Health Data Ecosystems
Contested Valuations in Denmark

John Mark Burnett
IT University of Copenhagen
Business IT, Year 2022

Abstract

Personal health data – in its contested definitional space – comes with significant individual and societal promise in a time of computational advance. With lofty promises of healthy citizens and national economies this narrative is attracting significant investment in the EU as the region looks to overcome a bleak prognosis of population ageing and become a geopolitical force in the data economy. Unleashing the promise of personal health data is however not straightforward as it implies an intensified use of what GDPR recognises as ‘sensitive’. Data-sharing initiatives in the UK and other nations have moreover shown how the transition from promise to controversy can be slim when claims to public values are contested. How to put personal health data to use responsibly, and who decides, is in this respect of imperative importance to the promise of data and yet currently uncertain. This doctoral study attends to this promissory space of uncertainty through the following research question: How is data-intensification sought and brought about by infrastructural initiatives in the Danish ‘health data ecosystem’, and what tensions arise in the pursuit of valuations assigned to personal health data?

The ‘health data ecosystem’ is a conceptual and heuristic device I use to frame the field I study and the contributions that follow. It foregrounds the field as a moving, living target that evolves in relation to movements in the data sources, capabilities and stakeholders. What follows from this conceptualisation is a need for empirical research to keep debates, policies and regulations close to concurrent practices and tensions. Guiding this research is a theoretical framework that focuses on the infrastructures, life cycles and liveliness of data through an interdisciplinary body of scholarship, bringing sciences and technology studies (STS) and adjacent fields into conversation with ecology. The thesis argues that, as the health data ecosystem evolves, so do the relation between citizen, state and market, which in turn engenders new questions of societal and political importance. In Denmark, these movements are partly becoming visible where tech actors are looking to repurpose the national health data registries otherwise known as a ‘scientific goldmine’, which are examined more closely through infrastructural case studies.

The research question is examined through two core case studies based on ethnographic fieldwork conducted between 2017 and 2021 through interviews, participatory observations, document analysis, and digital methods. The first case is The Danish General Practice Database (Danish: *Dansk AlmenMedicinsk Database)* that following its national implementation in 2007 became a central infrastructural node in Danish healthcare before its collapse in 2015. The second case focuses on the rise of data intermediaries that are most closely examined through the infrastructural initiatives of Data for Good Foundation and Digi.Me. Whereas the former case represents a traditional model of ‘state-centred’ data-intensification in Denmark, the latter case proposes an alternative model that is by contrast centred on ideas of ‘individual control’. Seen together, the two cases provide a mix of post-mortem and in-the-making examinations of infrastructures that either claim to provide a sustainable model for data-intensification or perished in their attempt to build one.

The thesis describes what the aforementioned movements look like and interrogate the questions they enkindle from a Danish perspective. Overall, the articles make a contribution by studying the *life cycle of data within the emerging ecosystem*, from its collection to its use and deletion. The articles deliver empirical, conceptual and methodological contribution for understanding efforts to make use of personal health data in Denmark. Empirically, through my two case studies, I analyse the failure of one infrastructure on the grounds of ethical and re-use concerns, and how claims to ethical data sharing also have an infrastructural dimension that is under-exposed in the literature. Conceptually, I enrich conversations that pertain to how data is thought of and managed, from the deletion and retention of data to its 'call'. Finally, my methodological contribution, with colleagues, is to scholarship on industry conferences, reviewing different modes of participant observation as a means of studying the role of conferencing. The thesis engages conversations on the social sustainability of data practices to which it provides an analytical and empirical contribution. What this contribution points to is that data practices need to be cultivated with care, which I concludingly suggest should entail social and environmental considerations of sustainability.

