



# The practical ethics of repurposing health data: how to acknowledge invisible data work and the need for prioritization

Sara Green<sup>1</sup> · Line Hillersdal<sup>2</sup> · Jette Holt<sup>3</sup> · Klaus Hoeyer<sup>4</sup> · Sarah Wadmann<sup>5</sup>

Accepted: 4 November 2022 / Published online: 19 November 2022  
© The Author(s), under exclusive licence to Springer Nature B.V. 2022

## Abstract

Throughout the Global North, policymakers invest in large-scale integration of health-data infrastructures to facilitate the reuse of clinical data for administration, research, and innovation. Debates about the ethical implications of data repurposing have focused extensively on issues of patient autonomy and privacy. We suggest that it is time to scrutinize also how the everyday work of healthcare staff is affected by political ambitions of data reuse for an increasing number of purposes, and how different purposes are prioritized. Our analysis builds on ethnographic studies within the Danish healthcare system, which is internationally known for its high degree of digitalization and well-connected data infrastructures. Although data repurposing ought to be relatively seamless in this context, we demonstrate how it involves costs and trade-offs for those who produce and use health data. Even when IT systems and automation strategies are introduced to enhance efficiency and reduce data work, they can end up generating new forms of data work and fragmentation of clinically relevant information. We identify five types of data work related to the production, completion, validation, sorting, and recontextualization of health data. Each of these requires medical expertise and clinical resources. We propose that the implications for these forms of data work should be considered early in the planning stages of initiatives for large-scale data sharing and reuse, such as the European Health Data Space. We believe that political awareness of clinical costs and trade-offs related to such data work can provide better and more informed decisions about data repurposing.

**Keywords** Automation · Data friction · Data repurposing · Data work · Denmark · Digitalization · European Health Data Space · Healthcare · Healthcare priorities · Practical ethics

---

✉ Sara Green  
sara.green@ind.ku.dk

Line Hillersdal  
line.hillersdal@anthro.ku.dk

Jette Holt  
jho@ssi.dk

Klaus Hoeyer  
klho@sund.ku.dk

Sarah Wadmann  
sawa@vive.dk

<sup>2</sup> Department of Anthropology, University of Copenhagen, Øster Farimagsgade 5, 1353 Copenhagen K, Denmark

<sup>3</sup> Infectious Disease Epidemiology & Prevention, The National Center for Infection Control (CEI), Artillerivej 5, 2300 Copenhagen S, Denmark

<sup>4</sup> Centre for Medical Science and Technology Studies, Department of Public Health, University of Copenhagen, Øster Farimagsgade 5, 1014 Copenhagen K, Denmark

<sup>5</sup> The Danish Center for Social Science Research, VIVE, Herluf Trolles Gade 11, 1052 Copenhagen, Denmark

<sup>1</sup> Section for History and Philosophy of Science, Department of Science Education, University of Copenhagen, Niels Bohr Building (NBB), Universitetsparken 5, 2100 Copenhagen Ø, Denmark

## Introduction

Many countries throughout the Global North currently invest in data intensive resourcing in healthcare (Hogle 2016; Hoeyer et al. 2019). Health data has become a kind of capital that is “collected, stored, and traded for the future benefits it is believed to bring” (Barilan and Brusa 2022, p. 2). Comprehensive health-data infrastructures are hoped to simultaneously improve medicine through strengthened research and capitalize on new data economies for pharmaceutical innovation in precision medicine (Tarkkala et al. 2019; Boniolo 2022). In the USA, for example, massive funding has been channeled into digitalization through the stimulus package enacted in the wake of the financial crisis. In the UK, there have been multiple initiatives to facilitate research, including the care.data initiative, the 100,000 Genomes Project and collaborations with Google DeepMind and more recently Palantir. In Australia, investments have been made into an infrastructure called My Health Record that gathers health data on a nationally integrated platform. Also in Europe, there are many national initiatives, such as the French Health Data Hub and the Finnish Findata. Recently, the European Union (EU) has also initiated an infrastructure for reuse of pharmaceutical data called EU Darwin. An even more ambitious initiative is the planned European Health Data Space (EHDS), a data infrastructure envisioned to facilitate data sharing and reuse of health data for citizens, health professionals, administrators, researchers, and industry across the union’s member states. Common to these diverse initiatives is a shared policy vision: to use digitalization to pave the way for integration and repurposing of health data for administration, political governance, public health surveillance, research, innovation, and economic growth (European Commission 2020b, c). Political strategies thereby seem to assume a straightforward compatibility between multiple uses of health data. But what does the political goal of repurposing of health data for non-clinical purposes entail for clinical work? Which costs and trade-offs may be involved for those who produce data in the first place?

The policy goal of data repurposing is fueled by a widespread discourse around health data being an untapped resource for bioeconomic policies (OECD 2013), as “repositories for data ready for statistical analysis” (Barilan and Brusa 2022, p. 2). In the Nordic countries, health data are often framed as an unexploited “goldmine”, the data being “gold” that could be extracted from the existing—and highly integrated—data sources, to promote health and wealth (e.g., Nordfors 2014; see also Tarkkala et al. 2019; Tupasela 2021). In the UK, the Department of Health has stressed the need to “liberate” national health

data for additional purposes, via new infrastructures that will improve the “flow of information between organizations” (UK Department of Health 2011: 48). In the US, the National Academy of Science has suggested that drug development can be speeded up by making clinical data immediately available for research via “digital commons” (NAS 2011). Similarly, the European Commission envisions that “a single market for data will allow it to flow freely within the EU and across sectors for the benefit of businesses, researchers and public administrators” (European Commission 2019: introduction; see also European Commission 2010, 2016, 2020a, b, c, p. 9). The European Health Data Space is, for example, framed in these documents as “unleashing the potential” of health data—as if the many objectives can be reached at no cost, simply by removing “barriers.”

The metaphor of “data flow,” often used in policy reports, suggests that integration and reinterpretation of data are about ensuring that nothing “stops” the flow, as if data were water moving in pipes. However, a growing literature in philosophy of science and social science demonstrates that data integration and repurposing are far from straightforward but require meticulous data work and expertise to succeed (e.g., Hogle 2016; Leonelli 2014; 2016; Bossen et al. 2019; Gabrielsen 2020; Pine et al. 2020; Hoeyer 2023). While ethical debates about data reuse have raised important points about privacy, autonomy, discrimination, and inequality (see below), the reframing of health data as “assets” for administration, research, and innovation can also include costs and trade-offs in need of ethical attention (see also Hunt et al. 2017; Vezyridis et al. 2017; Birch et al. 2021; Pinel 2021). We contend that to minimize the friction between clinical needs and the aim of data repurposing, policy makers need to set priorities early on, in the planning stages of new infrastructural initiatives. The main aim of our paper is to explore what such costs and trade-offs consist of. We do so by unpacking the “invisible” data work (Star 1991; Bowker and Star 1999) in clinical practice that is associated with new strategies aimed at using health data for an increasing number of purposes.

The functions of health records have expanded significantly in the last decades. From being primarily a tool for clinical record keeping and communication, electronic health records increasingly also serve purposes such as quality analysis and management, financial administration, as well as research and innovation (Winthereik et al. 2007; Vezyridis and Timmons 2021). Political visions to reuse health data for multiple purposes are facilitated by digitalization of healthcare systems, but they also shape the digitalization process itself through choices of data infrastructure design. Infrastructure design, in turn, impact working conditions for the users. Zuboff documented already in 1988 how digitalization and automation via “smart machines” can

improve information processing and work productivity but also negatively disrupt working conditions and collaborations among employees (Zuboff 1988). Zuboff stresses that digital technologies can “take many forms depending upon the social and economic logics that bring it to life” (Zuboff 2019, p. 15). As in Zuboff’s work, the target of our critical analysis is not technologies or digitalization as such but the “logic that imbues and commands it into action” (Ibid, p. 15). We are interested in understanding the consequences of using large-scale digitalization and data integration as means for various purposes. We therefore focus on strategies that are motivated and shaped by the political desire to create uniform and reusable data via a system that simultaneously can cater for multiple purposes.

