at play in data politics. Nevertheless, it is not my aim with this book to suggest one more diagnostic label. I am not planning to diagnose society. Diagnostic labels are too one-sided in their description to fit the developments I observe. For example, I could say that we live in a

society governed by data with equal conviction as I could claim that we live in a society where emotions hold greater political strength than ever before (Durnová 2019). I therefore aim for a different type of generalization than an overarching “diagnostic”. I believe that data politics cannot be known in an abstract and generic way, but just as important, I believe that thinking with paradoxes can be much more analytically stimulating. What do I offer the reader with this book, then?

THINKING WITH PARADOXES: QUESTIONS AS ANALYTICAL PRODUCTS

This takes me back to why I believe that thinking with paradoxes can be helpful. In place of *one* diagnostic label, they attune analysts to think about several coexisting developments. Paradoxes go beyond simple answers. With my insistence on paradoxes, I aim to stimulate better questions, while at the same time avoiding definite answers that preempt local curiosity. My ambition is in line with Fortun’s (2008: 288) reflections after studying genomic promises in Iceland: “What is needed is an analytical adroitness, a tolerance for contradictions and paradoxes [. . .] and sustained critical involvement.”

Many STS scholars will understand my view of knowledge as an ability to pose relevant questions. It is, however, important to be explicit about this when studying organizations that aim to become *data-driven*. These organizations are permeated by ideas about knowledge as *answers*—answers of positivist certainty. Data projects thrive on a lure of “evidence” without explicating the type of transferability between contexts that should constitute that evidence. The contemporary rhetoric about data makes it necessary to be much more explicit about what you can learn from a specific study of a data practice. There is a lot to learn from case studies. They can make us understand mechanisms of more general significance. We just never know for sure what is relevant in a new case. All situations have unique features. It does not mean that scholars, policymakers, or administrators should abandon the use of data or stop learning from experience. On the contrary, it should make them more curious about which data they should request locally.

If I propose better questions as my main contribution, how did I myself arrive at those questions? Data science typically relies on either hypothesis testing or pattern recognition. As I wrote this book, I did something very different. I conducted a form of ethnographic fieldwork at home. For more than five years, I have participated, observed, interviewed, and collected

cases from news and networks, traced cases, and looked into their background. These types of research data are very different from those data that make it into public registries and health records. They are hermeneutic and phenomenological, discursive and reflective. I kept recalibrating my curiosity with my informants (Marcus 2021). The methodology is elaborated in chapter 4.⁵ I compared notes from various sites to look for commonalities and differences, tensions and contradictions. In this way, I gradually produced lists of drivers and implications. In the course of this, I noted opposing stories about such drivers and implications and then thought about how and why they might all be at least partly true. Seeming opposites coexisted side by side; similar drivers could be related to seemingly opposite effects. In short, I identified perplexing paradoxes. A paradox does not close a case. It is a *both-and* that invites the analyst to linger, contemplate, and explore what remains unruly (Ballesterio and Winthereik 2021). I hope the paradoxes I have compiled can inspire people to look for similar complexities in other settings without being simply lost in nuances.

With awareness of paradoxes, I aim to create a more cautious as well as more playful attitude to data. Why playful? As C. Wright Mills (2000 [1959]) once argued, it is important to avoid using methodological rigor as a reason to stop thinking. I think that awareness of paradoxes can help creating the needed playfulness. My hunch is that paradoxes do a better job of sparking curiosity than p-values and confidence intervals. The point is not to replace or downplay the relevance of p-values and confidence intervals; but to let these methodological tools serve the sense of judgment. Clinicians, administrators, and policymakers should be no more data-driven than researchers. They should use data to exercise their judgment.