Resumé

I en samtid fyldt med digitalisering er personlige sundhedsdata blevet associeret med betydelige individuelle og samfundsmæssige løfter. Disse løfter om sunde borgere og voksende nationale økonomier udgør et narrativ der tiltrækker signifikante investeringer i EU-området, i forlængelse af regionens forsøg på at imødekomme udfordringerne ved en aldrende befolkning og i bestræbelsen på at blive en geopolitisk magt i data-økonomien. At forløse de løfter som der rejses omkring personlig sundhedsdata er dog ikke en ligetil opgave, da det implicerer en intensiveret brug af lige præcis de data som GDPR betegner som 'følsom'. Datadelingsinitiativer i Storbritannien og andre lande har desuden vist hvordan overgangen fra løfte til kontrovers let kan ske, særligt når initiativers værdiskabelse udfordres. Hvordan personlige sundhedsdata kan bruges ansvarligt, og hvem der bestemmer dette, er i denne henseende essentielt i forhold til de løfter og forhåbninger som data rejser, men som endnu er usikkert. Denne PhD-afhandling undersøger dette rum af løfter og usikkerhed igennem følgende forskningsspørgsmål: Hvordan udsøges og gennemføres dataintensivering af det danske 'økosystem for sundhedsdata' igennem infrastrukturelle initiativer, og hvilke spændinger opstår der i forsøget på at værdisætte personlige sundhedsdata?

'Økosystem for sundhedsdata' er et konceptuelt og heuristisk udtryk som jeg bruger til at indkredse det felt jeg studerer og afhandlingens bidrag. Det fremhæver feltet som levende og dynamisk, og noget der udvikler sig i relation til bevægelser i datakilder, kapaciteter og interessenter. Fra denne konceptualisering følger det, at der er behov for empirisk forskning, for at sørge for at debatter, politikker og regulering sker tæt på de faktisk praksisser og spændinger. Afgørende for forskningen som denne afhandling bidrager med, har været et teoretisk rammeværk med fokus på de infrastrukturer, livscyklusser, og dynamikker der kendetegner data, gennem en interdisciplinær forskningslitteratur, der bringer Videnskabs- og Teknologistudier (Science and Technology Studies) og tilstødende felter i samtale med økologi. Afhandlingen argumenterer for at i takt med at sundhedsdata økosystemet udvikler sig, gør relationen mellem borgeren, staten og markedet det også, hvilket skaber nye spørgsmål af samfundsmæssig og politisk betydning. I Danmark er disse forandringer ved at blive delvist synlige, i de tilfælde hvor 'tech'-aktører forsøger at omdefinere formålet ved de nationale sundhedsdataregistre der ellers primært er kendt som 'videnskabelige guldminer.' Dette undersøges nærmere igennem infrastrukturelle case studier.

Afhandlingens forskningsspørgsmål undersøges gennem to centrale cases, baseret på etnografisk feltarbejde udført imellem 2017 og 2021 via interviews, deltagerobservation, dokumentanalyse og digitale metoder. Den første case omhandler Dansk AlmenMedicinsk Database, en database der blev en central infrastruktur i det danske sundhedsvæsen efter dens implementering i 2007, før den så kollapsede i 2015. Den anden case fokuserer på fremkomsten af 'dataformidlere', der undersøges ved at undersøge de infrastrukturelle tiltag af Data for Good Foundation og Digi.Me. Hvor den førstnævnte case repræsenterer en traditionel model af 'stats-baseret' dataintensivering, så repræsenterer den sidstnævnte case en alternativ model, der er centreret på en idé om 'individuel kontrol.' De to cases er begge undersøgelser af infrastrukturer, hvor den ene er et post-mortem af et mislykkedes forsøg på

dataintensivering, mens den anden er en diagnostik af et samtidigt forsøg på at bygge en ny, bæredygtig model for at gøre det.

Afhandlingen beskriver hvordan de førnævnte bevægelser ser ud, og udforsker de spørgsmål de rejser fra et dansk perspektiv. Artiklerne i afhandlingen repræsenterer et videnskabeligt bidrag ved at studere den *livscyklus data har i de emergerende økosystemer*, fra dens indsamling til dens brug og sletning. Artiklerne yder empiriske, konceptuelle og metodologiske bidrag til forståelsen af forsøg på at nyttiggøre personlige sundhedsdata i Danmark. Empirisk bidrager jeg via mine to cases en analyse af hvordan bekymringer om etik og data genanvendelse førte til lukningen af én infrastruktur, og en analyse af hvordan projekter med fokus på etisk datadeling også har infrastrukturelle dimensioner, der er underbelyst i litteraturen. Konceptuelt bidrager jeg til diskussioner om hvordan data opfattes og styres, ved at berøre emner fra sletning og vedligehold af data til dets 'kald.' Endelig bidrager jeg metodologisk, i samarbejde med kollegaer, til studier af industrikonferencer, hvor vi skaber overblik over forskellige måder at lave deltagerobservation i teknologi-konferencer. Afhandlingen engagerer sig med diskussioner om datapraksisser sociale bæredygtighed, hvilket den foretager analytiske og empiriske bidrag til. Dette bidrag peger på at datapraksisser skal formes med omsorg, hvilket jeg afslutningsvis påpeger bør indebære hensyn til social og miljømæssig bæredygtighed.