The study builds on ethnographic studies in the Danish healthcare system. Denmark is a small welfare state with a population of 5.8 million citizens. The Danish healthcare system is particularly apt for studying what data repurposing involves because it is highly digitalized and integrated (Schmidt et al. 2019; United Nations 2020). Denmark is considered an international pacesetter in the use of electronic health records, which are mandatory in both primary and secondary care, and there is a high degree of public trust in data handling by health officials. Healthcare is universal and accessible to all registered citizens with a CPR-number, a personal identifier that is also used to link information on all citizens’ encounters with public and private health providers (Ministry of Health 2017). Since 1968, the personal identifier has been assigned to all Danes at birth and enables the establishment of lifelong data trajectories at the individual level and across sectors—a feature that has made Denmark internationally known as “the epidemiologist’s dream” (Frank 2000; Bauer 2014; Schmidt et al. 2019). Tupasela (2021) describes the political strategies in Nordic countries, including Denmark and Finland, as *population branding* that reframes healthcare systems, health data, and populations as assets for developments of and investments in genomic medicine. These include centralized health data registers, wide healthcare coverage, the relatively high genetic homogeneity of the populations, as well as the high public trust in data collection. These features make Nordic health databases significantly different from countries with more diverse populations and more fragmented and less inclusive healthcare systems, such as the US (Dawes 2020). Because of these features, Denmark has been promoted as a “digital frontrunner” for the European Health Data Space (Digitaliseringspartnerskabet 2021, p. 19). Since other countries may pursue similar paths, the experience gained in Denmark is therefore likely to be of wider international relevance.

The paper is structured as follows. First, we briefly introduce some of the academic discussions about repurposing of health data before we present the Danish case and our methods. In the analysis, we outline five types of clinical

data work that proliferate with increased emphasis on data repurposing. In the discussion and conclusion, we highlight the need for political priority-setting when planning large-scale infrastructures for data repurposing.

## The practical ethics of repurposing health data

Ethical debates about the repurposing of health data focus prominently on principal values related to the rights of data subjects, like privacy and autonomy. For instance, in an influential review of ethical debates on Big Data practices, Mittelstadt and Floridi (2016) highlight the issues of informed consent, privacy, data ownership, epistemological challenges, and inequality in power as key concerns when data are used for profiling and surveillance. These are important themes. Yet, other kinds of ethical issues may also be at stake in the daily work of those who are to produce data. Rather than being a question of principle, such issues may be expressed as trade-offs or frictions that are to be handled as part of everyday practices; what other scholars have referred to as *practical* or *empirical ethics* (Hoffmaster 1992; Arribas-Ayllon et al. 2011; Pols 2015). By attending to practical ethical concerns related to the *production* and *use* of data in clinical practice, it is possible to point out challenges associated with, for instance, incompatible data formats and limited interoperability of IT solutions developed for different needs (Kruse et al. 2016). Scholarly attention has been already given to the repurposing of health data for research (Tempini 2020) and for commerce (Birch et al. 2021; Pinel 2020; Vezyridis et al. 2017); and the use of clinical data for administrative or juridical purposes has also been addressed (Hunt et al. 2017; Wiener 2000). These studies point out that the alleged benefits of data purposing also come with potential costs. We show how *frictions* may arise when additional *data work* is required in clinical settings as a precondition to facilitate repurposing of health data.

The concept of *data work* refers to the skilled and distributed labor involved in producing, documenting, curating, storing, and disseminating data, as well as the efforts required to make sense of them (Bonde et al. 2019). Berg and Goorman’s (1999) seminal paper on the repurposing of health data highlighted how the amount of data work increases with the *number of* and *distance between* different uses of data:

The further information has to be able to circulate (i.e., the more diverse contexts it has to be usable in), the more work is required to disentangle the information from the context of its production (Berg and Goorman 1999, p. 51).

Berg and Goorman termed this statement the “law of medical information.” Similarly, Leonelli (2016) has highlighted the requirements of advanced infrastructures, long-term planning and skilled labor to “package data for travel” in the context of biological and biomedical research. Rather than the metaphor of “data flow,” Leonelli prefers the notion of “data journeys” to highlight that data integration is often delayed, disrupted, or retransferred due to a lack of resources for data curation (Leonelli 2014). Data curation requires expertise in the local context of data production, as well as the epistemic interests of new users, and thus involves tasks that are not easy to automate (see also Akrich 1992; Leonelli and Tempini 2020; Tempini et al. 2020).

Using electronic health information is considered an important component of evidence-based medicine, and non-clinical purposes can benefit clinical work, e.g., through quality control and strengthened biomedical research and innovation for development of future treatments. At the same time, digitalization often does not improve patient care or clinical practices in a straightforward manner or without costs (Fiander et al. 2015). While the introduction of new data infrastructures is intended to improve information flow and reduce data work in the clinic, the experience in practice is often the opposite (Vikkelsø 2005; Downing et al. 2018). Because of existing discontinuities in data formats and local differences in reporting standards, new infrastructures often redistribute, rather than eliminate, data work. New forms of data work are required to collect, check, clean, store, and reformat data in ways that comply with new systems and additional users to make data meaningful. Data work in clinical practice following the introduction of new IT systems has been described as “invisible” or “hidden” (Star 1991), in the sense that it is often taken for granted, or not included, in rationalized models of how IT systems influence work tasks or measurements of hospital productivity (Bowker and Star 1999, p. 245; Timmermans and Berg 2003; Bonde et al. 2019; Fiske et al. 2019; McVey et al. 2021). When data work is experienced as draining resources from other tasks in scientific or medical practice, it can lead to what has been described as *data friction* (Edwards 2010; Edwards et al. 2011). The notion of data friction highlights that the transformation and movement of data always consume energy. Data friction, like physical friction, can be productive in the sense that the energy consumed can be converted into new possibilities (Bonde et al. 2019). Yet, in some cases, the frictional cost of data transformation may exceed the resources available and lead to a decline in productivity or system collapse. Disruptions in data work can therefore be considered as a *practical ethical problem* for the functioning of healthcare systems.

## Methods

Individually and collectively, we have explored data-integration aimed at repurposing data in the Danish healthcare system through observations of data work, interviews and informal discussions with clinicians, data analysts, and policy makers, as well as analysis of policy debates as they are expressed in strategy papers, health policy magazines, and the bulletins of health professionals’ associations. Wadmann has undertaken 36 h of observation and conducted nine semi-structured interviews with clinical staff and managers in two psychiatric centers. Hillersdal has observed patient consultations and the daily activities of oncologists and nurses at the cancer research unit for experimental drug trials. In addition, Hillersdal has observed clinical practice and interviewed nine clinical staff, five research nurses, three data consultants, three industry partners, and 16 participating patients. Holt has carried out observation and semi-structured interviews with 23 infection prevention and control nurses, 10 clinical microbiologists and a doctor in 14 infection control units. Hoeyer has interviewed data analysts, administrators and policymakers working with data integration across municipal, regional, and state levels, which has complemented our understanding of the strategy papers, though we do not specifically quote these interviews here. Moreover, we have all participated in public and politically organized meetings, where the organization of and ambitions for the future Danish healthcare system have been discussed.

This paper draws on ethnographic studies carried out by the authors towards various individual research ends (Wadmann and Hoeyer 2018; Holt 2020; Hillerdal and Svendsen 2022; Hoeyer 2023). We found similar issues arising in different contexts and decided to begin comparing systematically across case studies. For example, particular large-scale investment in a new electronic health record system from the American supplier EPIC came up in all our studies as presenting similar challenges. In analyzing our material, we categorized the types of data work involved through an iterative process of identifying themes (Madden 2010) and revisiting materials to look for differences and similarities across sites. All translations from Danish to English were made by the authors. In Denmark, qualitative research is not subject to approval from an ethics committee. The collection and use of empirical examples comply with the requirements of the EU’s General Data Protection Regulation (GDPR).