Furthermore, those working with data initiatives need to study data practices. As I noted earlier, policymakers rarely define what counts *as* data: they encourage people to *see with* data, but do not *look at* data. With this book, I turn the gaze around to look *at* data rather than just *with* data. We have to care about how data are created, collected, curated, stored, and exchanged, and look at what those processes do. To use data well, those in charge of data analysis in each organization need to take into account the social and political dynamics of data. One of the most striking things about the many reports and strategy papers that propel data promises is probably the absence of discussion of such sociopolitical dynamics. I have read countless consultancy reports and strategy papers claiming to outline opportunities

and risks associated with new data initiatives, and not one of them have mentioned well-known insights such as Berg and Goorman's "law of medical information" or Markus's "theory of knowledge reuse" (Berg and Goorman 1999; Markus 2001). In 1999, well before any of the strategy papers referred to previously in this text, Berg and Goorman presented their iconic observation on data work in healthcare. "The further information has to be able to circulate (i.e. the more different contexts it has to be usable in) the more work is required to disentangle the information from the context of its production." They used this observation to pose a very relevant question concerning the sociopolitical dynamics of health data: "who has to do this work and who reaps the benefits?" (Berg and Goorman 1999: 52).

The information systems researcher Markus drew together available insights on the reuse of data sources back in 2001, identifying four types of reuse that all depend on tacit knowledge about what data mean if users are to arrive at accurate conclusions. She also suggested that to successfully reuse data, the organization needs to analyze its costs and incentives, as well as how it interacts with professional roles among those affected. Insight into the sociopolitical dynamics of data reuse is really not new; it just seems to have escaped the preachers of the messianic gospel about becoming a data-driven society. There is an important role for the social sciences and humanities in recovering this type of lost knowledge (Hoeyer and Winthereik 2022).

THE CHAPTERS

The first five chapters in this book present different aspects of intensified data sourcing. Taken together, they tell a story about what is driving intensified data sourcing and what the implications are for policymakers, staff, and patients. Each chapter develops a conceptual approach to the given aspect: promises, living, work, experiences, and wisdom.

Chapter 1, on *data promises*, gives an introduction to the political promises associated with data intensity and data integration sweeping across all healthcare systems. It outlines how global drivers for intensified data sourcing interact with local opportunities and struggles, describes what makes data appealing to so many people in and around the health services, and presents a categorization of dominant goals with data that the rest of the book draws upon. Data are used for *research*, *clinical*, *administrative-political*, and *industrial* purposes. These purposes relate to four goods: *knowledge*,

health, good governance, and wealth. I also argue that it is important to accept the gravity of the problems people face—their proclaimed data needs—without necessarily accepting the claimed power of the proposed solutions. Rather than buying into the narrative of the technological power characterizing contemporary data promises, I suggest that in many instances, the political power of data revolves less around what data can actually do and more around what it *promises* to produce in the present. Data tools do not always deliver. Instead, they buy time. They make it legitimate to postpone initiatives aimed at helping citizens. Often, data are left unused.

Chapter 2, on *data living*, moves from the high-flying promises to the more mundane implications on the ground. It gives insight into the everyday experience of living in Denmark as a place characterized by a high degree of data integration. To understand what data intensification produces at the level of everyday living involves appreciation of context. I present five types of contexts: new conceptions of health and illness, global rankings, descriptions of key data infrastructures, and narratives of living in the web of a civil registration system assigning a number to each individual. I also argue that global trends in data-intensive medicine acquire local form and meaning as they interact with national infrastructures and the hopes and concerns of local communities.

Chapter 3, on *data work*, describes four types of work that all proliferate in health services as a consequence of data intensification: *production, analysis, instruction, and use.* Despite promises of less work and more automation, many data initiatives involve more work or shift around who does the various types of work. Whereas chapter 2 focused on patients and other citizens, this chapter focuses on health professionals and how increased amounts of data work affect the clinical gaze. Health professionals often consider data work as meaningless and frustrating. Nevertheless, it is not uncommon for clinicians to respond to these frustrations by setting up their own data-gathering initiatives—and produce even more data. I suggest that data have become the lens through which they see their own work, and this makes the grinding data mill accelerate.