Overview of articles and publication status

No.	Title, author(s), journal	Status
I	**"Let's make it happen!":** An STS ethnographer's guide to tech conferences Douglas-Jones R, Burnett J, Fritsch E and Hockenhull M Science & Technology Studies	Revise and re-submit
II	**Sentinel vision:** Data collection, disease prevention and professionals in Danish diabetes healthcare Burnett J and Douglas-Jones R Science Technology & Human Values	Pre-submission
III	**The call of data:** A post-mortem of the Danish General Medicine Database Burnett J Big Data & Society	Accepted
IV	**Refusing retention:** Health data and its deletion in the Danish welfare state Burnett J and Douglas-Jones R Social Studies of Science	Accepted
V	**Infrastructures in the promotion of ethical tech:** The rise of data intermediaries in the Nordic market for digital health Burnett J Big Data & Society	Pre-submission

Table of content

PART 1

1. Introduction: The promise, tension and uncertainty of personal health data

Personal health data (PHD) – in its contested definitional space – comes with significant individual and societal promise in a time of computational advance. With lofty promises of healthy citizens and national economies this narrative is attracting significant investment in the EU (European Commission, 2020; The Danish Ministry of Health, 2018; see also OECD 2015) as the region looks to overcome a bleak prognosis of population ageing while asserting itself as a geopolitical force in the data economy. Unleashing the promise of PHD is however not straightforward as it implies an intensified use of what the European General Data Protection Regulation (GDPR hereafter) considered 'sensitive' (European Commission, n.d.). Data-sharing initiatives in the UK (e.g. Vezyridis and Timmons, 2017; Sterckx et al., 2016; Powles and Hodson, 2017), the Netherlands (Smits, 2013) and other nations (e.g. Garrety et al., 2014) have moreover shown how the transition from promise to controversy can be slim when claims to public values are contested. How to put PHD to use responsibly, and who decides, is in this respect of imperative importance to the promise of data and yet currently uncertain. It is within this space of uncertainty that the motivation and contribution of this thesis is located.

This doctoral study attends to this uncertainty by conducting empirical research to describe how infrastructural data-sharing initiatives navigate the tensions, values and politics that inhabit this space in Denmark. With a scholarly base in science and technology studies (STS), critical data studies (CDS) and digital health, the study entails a mix of in-the-making and post-mortem examinations of infrastructures that either claim to be able to navigate this uncertainty or died trying. Two infrastructural approaches can be delineated that I refer to as 'state-centred' and 'citizen-centred': The first approach is traditional in the sense that it centres on the state-funded healthcare system for the collection and management of data while these tasks are in the second approach largely centred on the citizen. The empirical material was collected between 2017 and 2020 through a range of ethnographic methods. In terms of the contribution, the thesis engages conversations in STS and CDS where it provides sociotechnical insights on the life of data, from their collection through (re)use to deletion.

Expanding upon the brief summary above, what comes next in this introduction is a more detailed overview of the dissertation conveyed in four steps. First, I situate the doctoral study within the research project from which it originates and outlines the chosen theoretical framework. The second step explains why Denmark, in this specific moment in time, is a relevant site for exploring the health data ecosystems and specifies the main research question that the dissertation sets out to answer. The third step introduces two case studies that are interrogated to answer the research questions. The fourth and final step establishes significance and aim of the dissertation, finishing with a roadmap on how the dissertation will progress through the subsequent sections to the conclusion.

1.2 An infrastructural study of the health data ecosystem

The thesis written in this doctoral study is part of a research project called *Data as Relation: Governance in an Age of Big data, s*ponsored by the Velux foundation. The formulated aim was to craft ethnographic knowledge about the emergence of big data in the government practice of Denmark known as a front-runner in public sector digitisation. As the subtitle also signals, one could say that the research project sought to take stock of a particular 'data moment' (Maguire et al., 2020) by engaging a trend that is sometimes used to characterise the age of contemporary society. With sub-projects in tax, citizens centres, data centres, humanities faculties and health, the research project also puts forth an additional claim that conversations are needed across different sectors.