## Context: policy visions and data work in a highly integrated data infrastructure

While Denmark is ahead of many other countries in terms of integrating infrastructures for health data, issues related to the prioritizing of financial investments into data repurposing are of relevance far beyond the Danish setting. Such investments relate, for instance, to the development and implementation of digital equipment and shared communication standards. As mentioned in the introduction, an ambitious example is the European Health Data Space initiative, which involves the harmonization of standards for electronic health records (European Commission 2020c). Harmonization is needed to enable automated data transfers, and automation is needed to ensure seamless data availability, completeness, and ease. Such efforts challenge the old distinction between “primary” and “secondary” data use (e.g., Markus 2001), because data are intended to take formats that work equally well for multiple purposes. The Danish national strategy for digitalization from 2018 states illustrates this by highlighting that “with the new data-driven technologies, the [clinical and non-clinical] purposes increasingly supplement each other,” fostering a growing “reciprocity” in the use of health data (Sundheds- og Ældreministeriet, Finansministeriet, Danske Regioner and KL 2018, p. 4, our translation). Similar views are expressed in ambitions to develop infrastructures enabling health data to be used for research and innovation in addition to clinical purposes (Danske Regioner 2015; Danske Regioner og Dansk Industri 2019; Digitaliseringspartnerskabet 2021; Ministry of Health 2016; Regeringen 2021). The most recent Danish digitalization strategy, from May 2022, highlights a vision for “Better use of health data for the benefit of Danish patients, as well as research and development of innovative life science solutions through, among other things, realizing a vision for better use of health data, [and for] one common access point to health data for research and innovation, etc.” (Regeringen 2022: 37, our translation). Thus, the clinical aims for producing and using data no longer hold primacy in defining data standards.

Political strategy papers brand Denmark as an ideal context for life science research and drug development (Danske Regioner 2015; Sundhedsministeriet og Danske Regioner 2021; Ministry of Foreign Affairs 2014). Like other Nordic countries, such as Finland, it is a key policy vision to use health data to attract international commercial investments in research and innovation (Tupasela 2021). To attract biomedical companies, Denmark established an infrastructural project termed *Trial Nation* in 2018. It is a merger of previous initiatives to attract global investments in clinical trials in Denmark, by offering pharmaceutical companies a single-entry point to health data, thus making it easier to

identify candidate patients for trials.<sup>1</sup> Another milestone was the launching of the National Genome Center in 2019 to facilitate population-wide collection and integration of genomic and health data (Novo Nordisk Foundation 2018; Danish National Genome Center 2019). Such initiatives received further support with a new Life Science Strategy, published by the Danish government (Regeringen 2021). It is a common feature, globally, of data integration initiatives that countries compete against each other (Vezyridis and Timmons 2021) and increasingly use health data as assets aimed at “branding” (Tupasela 2021). The ambition to reuse health data for research, innovation, and administration installs new demands on data quality and availability, as well as on the standardization and completeness of datasets (Petersen 2019).

Record keeping of health data has been mostly digitized for decades in Denmark, but ambitions to develop a nationwide platform for electronic health records have not yet materialized. Danish hospitals are managed in five regions with their own political levels of management. In 2016, the IT system *Sundhedsplatformen* (Danish for “the health platform”) was delivered by the American EPIC company in two of these five Danish regions. It was presented as a move to further integrate a range of different systems already in use, and to facilitate effective and fast data repurposing (Bentzen and Rosenberg 2021, p. 22). The system was also said to improve continuity and transparency in patient information and patient safety, as well as to optimize workflows and data reporting by requiring hospital doctors to write directly into the patient record (Drachman and Davidsen-Nielsen 2018). This was the largest IT investment in Danish healthcare (2.8 billion DKK, or 458M US\$) and therefore it is worthwhile studying what it entailed in more detail. We use this IT-system to study the consequences of implementing a centralized data infrastructure, which was presented as a way to improve documentation and quality control for clinical purposes, but whose design is shaped by the desire to create uniform and reusable data for functions beyond these. A highlighted virtue of the new system was the availability of automation functions to ease the so-called documentation burden and time spent on data reporting for clinicians. This “automation” was simultaneously expected to introduce standards that could facilitate reuse. Yet, the practical ethical concerns experienced by healthcare staff were not adequately addressed via the suggested automation practices. To explain why, we now turn to the empirical analysis where we outline the five types of data work we identified and the frictions they entailed. The data work relates to the production, completion, validation, sorting, and recontextualization of data. We feature examples from the introduction of the EPIC

<sup>1</sup> For more information, see <https://trialnation.dk>.

system because of the high political expectations of, and the vast investments into, this IT solution.

### **Production: data work proliferates through parallel registrations**

The first observation we made across our ethnographic studies was that the repurposing of clinical data, via new software, increases the time that healthcare staff spend on data work. IT systems that prioritize features for data repurposing often come with complicated procedures for data registration, and moreover, an increasing amount of data are to be registered. Yet, despite more data being registered, the systems do not always make it easier for healthcare staff to *find* the information needed for clinical care, because the system is not necessarily designed with the “primary” user in mind. We observed how parallel documentation systems tended to emerge, when new user interfaces contained too little or too much information, making it difficult for healthcare professionals to get an overview of the patient’s current health issues. For example, Wadmann observed how nurses in a psychiatric center not only registered the results of electroconvulsive therapy in the IT system (Sundhedsplatformen) but also printed and displayed the graphs on a paper card for each patient. Because the software platform did not provide a chronological overview of the patients’ responses to a series of treatments, the health professionals continued with their own analogue solution in parallel to the digital platform. Other examples of what the healthcare staff termed “handheld data” were manually registered systems of information needed to keep an overview of treatment capacity and patient transfers. Although these types of data are central to the workflow of the clinics, the IT system did not provide an accurate overview of data to inform clinical decisions. Despite being more centralized, Sundhedsplatformen provided a more fragmented picture for the clinical user.

Although parallel data work is often necessary for clinical purposes to fix a problem introduced by a new infrastructure, this work remains largely invisible for data users outside the clinical setting. They see the data they request, not the work it takes to produce handheld registrations. The observation of Wadmann in the psychiatric center resonates with experiences of clinicians in other psychiatric units, who also reported on difficulties of retrieving clinically relevant information and a substantial increase in data work after the implementation of Sundhedsplatformen (Overlægerådet i Region Hovedstadens Psykiatri, 2018). This type of “frictional cost” associated with implementing the EPIC system was substantial also beyond psychiatric care. It should here be mentioned that some of the challenges may relate to the lack of knowledge among some users about functions supported by the system, and that improvements to

Sundhedsplatformen have been continuously made to optimize clinical functions. However, the need for additional training in how to report and use data in new IT-systems, as well as the development of and training in additional functions to meet user needs, can also be considered a type of data work that implies a “production cost”. In this case, according to a national audit, the implementation of the EPIC system increased data work and reportedly led to a decline in productivity, concerns about patient safety, and staff burnout (Rigsrevisionen 2018; Bentzon and Rosenberg 2021).

### **Completion: extending data work to support research and administration**

Non-clinical purposes of data use, like research and administration, are often more dependent on data completeness than clinical work. A second form of data work observed in the clinical sites was therefore related to data completion. Clinical research has always depended on data. However, the increasing use of health data as assets to attract investments (Vezyridis and Timmons 2021) means that data production in clinical settings is taken to another level. Completeness of health datasets become an end in itself, because it is considered as a resource for economic growth via pharmaceutical investments and innovation. Danish cancer treatment trials exemplify the attempt to brand Danish clinics as the ideal sites for investments by the life science industry. In return for sponsoring clinical trials, companies gain access to very detailed, high-quality datasets. This involves extensive questionnaires, repeated testing and sampling with increased precision, as well as reporting according to the standards relevant for research. This granularity and complexity of data go beyond traditional clinical trials. We illustrate this through observations by Hillersdal, who studied data work in a cancer clinic that specializes in early phase-one drug trials of targeted treatments.