Chapter 4, on *data experiences*, continues the discussion of drivers for and implications of intensified data sourcing by focusing on the embodied and emotional reactions to data work. It thereby goes beyond assessing data for their epistemic values (the information they convey or fail to convey) and power effect (what they do as governance tools), and suggests exploring

the experiential dimensions of working with data in a phenomenological sense. In relation to each of the four types of data work discussed in chapter 3 (production, analysis, instruction, and use), people have embodied and emotional experiences. They are affected by their data work. They react emotionally to representations of data, such as to graphs, tables, and color-coded maps. When data are seen merely as epistemological tools, the human engagement with them is reduced to matters of data literacy. The best route to awareness of a more phenomenological understanding of data, I believe, is by working reflectively with our own engagement with data. I therefore build the point on an introspective approach to my own methodology and my own data. Readers who are curious as to my methodology can go directly to this chapter. I conclude the chapter with a presentation of five lessons about data experiences of relevance to researchers as well as practitioners.

Chapter 5, on *data wisdom*, returns once again to the list of purposes with datafication from chapter 1 and addresses the epistemological and normative challenges with reaching these goals. It describes various forms of knowledge and how clinical work depends not only on data, but also on knowledge forms that cannot be datafied. I provide examples of potential errors in a data analysis and use. Wise data use requires forms of expertise beyond what is offered by data scientists. STS, anthropology, and related disciplines can contribute in important ways to build such expertise, and thereby help form “Data Wisdom.”

Chapter 6, *data pandemic*, takes the COVID-19 pandemic as a case that exemplifies the themes and paradoxes from the preceding chapters. The pandemic constitutes a tumultuous and globally significant period where data predictions came to put their mark on societies all over the world. Many citizens became data consumers in new ways, discussing curves and interpretations of test results, death tolls, and even data lingo such as the “R-number.” With colleagues in epidemiology, I initiated a collaborative research project with questionnaires in Denmark, the United Kingdom, Netherlands, and France (for a total of 200,000 respondents) to map the impact of the pandemic. We also did telephone interviews with Danish citizens from the beginning of the lockdown in March 2020 and onward (Clotworthy, Dissing, Nguyen, Jensen et al. 2020; Varga, Bu, Dissing, Elsenburg et al. 2021). I draw on this material to revisit each of the themes from the previous chapters (promise, living, work, experience, and wisdom) and reflect on how the pandemic both confirmed and challenged some of my previous findings. I gradually came

to realize that I had hitherto primarily viewed data as depoliticizing instruments, but the pandemic now clearly showed that attempts of governing through data opens up a Pandora's box of moral and political contestation. We are likely to have to live with the political repercussions of this moment of data politics long after the virus has ceased to dominate everyday life.

The conclusion of this book sums up the themes as a set of data paradoxes. Data intensification revolves around integration of information infrastructures, and this integration implies interactions among various domains, many of which have not previously been connected. If the drivers for and implications of intensified data sourcing give rise to paradoxical stories, how then are policymakers and administrators to anticipate what a data initiative might produce in their organization? To inspire curiosity (and a healthy dose of caution), I suggest a new metaphor to replace the current preference in policy papers for seeing data as "oil" or "gold mines." The linguists Lakoff and Johnson (1980) taught us how metaphors shape our actions, desires, and analytical frameworks. Data are abstract and intangible; metaphors make them more like things—more comprehensible for the human mind. Metaphors such as "oil" or "gold," however, direct the attention to the creation of wealth. It is limiting. I therefore suggest a rather different metaphor: "drugs." If we analyze (and regulate) data with the same care and attention as we analyze and regulate drugs, I believe we can learn to pose more relevant questions than when we think of them as hidden sources of wealth to be mined or drilled out of the ground. I end the conclusion with a critique of the ethics of intensified data sourcing and a call for new ways to think about regulation. The dominant ethics discourses have tended to focus on individual choice (and thereby individual responsibility). I believe that we need to rethink how patients and citizens can gain from the benefits of intensified data without succumbing to the risks.