The timing of the project was according to the project proposal not insignificant, as it provided an entry "when the 'black box' of public sector digitization is still open, and where routine usage of new data sources has not yet been stabilized" (Wintereik and Gad, 2015: 1). While it should be noted that the novelty of "big data" as a material phenomenon is contestable (e.g. Hacking, 1991; Beer, 2016), the notion itself comes with a particular discourse (Beer, 2016) and ideology (Maguire et al., 2020) in a moment of computational advance. With a theoretical base in infrastructure studies, the research project took the following as its hypothesis: 'that digitization is not simply about rewiring and technologically upgrading government institutions, but about reinventing society' (Wintereik and Gad, 2015: 1)[1].

Central to the work in this thesis are burgeoning conversations on the evolving nature of the contemporary 'health data ecosystem' (Vayena and Glasser, 2016; 2017; Vayena and Blasimme, 2018; Vayena, Dzenowagis, Brownstein; Sharon and Lucivero, 2019). While the metaphor of an 'ecosystem' also appears as an emic term in the field I study, it is its etic use that I introduce here. When scholars talk about the 'health data ecosystem' it is generally used as a device to describe and discuss the shifting terrains for the circulation of health data. These shifts are with minor variations predominantly categorised according to three interdependent elements: data sources, stakeholders and technological capabilities. While the conversation draws on different fields of research, they all contribute to novel understandings of what counts as 'health data' (Vayena and Gasser, 2016), by whom they are valued (e.g. Sharon, 2016; Sharon, 2018; Riso et al., 2017) and for what purposes (e.g. Blasimme, Vayena and Hoyweghen, 2019; see also McFall, Meyers and Hoyweghen, 2020). Within the wider project from which this PhD emerges, I take the question of what will count as health data, and how it is infrastructured, to be a site of a rise of a 'new canonical infrastructure', and with my colleagues, assert that their growing pervasiveness 'coincide[s with the coming into being of new types of citizens' (Winthereik and Gad, 2015: 2). What I take from these conversations is the imagery of an ecosystem in flux, which I use to conceptualise the field I study as a moving, living object. But how can one study something that is in constant flux and therefore essentially a moving target? I address this question as both one of method and contemporary research practice, through this Introduction.

---

[1] Looking beyond the proposal, a similar argument has been proposed in relation to the ongoing platformisation of society: 'Platforms do not reflect the social: they produce the social structures we live in (Couldry and Hepp 2016)' (Van Dijck, Poell and De Waal 2018: 24).

Taking an ecosystem to be in flux, it follows that there is a need for further empirical research since our understanding of the social and political life of health data must evolve in tandem with its ecosystem. This knowledge production has a vital role in keeping debates, policies and regulations close to concurrent practices and tensions. Scholars writing from a bioethicist tradition of research tend to stay close to a philosophical level of abstraction where the empirical seldom exceeds the form of examples (e.g. Vayena and Blasimme, 2018). On the other side of the spectrum are scholars who write from an ethnographic tradition where the empirical is used to illustrate but also *drive* the analysis (e.g. Lehtiniemi and Ruckenstein, 2019). With a commitment to descriptive sociology (Latour, 2005), this dissertation is aligned with the latter approach to explore the contemporary movements in the health data ecosystem of the Danish welfare state. This undertaking is conducted through ethnographic case studies of infrastructural data sharing initiatives where the infrastructural focus is chosen for two reasons. As stated by Knox and Nafus, infrastructures are sociomaterial constructs that shape the conditions of possibilities for data saturated worlds (Knox and Nafus, 2018). Infrastructures are thus a *part* of the ecosystem as they are built to marshal and manage the life cycle of health data with new data sources, stakeholders and technological capabilities. If infrastructures are part of the ecosystemic shifts then they can also be a *means* to study the ecosystem itself. Building on this logic, the ethnographic case studies of infrastructuring are then a way to explore emerging promises, practices and politics of broader ecosystems through situated accounts. This research design however beckons the question why Denmark is a relevant site for exploring the health data ecosystems in this specific moment in time?