Patients enrolled in one of the approximately 150 open trial protocols were meticulously monitored, such as through electrocardiograms (ECG), blood samples and the registration of performance status and symptoms. External data monitors were hired by the pharmaceutical company to control data quality while the generation of clinical data was undertaken by clinical research nurses in the unit. In interviews, the physicians and nurses in the clinical unit emphasized that a substantial amount of their time is spent on delivering complete data to qualify the unit for future trials. “Complete” datasets are crucial for securing upcoming “slots” in the competitive market of investments in cancer trials by big pharma corporations. This can give Danish cancer treatments new experimental treatment options. But this strategy also has a cost. The work to complete the

datasets for research purposes was experienced as a drain on resources for both health personnel and the cancer patients. Clinical nurses spent considerable time translating the trial protocol into a clinical “work sheet” and retrieving data or test results that were required to fulfill the demands of the research protocol. Physicians commented on the drain of resources in response to regulatory demands for standardization procedures and for more and repeated testing on an increasingly narrow patient population. The physicians stated that the increasing resource requirements for running the clinical research trial meant that the unit could treat fewer patients at a time, that is, that fewer terminal cancer patients could be offered a place in the experimental treatment protocol. Moreover, the physicians found it ethically challenging to expose these patients to the increasing test demands, and to underline the workload, trial participation was presented to patients as a part-time job! The example highlights how the ambition of the so-called “reciprocal” use of health data to attract research investments involves a substantial increase in data work, imposed not only on healthcare staff but also on patients. If this data work is not accounted for, the branding of health data can put additional pressure on clinical units. This is particularly the case in a context of high-speed global competition to attract commercial investments, and where the clinic must adapt to the industrial research agenda (Hillersdal and Svendsen 2022).

In the handheld *data production* discussed in the previous section, additional data work is required to support clinical tasks, because some functions of the IT system do not prioritize the needs of clinical users. *Data completion* involves additional reporting in existing systems because the use of health data for secondary purposes comes with a call for more data and of higher quality. Our example illustrates how what counts as “improved data quality” is dependent on the context of the user, as complete datasets are not always clinically relevant. The demand for data completion may be particularly evident in clinical research units. Nonetheless, it is a common experience that additional data are required to ensure completeness when health data are needed for non-clinical purposes, including also quality assessment and administration (Petersen 2019). For example, standardized IT systems for electronic health records often require the registration of vital signs (body temperature, blood pressure, pulse), because “completeness” of data is considered important for comparative data analysis. In the Danish electronic health record system, it is possible for health personnel to choose the “not relevant” option if such measurements are not deemed clinically relevant. Nevertheless, it still takes time to fill in all the mandatory entries. When data work consumes considerable time and energy, without adding value to what clinicians see as their primary work, data work can be experienced as “meaningless” (Hoeyer and Wadmann 2020). This “frictional cost” can become an

ethical challenge when data work needs to be prioritized over clinically relevant tasks.

### **Validation: data work proliferates to ensure authentication**

By *data validation* we refer to data work meant to ensure that the right data are reported in the relevant places for both clinical and non-clinical purposes. Data validation has always been part of record keeping in healthcare systems, but the increasing complexity of IT-infrastructures and additional uses of data also increase this type of data work. While new IT solutions often come with automation functions intended to reduce the need for manual data work, we have observed how automation can also generate new tasks of data validation.

Sundhedsplatformen offers auto-generated text and suggestions for data to be included in medical notes and records, for instance test results retrieved from other parts of the health record. This type of automation is intended to help clinicians include relevant data in the electronic health records without having to search for the information elsewhere. However, the time saved on *retrieving* the relevant data is often countered by a need to *validate* the automatically retrieved data. A physician in a psychiatric unit explained that it required extra work to find out where the data came from and whether they were relevant and valid for the specific patient encounter. The physician referred to the auto-generated data as “noise” because their clinical relevance could be questionable, or it was unclear whether the data were up to date. In such cases, the physician had to spend additional time to find out where the data originated from, when the data were registered, and judge their relevance for the particular patient. Thus, the implementation of an automation strategy *intended to reduce* manual data work instead created a need for *additional tasks* of data validation. These issues also have led to concerns about patient safety, as discussed in several articles in a special theme on Sundhedsplatformen in the journal of the Danish Medical Association.<sup>2</sup>

The need for data validation also arose due to the redistribution of data work. To ensure “real-time data,” new IT systems are often designed to foster “direct registration” by health professionals during or immediately after patient contacts. In Denmark, this was also the case for Sundhedsplatformen. While medical secretaries previously had the tasks of transcribing dictated recordings, making requisitions, entering disease-specific codes, and making the bookings necessary for patient transfers, these tasks were redistributed

<sup>2</sup> <https://ugeskriftet.dk/tags/sundhedsplatformen>, accessed Nov 24, 2021.

to physicians. This change in the distribution of data work was envisioned to minimize delays in data registration and reduce the risk of errors, while also making the data work of medical secretaries obsolete. For political-administrative decision-makers, the possibility of removing the “extra layer” of data work performed by secretaries was part of the business case of the IT-investment, and the costs of implementing Sundhedsplatformen were expected to be paid off over 10 years due to increased productivity (Drachman and Davidsen-Nielsen 2018). Hundreds of medical secretaries were now officially made redundant, corresponding to about two percent of the total number of employees in the hospital sector (Bentzon and Rosenberg 2021). However, the redistribution of the data work from secretaries to physicians was not as seamless as envisioned.

Reporting errors grew as the physicians did not have the time required for careful reporting and identification of missing information, nor the administrative expertise required to code data correctly, for instance to link diagnostic codes to reimbursement codes.<sup>3</sup> As the financial consequences for the hospitals became clear (e.g., due to missing reimbursement), a re-hiring process of medical secretaries began. But the function of the secretaries changed: from a role as main data producers, the secretaries now had to *verify the physicians’ data production*. A secretary in a psychiatry unit described her new role as a “controller-function” to emphasize her primary task of data validation. Yet, she also commented that secretaries often took on administrative tasks assigned to the physicians (e.g., sending referrals or adding reimbursement codes), because physicians struggled to use the new system. Indeed, hospital administrators have recently highlighted that the need for employees to register and manage data is even higher than before the implementation of Sundhedsplatformen (Tiirikainen and Rasmussen 2021).<sup>4</sup>

### Sorting: data work proliferates to make data findable

The policy goal of repurposing health data comes with demands for the registration of increasing amounts of data, but also with suggestions for how to minimize data work

through IT systems offering automation and easier access by multiple users. Integrated IT systems are intended to provide health personnel with the ability to access *all* data on a specific patient from one entry point. For example, Sundhedsplatformen was intended to provide more continuity in data access through a single entry, instead of having to log on several times to multiple systems. Yet, automation functions and access to more data now came with the trade-off of more data work related to *sorting* information.

Several of the physicians interviewed explained how Sundhedsplatformen provided a “tangle of notes” and resulted in “data overload.” A chief physician described the data as “unfiltered and unstructured” and explained that it took additional time to sort and find the relevant data to support clinical decision-making. The need for sorting arose when the complexity of user interfaces made it difficult for healthcare staff to find the *relevant* information. Moreover, some of the automation functions designed to ensure data completeness meant that healthcare staff were often presented with a volume of data that exceeded their needs. Ironically, this challenge was brought about by automation strategies that were envisioned to ease data reporting and improve information transfer across health units. For example, templates or “smart text” consisting of standard phrases (inserted via a shortcut key) were to be used in referral situations to ease data registration and reduce the loss of information across providers, such as between primary and secondary care. However, as the volume of data in referral letters increased, it took more and more time to get a “quick” overview of the patient’s current medical condition. For GPs to cope with the vast amount of data in referral letters from hospitals, a new algorithm had to be developed to highlight only the data of relevance to the GPs (Allen 2019). It is telling how this sorting algorithm had to be introduced to cope with the data overload produced by another automation strategy originally intended to *reduce* the need for data work.

The integration of multiple IT systems and the implementation of automation functions are envisioned to make more data available for clinical decision making, as well as for secondary users. However, automated data sharing and the increasing use of smart phrases and copy-paste functions have also been associated with a risk of *note bloat*, that is, user interfaces ending up containing too much (clinically irrelevant) information while the essential information gets buried in the details (Weis and Levy 2014; Wang et al. 2017; From et al. 2019). Thus, while automated solutions can save time on data work related to data production and transfer, they also risk generating new types of data work related to the sorting of information.

<sup>3</sup> In Denmark, diagnostic codes are linked to so-called Diagnosis-Related Groups (DRG), which are used for purposes of remuneration (Bossen 2011).