By exploring the policy, practice, and experience of data-intensive medicine, the entire book represents an ethnographic engagement with data, and like all ethnography, it needs to embrace the ambiguities. As an ethnographic engagement with data, the book draws on and inscribes itself in an emerging corpus of work in anthropology documenting the effects of datafication on healthcare (Adams 2016a; Biruk 2018; Cool 2016; Erikson 2012; Hogle 2016, 2019; Mason 2018; Merry 2016; Storeng and Behage 2017; Taylor-Alexander 2016), while being equally informed by, and intended as a contribution to, work in STS on the role of data in the health services (Hedgecoe 2004; Prainsack 2017; Pickersgill 2019a, 2019b; Sharon and Zandbergen 2016; Sharon

2016; Greene 2007; Tupasela et al. 2020), and data-intensive knowledge practices (Morigera and Palladino 2011; Leonelli 2016, 2012b; Strasser, 2019; Sætnan, Schneider, and Green 2018). The book can thereby be read as a contribution to what various scholars have called critical data studies (Iliadis and Russo 2016), critical (big) data studies (Wyatt 2022), the anthropology of data (Douglas-Jones, Walford, and Seaver 2021), and anthropology of (big) data (Levin 2019). Still, the ambition is not disciplinary: the book is written for anyone interested in how intensified data sourcing affects contemporary healthcare. My ultimate aim with this book is to pave the way for a carefully balanced approach to intensified data sourcing that dares to reach out for new opportunities—while remaining aware of unintended risks.

This is a section of [doi:10.7551/mitpress/14926.001.0001](https://doi.org/10.7551/mitpress/14926.001.0001)

Data Paradoxes

The Politics of Intensified Data Sourcing in Contemporary Healthcare

By: Klaus Hoeyer

Citation:

Data Paradoxes: The Politics of Intensified Data Sourcing in Contemporary Healthcare

By: Klaus Hoeyer

DOI: [10.7551/mitpress/14926.001.0001](https://doi.org/10.7551/mitpress/14926.001.0001)

ISBN (electronic): 9780262374156

Publisher: The MIT Press

Published: 2023

The open access edition of this book was made possible by generous funding and support from MIT Press Direct to Open



The MIT Press

© 2023 Massachusetts Institute of Technology

This work is subject to a Creative Commons CC-BY-NC-ND license.
Subject to such license, all rights are reserved.



The MIT Press would like to thank the anonymous peer reviewers who provided comments on drafts of this book. The generous work of academic experts is essential for establishing the authority and quality of our publications. We acknowledge with gratitude the contributions of these otherwise uncredited readers.

This book was set in Stone Serif and Stone Sans by Westchester Publishing Services.

Library of Congress Cataloging-in-Publication Data

Names: Hoeyer, Klaus, author.

Title: Data paradoxes : the politics of intensified data sourcing in contemporary healthcare / Klaus Hoeyer.

Description: Cambridge, Massachusetts : The MIT Press, [2023] |

Series: Infrastructures series | Includes bibliographical references and index.

Identifiers: LCCN 2022019995 (print) | LCCN 2022019996 (ebook) |

ISBN 9780262545419 (paperback) | ISBN 9780262374163 (epub) |

ISBN 9780262374156 (pdf)

Subjects: LCSH: Medical informatics—Denmark—Case studies. | Medicine—Denmark—Data processing—Case studies. | Medical policy—Denmark—Case studies.

Classification: LCC R858.A3 H64 2023 (print) | LCC R858.A3 (ebook) |

DDC 610.28509489—dc23/eng/20220830

LC record available at <https://lcn.loc.gov/2022019995>

LC ebook record available at <https://lcn.loc.gov/2022019996>