1.2 Danish health data: From a “Science Gold Mine” to “Denmark’s Greatest Innovation Potential”

While the idea of an *evolving* ecosystem may at first glance seem abstract, I experienced it to be fairly concrete during the first year of my fieldwork observations. Through the events I attended, interviews and desk research, I have encountered a range of metaphors that were all used to convey a perceived valuation of the Danish health data registers. While the metaphors are often not very original, the interests appended to them indicate a possible shift and something specific about the current ‘data moment’ (Magure, Langstrup and Danholt et al., 2020). This data moment has been characterised in my field as emergent and responsive to the General Data Protection Regulation (hereafter GDPR, see Marelli, Lievevrouw and Van Hoyweghen, 2020; Starkbaum and Felt, 2019) and the kind of environment for data more recently articulated through the *European Strategy for Data* (2020).

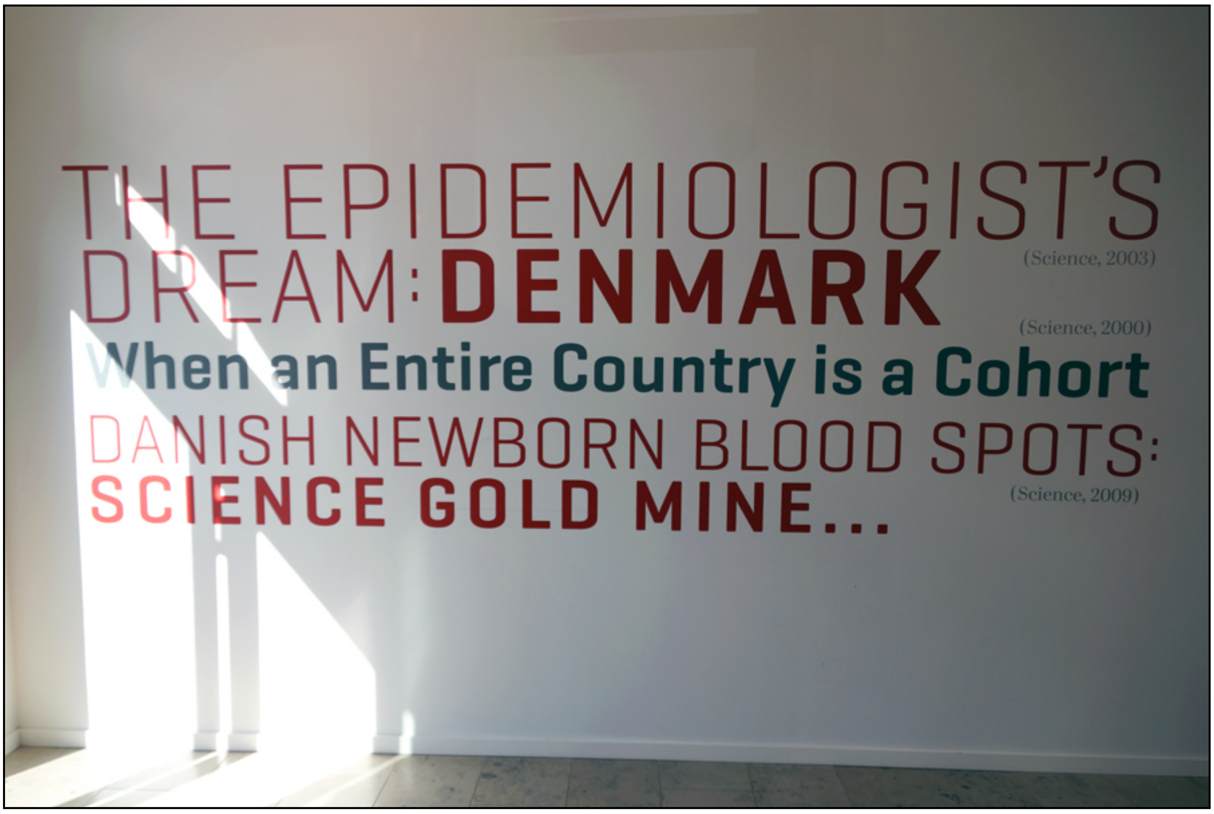


Figure 1. Photograph of the wall print in the Danish National Biobank taken from the cited field report of the Biomaterial Banks Working Group *Technology, Methods and Infrastructure for Networked Medical Research*