<sup>4</sup> Studies of data work in the American healthcare system also illustrate how increasing demands for data work have led to the emergence of new professions, such as medical scribes and clinical documentation integrity specialists, CDIS (Bossen et al. 2019; Pine and Bossen 2020). The latter have significantly higher salaries than average coders, because they have a clinical background and extensive clinical experience, thus illustrating a drain not only on financial resources but also clinical expertise.



## Recontextualization: the expertise needed to interpret health data

The political aim of repurposing health data presupposes that data can be analyzed, disseminated, and interpreted for use in new contexts. Leonelli (2014, 2016) has emphasized how the reuse of biological research data via large databases requires *decontextualization* and *recontextualization* of data. These processes involve data work such as reformatting data to comply with standardized annotation to minimize differences in data collected at different sites (decontextualization), as well as the compiling of additional information about a given dataset (metadata). These processes in turn enable the repurposing of data as evidence in different contexts (recontextualization). We include recontextualization as our fifth type of data work to highlight how medical expertise, contextual experience, and clinical resources are needed to ensure robust interpretation of health data beyond the original site of production.

From interviews, we have learned about the types of local expertise it takes to integrate and interpret seemingly simple data. Even something as straightforward as integrating test results from the measurement of blood cholesterol or blood pressure depends on clinical knowledge and awareness of local and historical contexts. Often, health data are not registered using the same digits or measured via the same instruments. Blood cholesterol has for instance been reported by some laboratories as being above or below a specific guideline level, rather than in absolute numbers, and guidelines for what is considered “normal” or “high” have changed over time. Similarly, disease and risk classification guidelines are continuously updated and changed, and diagnostic criteria for the diagnosis of, e.g., heart failure, type 2 diabetes, and hypertension, have changed over time (see also Ellingsen and Monteiro 2003).

Further, a clear illustration of the need for recontextualization arose in relation to Holt’s study of quality control in Danish hospitals. To prevent and control hospital-acquired infections, thereby both improving the cost-effectiveness and quality of care, and reducing patients’ suffering, a national automated incidence monitoring system was launched in 2015. The database called HAIBA (Hospital Acquired Infections dataBase) is accessible online and exemplifies the political vision of ensuring greater transparency in quality improvement and patient safety for patients as well as professionals. Producing data on hospital-acquired infections was intended to document treatment trajectories for specific patients, as well as to monitor developments within specific units. However, administrators have also seen potentials for using HAIBA to compare the performance of different clinical units—a kind of benchmarking.

While infection monitoring may seem like a straightforward way to repurpose already available health data, Holt’s fieldwork reveals a more complicated picture. A physician specializing in clinical microbiology highlighted the risks of uncritically interpreting aggregated data without proper insight into how the data were produced. Reading a report that incorporated HAIBA data to evaluate infection control in different hospitals, the physician was surprised that a particular kind of infection seemed to be on the rise in his region. This was indicated with an alarming red arrow in the report, and the result was followed by political calls for immediate action to bring the numbers down. Surprised by the dramatic increase in infections, he decided to conduct his own analysis of the data. In this process, the physician discovered that data from the hospital he was employed at stood out with a rapid increase in infection rates. He noticed that the dates of the documented peaks were associated with two important changes in testing procedure and capacity at his hospital. During this period, the hospital had introduced a more sensitive testing procedure and increased the number of total tests, because they had taken over test analysis previously done by another lab. Without this contextual knowledge, however, the data misleadingly signaled that hospital-acquired infections were out of control.

It may be argued that the inclusion of metadata about changes in testing procedures could potentially have avoided the misunderstanding in our example—and thus the resources needed to discuss calls for action and the subsequent data work undertaken by the physician to question the findings in the report. If this is indeed the case, it would, however, only underscore the point that the resources required for repurposing of health data are substantial, as this typically requires the production of metadata—a cost that is often not accounted for in rationalized models of the benefits of reusing health data. Moreover, the availability of high-quality metadata is not always sufficient to ensure robust recontextualization of data. In this specific case, the physician was generally skeptical of the ability of non-clinical users to make sense of clinical data and emphasized the multitude of contextual factors hidden in health data: “There are so many parameters, and we only have to change a few for the numbers to change. It is therefore very difficult to say if this even reflects the underlying reality.” If these challenges arise in the context of infection prevention and control, one should not underestimate the resourced and contextual expertise required to recontextualize more complex health data.

## Discussion: foreseeing the unintended consequences of hopeful policies

The phenomenon that IT-technologies that are intended to increase productivity instead result in a productivity *decline* is not rare, nor unique to IT-systems in healthcare. What is

sometimes referred to as the Solow Paradox, or the Productivity Paradox, refers to an observation made already in 1987 by the economist Robert Solow that “You can see the computer age everywhere but in the productivity statistics” (Solow 1987, p. 36). Similarly, we have commented on how the IT-system Sundhedsplatformen was intended to improve clinical information “flows” and “productivity” but was often experienced by healthcare staff as time consuming. Importantly, our aim is not to be critical of digitalization as such, nor to dismiss the possibility that an increase in data work can be justified. Indeed, our informants recognize many benefits of digitalizing and integration of patient data, such as providing evidence-based strategies to improve quality of care and more cost-effective administration of healthcare systems for the future. The primary concern is rather the political expectation that IT-systems to facilitate multiple uses of data will be seamless and that data are already there—ready to be reused. We find that the documented challenges pose ethical concerns.

Some of the challenges we describe can be interpreted as trade-offs in the *clinical usability* of health data infrastructures designed to prioritize data repurposing (Hoeyer 2023). When optimizing the formatting and integration of data for one purpose, it often results in data friction elsewhere in the system. We have identified five types of data work related to the production, completion, validation, sorting, and recontextualization of health data. This typology partly overlaps with other studies of data work, some of which also examine data-related tasks conducted by patients and specialized data managers (Bossen et al. 2019; Fiske et al. 2019; Pine and Bossen 2020; Torenholt et al. 2020). Our ambition is not to establish an exhaustive list of types of data work, but rather to encourage more discussion of the consequences of the political aims of repurposing. We propose that the proliferation of data work in clinical practice is addressed not only as a practical problem but also as an *ethical challenge*, because it involves trade-offs in terms of prioritization of clinical resources, including the time spent with patients versus data documentation. Failing to acknowledge tradeoffs and the need to make priorities can have important consequences for patients and health professionals, such as reduced resources for patient care and occupational burnout among health professionals (Rigsrevisionen 2018; Downing et al. 2018).

One type of trade-off concerns divergence in what different users may view as “good data.” What counts as good data for health professionals and secondary users can differ, as seen in the sections on Completion and Sorting of data. Commenting on the introduction and widespread use of the Epic-system in the US, medical doctor Gewande (2018) criticizes that design choices are more politically than clinically motivated. He argues that doctors and administrators have different views on what functions and information should be prioritized. What is relevant for audit or research is not

always clinically relevant, and vice versa (see also Hoeyer and Wadmann 2020). Moreover, the standardization required for data integration and reuse sometimes conflicts with the local needs of health professionals, such as when flexibility is required to account for the iterative and temporal aspects of disease diagnostics (Winthereik 2003) or information to account for the specific patient’s narrative (Hunt et al. 2017; Wachter 2017). New IT solutions, including the Epic system examined in this paper, often place constraints on free text spaces, because non-standardized terminology is not straightforwardly machine-readable and therefore conflicts with the aim of data repurposing (Pine and Bossen 2020). We must therefore acknowledge the trade-offs documented above and in other studies reporting how the use of predefined default options can negatively affect qualitative aspects of patient care, such as the inclusion of relevant information concerning the specific circumstances of the individual patient (Fogelberg et al. 2009; Petrovskaya et al. 2009; Robichaux 2019; Siegler and Adelman 2009).

Non-standardized information can also be essential for the reuse of data, as seen in the section on Recontextualization, because it is often required to validate structured data entries and to avoid misinterpretation when data are analyzed outside the context of data production (see also Schmidt et al. 2019; Weiskopf and Weng 2012). Health data cannot be interpreted without “human input to recontextualize knowledge” (Greenhalgh et al. 2009, p. 729), as the meaning and accuracy of data need to be understood in relation to the specific circumstances of production and use. That data gain meaning only when understood in their context of production is by no means a new insight nor unique to medicine (Latour and Woolgar 1979). Even seemingly standardized data, such as genomic data in biobanks, require a “learned intermediary” to become recontextualized in new settings (Reardon 2017, p. 135). It therefore takes additional work, medical expertise, and contextual knowledge to package health data for reuse (Leonelli 2016). With the COVID-19 pandemic, challenges of recontextualization became vividly clear, even to the public, through discussions of the limitations in the comparability of data from countries with different testing procedures, age distributions, containment measures, and levels of public trust (COVID-19 National Preparedness Collaborators 2022). Still, the resources needed for proper probing of health data are rarely addressed in political reports.

The policy of *multiplication of purposes* leads to *multiplication of data work*, which affects clinical practice. An evaluation of the benefits of data repurposing must therefore also include a consideration of the costs. While digitalization and data repurposing may be particularly comprehensive in Denmark, the issues of resource-demanding data work are not confined to this setting. Similar problems of fragmented patient information and needs for double registration and

sorting of information have been described in other contexts (Ellingsen and Monteiro 2003; Sheikh et al. 2011; Gewande 2018; Pine and Bossen 2020), and many studies report on how the introduction of new IT systems to facilitate the reuse of health data increases the time spent by healthcare staff on data registration (e.g., Morrison et al. 2013; Kuhn et al. 2015; Downing et al. 2018; McVey et al. 2021). Nevertheless, the image of seamless data integration and repurposing keeps flourishing in policy reports. Though it is increasingly acknowledged by policy makers that repurposing requires resources for data curation, they typically focus on data work conducted at data repositories, thus leaving the data work conducted in clinical settings strikingly “invisible” (e.g., European Commission 2019, 2020c). What is more, some strategies envision how data curation at repositories can be minimized via the implementation of more comprehensive standardization strategies. An example is the proposal to develop pan-European standards for health data to facilitate easier data integration in the planned European Health Data Space (European Commission 2020c). Following this harmonization logic, standardized reporting is built into the infrastructure, and data do not need to be transformed or travel to be reused. However, given the challenges we have outlined, this strategy would have significant impact on data reporting in the clinic and is likely to involve substantial costs and trade-offs for the primary users in clinical settings.

We hope that our examples, and descriptions of the different types of data work, can help create awareness about possible trade-offs and resource demands to be considered in future analyses and business cases when developing strategies and infrastructures for digital healthcare systems. As a minimum, the requirement of skilled data work in clinical settings and counselling in reinterpretation must be considered in political strategies as a foreseeable cost. With this focus, we emphasize that the ethical questions to consider in relation to data repurposing should be expanded beyond the important issues of privacy, autonomy, and risk to include also issues of prioritization. Social science and medical humanities have an important role in making it possible for policy makers to balance costs and gains in a careful manner: only when the invisible work has been made visible—and brought into focus—can the costs be acknowledged and dealt with. This should be a key task for a practical ethics.

## Conclusion

Political ambitions of data repurposing currently pull medicine in many different directions, because the benefits of integrated information technology come with costs in terms of extra data work and trade-offs in usability for some users. If the current modus operandi in healthcare digitalization ignores the need for data work in clinical settings, attention

may be shifted away from patient needs and the validation of data may be undermined. We therefore propose that the trade-offs related to clinical data work should take a more prominent space in the ethical and political debates about the repurposing of health data. From the perspective of practical or empirical ethics (Hoffmaster 1992; Pols 2015), it is necessary to move close to the actual work practices and articulate the dilemmas at hand. In this article, we have pointed to examples where attempts to repurpose data drain resources from clinical care, where administrative needs consume clinical resources, and where the automation strategies can undermine data quality and validity. The analysis also suggests that the challenges do not stem from lack of investment, but rather from lack of acknowledgement of existing practices of data use in the clinic. If data integration is not just a means to enhanced efficiency, the critical question for policy makers is: which purposes should take priority? Depending on how different user needs are prioritized, there is a risk that secondary uses overrule primary ones: when and on which grounds can this be justified? Ambitions to repurpose health data raise fundamental questions about what counts as relevant information, for whom, and why. We therefore encourage a practically engaged form of ethics that can engage how to prioritize user needs and healthcare resources.

**Acknowledgements** We are grateful to our informants for sharing their experiences with us. We would like to thank two anonymous reviewers for helpful feedback on a previous version of this article.

**Funding** Sarah Wadmann and Klaus Hoeyer have received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Grant Agreement Number 682110). Green's research is supported by the Independent Research Fund Denmark (Grant Agreement No. 0132-00026B).

## Declarations

**Conflict of interest** The authors declare no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

## References

- Akrich, Madelaine. 1992. The de-scription of technical objects. In *Shaping technology/Building society*, ed. Wiebe E. Bijker and John Law, 205–224. Cambridge: MIT Press.
- Allen, Arthur. 2019. Lost in translation: Epic goes to Denmark. *Politico*. Available online: <https://www.politico.com/story/2019/06/06/epic-denmark-health-1510223>
- Arribas-Ayllon, Michael, Katie Featherstone, and Paul Atkinson. 2011. The practical ethics of genetic responsibility: Non-disclosure and the autonomy of affect. *Social Theory and Health* 9: 3–23.
- Barilan, Michael Y., and Margherite Brusa. 2022. Precision and persons in medicine. In *Can Precision Medicine be Personal; Can*