One of the more durable examples involves the metaphor of a "gold mine", which has traditionally been invoked to convey the *scientific* value of the Danish data registers (Bauer, 2014; Tupasela, Snell and Tarkkala, 2020; Tupasela, 2021). One of the more striking examples of this value articulation was encountered during the first year of my fieldwork that has also been noticed by STS scholar Aaro Tupasela (2021; see also Burnett, 2018). In early September 2016, the Biomaterial Banks Working Group from a German-based not-for-profit association, named *Technology, Methods and Infrastructure for Networked Medical Research*, visited the Danish National Biobank in Copenhagen. In their field report, the working group described how they were "impressed" how population wide data from "cradle to grave" was routinely registered with a personal identification number (the CPR number) that allowed "lifetime epidemiology" across medical and social registries (Technology, Methods and Infrastructure for Networked Medical Research, 2016). Among the documentation of their visit was a photograph of a wall print with the caption "Dreamland for epidemiologists". As Tupasela (2021) has noted, the photo is located by the entrance of the the Danish National Biobank that features three publication titles from the journal *Science* in large print: "The Epidemiologist's Dream: Denmark (Science, 2003)", "When an Entire Country is a Cohort (Science, 2000)", and "Danish Newborn Blood Spots: Science Gold Mine... (Science, 2009)" (See Figure 1). All publications are from the first decade of the 21st century and references the name of its renowned journal - rather than author - to establish scientific credibility. As Tupasela (2021) observes, the first two publications are authored by a Danish scientific journalist, Lone Frank. Visitors at the Danish National

Biobank who actually read the publications will learn how although 'Other Scandinavian countries have created powerful database systems', it is ‘Denmark [that] has earned a preeminent reputation for possessing the most complete and interwoven collection of statistics touching on almost every aspect of life’ (Frank, 2000: 2398).[2] The example illustrates two things. First, how national branding of populations (Tupasela, 2021) draws on narration practices that target the specific audience in places where they convene - be it in scientific journals or the entrance of the Danish National Biobank. Second, how metaphors can be an effective component in the circulation and solidification of particular narratives to for instance brand Denmark as a “Science Gold Mine”.

Figure 2. Image promoting health data as ‘Denmark’s Greatest Innovation Potential’. The image is taken from the official event invitation for the conference *More patients near welfare with health data* accessed on Facebook.

As stated, the quotations in the wall print are all from the first decade of the 21st century. More recently, existing metaphors have been recycled while new ones have emerged to rearticulate the value of Danish health data beyond science to highlight its potential for innovation and economic growth. One explicit illustration from my fieldwork was encountered at a conference entitled *More patient near welfare with health data*[3], which took place in Copenhagen during late November 2017. The conference invitation was pre-circulated online with an image of a golden egg in the foreground and geographical illustration of Denmark in the background (See Figure 2). On the golden egg, adjacent to the conference title, potential conference participants could read the following print: ‘Denmark’s Greatest Innovation Potential’.[4] With data in abundance and a highly digitised society, the Ministry of

---

[2] The reason why the Danish data registries are touted to be a ‘plum research tool’ (Frank, 2000: 2398) is however not limited to the comprehensive and longitudinal qualities of the nearly 200 databases dating as far back to the 1930s. It is the fact that population records include a personal identification number (CPR number), which enables researchers to combine data across registers for various cohort studies.

[3] Author’s translation from *Mere borgernær velfærd med sundhedsdata.*

[4] Due to a declined conference registration, my research of the event was conducted via online and requested documentation, including conference programme, presentation slides and a video

Health asserted that Denmark is favourably positioned to embrace the “digital revolution”. While the development of new health technologies was identified as a key response to the contemporary socio-economic challenges, the Minister stressed a need for *homegrown* solutions that reflect the values ingrained in the Danish welfare model. Speaking on the national economic prospects, one presenter argued that the ‘health data is not like oil - it is much better!’.