- Personalized Medicine be Precise?*, ed. Michael Y. Barilan, Margherita Brusa, and Aaron Ciechanover, 1–9. Oxford: Oxford University Press.
- Bauer, Susanne. 2014. From Administrative Infrastructure to Biomedical Resource: Danish Population Registries, the “Scandinavian Laboratory”, and the “Epidemiologist’s Dream.” *Science in Context* 27 (2): 187–213.
- Bentzon, Niels and Jacob Rosenberg. 2021. *Destruktiv digitalisering – En debatbog om Sundhedsplatformen 2016–2021*. København: FADL’s Forlag.
- Birch, Kean, D.T. Cochrane, and Callum Ward. 2021. Data as asset? The measurement, governance, and valuation of digital personal data by Big Tech. *Big Data & Society* 8 (1): 205395172111017308.
- Bonde, Morten, Claus Bossen, and Peter Danholt. 2019. Data-work and friction: Investigating the practices of repurposing health-care data. *Health Informatics Journal* 25 (3): 558–566.
- Boniolo, Giovanni. 2022. The problematic side of precision medicine. In: *Can Precision Medicine be Personal; Can Personalized Medicine be Precise?* Eds. Barilan, Michael Y., Brusa, Margherita, and Aaron Ciechanover, pp. 48–59. Oxford University Press.
- Bossen, Claus. 2011. Accounting and co-constructing: The development of a standard for electronic health records. *Computer Supported Cooperative Work (CSCW)* 20 (6): 473–495.
- Bossen, Claus, Yunan Chen, and Kathleen H. Pine. 2019. The emergence of new data work occupations in healthcare: The case of medical scribes. *International Journal of Medical Informatics* 123: 76–83.
- Bowker, Geoffrey C. and Susan Leigh Star. 1999. *Sorting things out. Classification and its consequences*. Cambridge, MA: MIT Press.
- COVID-19 National Preparedness Collaborators. 2022. Pandemic preparedness and COVID-19: An exploratory analysis of infection and fatality rates, and contextual factors associated with preparedness in 177 countries, from Jan 1, 2020, to Sept 30, 2021. *The Lancet* 399: 1489–1512.
- Danske Regioner. 2015. *Handlingsplan for Personlig Medicin*. København: Danske Regioner.
- Danske Regioner and Dansk Industri (2019). *HealthTech:DK. Danmark som førende HealthTech-nation*. Copenhagen, Denmark.
- Dawes, Daniel E. 2020. *The Political Determinants of Health*. Johns Hopkins University Press.
- Digitaliseringspartnerskabet. 2021. *Visioner og anbefalinger til Danmark som et digitalt foregangsland*. København: Digitaliseringsstyrelsen. Available online: <https://fm.dk/udgivelser/2021/oktober/visioner-og-anbefalinger-til-danmark-som-et-digitalt-foregangsland/>
- Downing, Lance N., David W. Bates, and Christopher A. Longhurst. 2018. Physician burnout in the electronic health record era: Are we ignoring the real cause? *Annals of Internal Medicine* 169: 50–51.
- Drachman, Hans, and Hans Davidsen-Nielsen. 2018. Regionsledelse så stort på alle advarsler om it system. *Politiken* June 20, 2018.
- Edwards, Paul N. 2010. *A vast machine: Computer models, climate data, and the politics of global warming*. Cambridge: MIT Press.
- Edwards, Paul N., Matthew S. Mayernik, Archer L. Batcheller, Geoffrey C. Bowker, and Christine L. Borgman. 2011. Science friction: Data, metadata, and collaboration. *Social Studies of Science* 41 (5): 667–690.
- Ellingsen, Gunnar, and Eric Monteiro. 2003. A patchwork planet: Integration and cooperation in hospitals. *Computer Supported Cooperative Work* 12: 71–95.
- European Commission. 2010. *Riding the wave: How Europe can gain from the rising tide of scientific data*. Final report of the High Level Expert Group on Scientific Data.
- European Commission. 2016. *Open innovation, open science, open to the world a vision for Europe*. Luxembourg: Publications Office of the European Union.
- European Commission. 2019. *European data strategy*. Available online. [https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/european-data-strategy\\_en](https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/european-data-strategy_en)
- European Commission. 2020a. *Europe’s moment: Repair and prepare for the next generation*. Brussels: The European Commission.
- European Commission. 2020b. *A European strategy for data. Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions*. Brussels: The European Commission.
- European Commission. 2020c. Proposal for a regulation of the European Parliament and of the Council on European data governance (Data Governance Act). Available at: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52020bPC0767>
- Fiander, Michelle, Jessie McGowan, Roland Grad, Pierre Pluye, Karin Hannes, Michel Labrecque, Nia W. Roberts, Douglas M. Salzwedel, Vivian Welch, and Peter Tugwell. 2015. Interventions to increase the use of electronic health information by health-care practitioners to improve clinical practice and patient outcomes. *Cochrane Database of Systematic Reviews* 3: CD004749. <https://doi.org/10.1002/14651858.CD004749.pub3>.
- Fiske, Amelia, Barbara Prainsack, and Alena Buyx. 2019. Data work: Meaning-making in the era of data-rich medicine. *Journal of Medical Internet Research* 21 (7): e11672.
- Fogelberg, M. Dahm., and Barbro Wadensten. 2009. Nurses’ experiences of and opinions about using standardized care plans in electronic health records. *Studies in Health Technology and Informatics* 146: 763–764.
- Frank, Lone. 2000. When an entire country is a cohort. *Science* 287 (5462): 2398–2399.
- From, Gustav, Martin Sølvkjær, and Jørgen. Bansler. 2019. Brug af copy and paste i elektroniske patientjournaler. *Ugeskrift for Læger* 181: V12180848.
- Gabrielsen, Ane M. 2020. Openness and trust in data-intensive science: The case of biocuration. *Medicine, Health Care and Philosophy* 23 (3): 497–504.
- Gewande, Atul. 2018. Why doctors hate their computers. *The New Yorker, Annals of Medicine*, November 12, 2018.
- Greenhalgh, Trisha, Henry WW. Potts, Geoff Wong, Pippa Bark, and Deborah Swinglehurst. 2009. Tensions and paradoxes in electronic patient record research: A systematic literature review using the meta-narrative method. *The Milbank Quarterly* 87 (4): 729–788.
- Hillersdal, Line, and Mette N. Svendsen. 2022. Cancer Currencies: Making and Marketing Resources in a First-in-Human Drug Trial in Denmark. In *Precision Oncology and Cancer Biomarkers. Issues at Stake and Matters of Concern*, eds. Bremer, A., and Strand, R. pp. 45–60. Cham: Springer.
- Hoeyer, Klaus. 2023. *Data Paradoxes. The Politics of Intensified Data Sourcing in Contemporary Healthcare*. MIT Press: <https://mitpress.mit.edu/9780262545419/data-paradoxes/>
- Hoeyer, Klaus, Susanne Bauer, and Martyn Pickersgill. 2019. Datafication and accountability in public health: Introduction to a special issue. *Social Studies of Science* 49 (4): 459–475.
- Hoeyer, Klaus, and Sarah Wadmann. 2020. ‘Meaningless work’: How the datafication of health reconfigures knowledge about work and erodes professional judgement. *Economy and Society* 49 (3): 433–454.
- Hogle, Linda F. 2016. Data-intensive resourcing in healthcare. *BioSocieties* 11 (3): 372–393.
- Hoffmaster, Barry. 1992. Can ethnography save the life of medical ethics? *Social Science and Medicine* 35: 1421–1431.

- Holt, Jette. 2020. *At gøre ting med tal: En nexusanalytisk afdækning af overvågningssystemet HAIBAS diskursive aflæsning*. Aalborg Universitetsforlag. Aalborg Universitet. Det Humanistiske Fakultet. Ph.D.-Serien. Available online: [https://vbn.aau.dk/ws/portalfiles/portal/404967144/PHD\\_JH\\_E\\_pdf.pdf](https://vbn.aau.dk/ws/portalfiles/portal/404967144/PHD_JH_E_pdf.pdf)
- Hunt, Linda M., Hannah S. Bell, Allison M. Baker, and Heather A. Howard. 2017. Electronic health records and the disappearing patient. *Medical Anthropology Quarterly* 31 (3): 403–421.
- Kruse, Clemens Scott, Rishi Goswamy, Yesha Jayendrakumar Raval, and Sarah Marawi. 2016. Challenges and opportunities of big data in health care: A systematic review. *JMIR Medical Informatics* 4 (4): e5359.
- Kuhn, Thomson, Peter Basch, Michael Barr, and Thomas Yackel. 2015. Clinical documentation in the 21st century: Executive summary of a policy position paper from the American College of Physicians. *Annals of Internal Medicine* 162 (4): 301–303.
- Latour, Bruno, and Steve Woolgar. 1979. *Laboratory life: The construction of scientific facts*. Princeton, NJ: Princeton University Press.
- Leonelli, Sabina. 2014. What difference does quantity make? On the epistemology of Big Data in biology. *Big Data & Society*. <https://doi.org/10.1177/2053951714534395>.
- Leonelli, Sabina. 2016. *Data-centric biology: A philosophical study*. Chicago: University of Chicago Press.
- Leonelli, Sabina, and Niccolò Tempini, eds. 2020. *Data journeys in the sciences*. New York: Springer.
- Madden, Raymond. 2010. *Being ethnographic. A guide to the theory and practice of ethnography*. London: Sage.
- Markus, Lynne M. 2001. Towards a theory of knowledge reuse: Types of knowledge reuse situations and factors in reuse success. *Journal of Management Information Systems* 18 (1): 57–93.
- McVey, Lynn, Natasha Alvarado, Joanne Greenhalgh, Mai Elshehaly, Chris P. Gale, Julia Lake, Roy A. Ruddle, et al. 2021. Hidden labour: The skillful work of clinical audit data collection and its implications for secondary use of data via integrated health IT. *BMC Health Services Research* 21 (1): 1–11.
- Ministry of Foreign Affairs. 2014. Invest in Denmark. Start with Denmark - The heart of life sciences for research and business. Copenhagen. Available online: <https://www.regioner.dk/media/3759/270916-startwithdenmark2016-fullreport.pdf>
- Ministry of Health and Danish Regions. 2016. *Personalized medicine for the benefit of patients. Clear diagnosis, targeted treatment, strengthened research. National Strategy for personalized medicine 2017–2010*. Copenhagen, Denmark. Available online: <https://www.regioner.dk/sundhed/medicin/personlig-medicin>
- Ministry of Health. 2017. Healthcare in Denmark. An overview. Available online: [https://sum.dk/English/~media/Filer%20-%20Publikationer\\_i\\_pdf/2016/Healthcare-in-dk-16-dec/Healthcare-engli sh-V16-dec.ashx](https://sum.dk/English/~media/Filer%20-%20Publikationer_i_pdf/2016/Healthcare-in-dk-16-dec/Healthcare-engli sh-V16-dec.ashx)
- Mittelstadt, Brent D., and Luciano Floridi. 2016. The ethics of Big Data: Current and foreseeable issues in biomedical contexts. *Science and Engineering Ethics* 22: 303–341.
- Morrison, Cecily, Matthew Jones, Rachel Jones, and Alain Vuylsteke. 2013. ‘You can’t just hit a button’: An ethnographic study of strategies to repurpose data from advanced clinical information systems for clinical process improvement. *BMC Medicine* 11 (1): 103.
- National Academy of Sciences (NAS). 2011. *Toward precision medicine. Building a knowledge-network for biomedical research and a new taxonomy of disease*. In: *Committee on A Framework for Developing a New Taxonomy of Disease*. The National Academies Press, Washington, DC.
- Nordforsk. 2014. Joint Nordic Registers and Biobanks – A goldmine for health and welfare research. Report by NORIA-net on Registers and Biobanks (NRB), available online: <https://www.nordforsk.org/2014/joint-nordic-registers-and-biobanks-goldmine-health-and-welfare-research>
- OECD. 2013. *ICTs and the health sector. Towards smarter health and wellness models*. Paris: OECD Publishing. <https://doi.org/10.1787/9789264202863-en>.
- Overlægerådet i Region Hovedstadens Psykiatri. 2018. *Rapport om Sundhedsplatformen og kvaliteten af journalføringen. En spørgeskemaundersøgelse om psykiatriske overlægers erfaringer efter 18 måneders brug*. København.
- Petersen, Alan. 2019. *Digital health and technological promise. A sociological inquiry*. New York, NY: Routledge.
- Petrovskaya, Olga, Marjorie McIntyre, and Carol McDonald. 2009. Dilemmas, tetralemmas, reimagining the electronic health record. *Advances in Nursing Science* 32 (3): 241–251.
- Pine, Kathleen H., and Claus Bossen. 2020. Good organizational reasons for better medical records: The data work of clinical documentation integrity specialists. *Big Data & Society*. <https://doi.org/10.1177/2053951720965616>.
- Pinel, Clémence. 2021. Renting valuable assets: Knowledge and value production in Academic Science. *Science, Technology, & Human Values* 46 (2): 275–297.
- Pols, Jeanette. 2015. Towards an empirical ethics in care: relations with technologies in health care. *Medicine Health Care and Philosophy* 18: 81–90.
- Reardon, Jenny. 2017. *The Postgenomic Condition*. University of Chicago Press.
- Regeringen. 2021. Strategy for life science. Copenhagen, Denmark. <https://em.dk/media/14179/strategi-for-life-science.pdf>
- Regeringen. 2022. *Danmarks digitaliseringsstrategi. Sammen om den digitale udvikling*. København: Finansministeriet.
- Rigsrevisionen. 2018. *17/2017. Rigsrevisionens beretning om Sundhedsplatformen afgivet til Folketinget med Statsrevisorernes bemærkninger*. Copenhagen, Denmark.
- Robichaux, Catherine, Mari Tietze, Felicia Stokes, and Susan McBride. 2019. Reconceptualizing the electronic health record for a new decade: A caring technology? *Advances in Nursing Science* 42 (3): 193–205.
- Schmidt, Morten, Sigrun Alba Johannesdóttir, Schmidt, Kasper Adalborg, Jens Sundbøll, Kristina Laugesen, Vera Ehrenstein, and Henrik Toft Sørensen. 2019. The Danish health care system and epidemiological research: From health care contacts to database records. *Clinical Epidemiology* 11: 563–591.
- Siegler, Eugenia L., and Ronald Adelman. 2009. Copy and paste: A remediable hazard of electronic health records. *The American Journal of Medicine* 122 (6): 495–496.
- Solow, Robert. 1987. We’d better watch out. *New York Times Book Review*, July 12, 36.
- Star, Susan Leigh. 1991. The Sociology of the Invisible: The Primacy of Work in the Writings of Anselm Strauss. In *Social Organization and Social Process. Essays in Honor of Anselm Strauss*, ed. David R. Maines, 92–97. New York: Aldine De Gruyter.
- Sundheds- og Ældreministeriet, Finansministeriet, Danske Regioner og KL. 2018. Et sikkert og sammenhængende sundhedsnetværk for alle. Strategi for Digital Sundhed 2018–2022. Copenhagen, Denmark.
- Tarkkala, Heta, Ilpo Helén, and Karoliina Snell. 2019. From health to wealth: The future of personalized medicine in the making. *Futures* 109: 142–152.
- Tempini, Niccolò. 2020. Data curation-research: Practices of data standardization and exploration in a precision medicine database. *New Genetics and Society*. <https://doi.org/10.1080/14636778.2020.1853513>.
- Tiirikainen, Morten and Mikkel Rasmussen. 2021. Først blev lægesekretærene fyret – nu vil politikere ansætte flere. TV2 lorry, Oct. 11, 2021.

- Timmermans, Stefan, and Marc Berg. 2003. *The gold standard. The challenge of evidence-based medicine*. Philadelphia: Temple University Press.
- Torenholt, Rikke, Lena Saltbæk, and Henriette Langstrup. 2020. Patient data work: Filtering and sensing patient-reported outcomes. *Sociology of Health & Illness* 42 (6): 1379–1393.
- Tupasela, Aaro. 2021. *Populations as Brands: Marketing National Resources for Global Data Markets*. New York: Springer.
- UK Department of Health. 2011. Liberating the NHS. An Information Revolution. Available online: [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/216664/dh\\_129580.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/216664/dh_129580.pdf)
- United Nations. 2020. *E-Government Survey 2020. Gearing E-Government to Support Transformation Towards Sustainable and Resilient Societies*. New York.
- Vezyridis, Paraskevas, and Stephen Timmons. 2017. Understanding the care.data conundrum: New information flows for economic growth. *Big Data & Society* 4 (1): 2053951716688490.
- Vezyridis, Paraskevas, and Stephen Timmons. 2021. E-Infrastructures and the divergent assetization of public health data: Expectations, uncertainties, and asymmetries. *Social Studies of Science* 51 (4): 606–627.
- Vikkelsø, Signe. 2005. Subtle Redistribution of Work, Attention and Risks: Electronic Patient Records and Organisational Consequences. *Scandinavian Journal of Information Systems* 17: 3–30.
- Wachter, Robert. 2017. *The digital doctor. Hope, hype, and harm at the dawn of medicine's computer age*. New York: McGraw-Hill Education.
- Wadmann, Sarah, and Klaus Hoeyer. 2018. Dangers of the digital fit: Rethinking seamlessness and social sustainability in data-intensive healthcare. *Big Data & Society*. <https://doi.org/10.1177/2053951717752964>.
- Wang, Michael D., Raman Khanna, and Nader Najafi. 2017. Characterizing the source of text in electronic health record progress notes. *JAMA Internal Medicine* 177 (8): 1212–1213.
- Wiener, Carolyn L. 2000. *The Elusive Quest. Accountability in Hospitals*. New York: Aldine De Gruyter.
- Weis, Justin M., and Paul C. Levy. 2014. Copy, paste, and cloned notes in electronic health records. *Chest* 145 (3): 632–638.
- Weiskopf, Nicole Gray, and Chunhua Weng. 2012. Methods and dimensions of electronic health record data quality assessment: Enabling reuse for clinical research. *Journal of the American Medical Informatics Association* 20 (1): 144–151.
- Winthereik, Brit R. 2003. “We fill in our working understanding”: On codes, classifications and the production of accurate data. *Methods of Information in Medicine* 42 (04): 489–496.
- Winthereik, Brit R., Irma Van Der Ploeg, and Marc Berg. 2007. The electronic patient record as a meaningful audit tool: Accountability and autonomy in general practitioner work. *Science, Technology, & Human Values* 32 (1): 6–25.
- Zuboff, Shoshana. 1988. *In the Age of the Smart Machine: The Future of Work and Power*. New York: Basic Books Inc.
- Zuboff, Shosana. 2019. *The Age of Surveillance Capitalism: The Fight for A Human Future at the new Frontier of Power*. London: Profile Books.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.