Contrary to oil, health data was (when digital) conceptualised as a non-depletable and non-rivalrous resource that would only grow, especially considering ‘exponential’ growth of private data sources and technological capabilities. Ultimately, the value of health data was argued to increase proportionally with its collection and use. It was based on this assumption that the presenter issued a call for action to intensify the collection and use of health data in Denmark. Although a similar recommendation was previously issued by OECD (2013: 19) in relation to Denmark’s ‘goldmine’ of state-collected data, the presenter argued that the “golden egg” (known from Aesop’s fable of *The goose that laid the golden eggs*) was a more fitting metaphor since it conveyed the *non-depletable* qualities of health data as a reusable resource for innovation and economic growth. One quote that speaks particularly well to the shift I observed is from STS scholar Klaus Hoeyer. Based on his own fieldwork in the Danish health sector, Hoeyer notes that ‘[...]though research uses of health data receive the most attention, research is not necessarily the main purpose with intensified data sourcing.’ (Hoeyer, 2016: 73). What this quote highlights is that the valuation of health data should not be taken for granted but elicited through the virtues of empiricism.

The two core case studies of this thesis illustrate the importance of this observation. They showcase how actors make strong advances to unleash the promise of data through infrastructural initiatives, but also how the valuations assigned to health data are vulnerable to change and contestation. With this glimpse into the research ahead, it is time to introduce the overarching research question of the present thesis:

> *How is data-intensification sought and brought about by infrastructural initiatives in the Danish ‘health data ecosystem’, and what tensions arise in the pursuit of valuations assigned to personal health data?*

This research question sets the stage for an explorative project that is empirically grounded in two core case studies that are introduced below.

1.3 Two Case Studies

The thesis is built from two core case studies, selected to highlight different phases and aspects of how a health data ecosystem is built, and explore the research question above. The case studies are empirically based on ethnographic fieldwork conducted between 2017 and 2021 through interviews, participatory observations, document analysis, and digital methods. The first case study concerns the *Danish General Practice Database* (hereafter

---

recording. According to the conference host, a representative from the national political network for the Liberal Party of Denmark (Venstre), the event was live-streamed due to an ‘overwhelming’ interest that exceeded the venue capacity in Christiansborg. The conference was also hosted by Aarhus University.

“DAMD” for *Dansk AlmenMedicine Database*) that was introduced to the Danish healthcare system in 2007 as a nationwide initiative. The aim of the initiative was to curb the socioeconomic burden of diabetes - and chronic illness - by supporting the uptake of data-intensive logics in general practice. By 2014, DAMD hosted a collection of personally identifiable data relating to 704 diagnoses on two-thirds of the Danish population from general practice. The collection of all this data in one place was unprecedented - even for a ‘research radical’ (Hoeyer, 2016: 89) nation as Denmark. During its life, the database accumulated new actors and interests to the point of collapse (Langhoff et al., 2016; Wadmann and Hoeyer, 2018). Coincidentally, the database was at one point also described as a “golden egg” by public administrators (Fischer and Tynell, 2014). This is both convenient and somewhat ironic since the case demonstrates the shortcomings of the economic model of linear valuation creation. While the infrastructure managed to facilitate the technical repurposing of data, it did so without negotiating the legitimacy of the new use cases among the relevant stakeholders (Wadmann and Hoeyer, 2018). In this case, the reuse of data was sought for conflicting interest that eventually resulted in a national controversy. Although intensified data reuse was explored as a means to extract a surplus value, it ultimately undermined the value for all stakeholders as the controversy culminated with the destruction of the database. From a sociotechnical perspective, it is neither a given that more value follows from more use nor that the data is inherently non-rivalrous and, perhaps even more surprisingly, non-depletable.

The second case attempts to take over where the first case fell short. In order to maximise the value extraction from Danish health data, the second case focuses on infrastructural initiatives that champion trending ideas of ‘individual control’ over data through different platform models (e.g. Hafen, Kossmann and Brand, 2014; Lehtiniemi T and Kortesniemi, 2017; Riso, Tupasela, Vears et al., 2017; Vayena and Blasimme, 2017; Prainsack, 2019b). During the course of my fieldwork, I have observed the rise and fall of several platform startups in Denmark - and the Nordic region more broadly. One of the central platforms - or what EU term ‘data intermediaries’ (European Commission, 2020) - in this thesis is the Danish startup *Data for Good Foundation* that sees the national health data as Denmark’s new oil adventure: ‘Just like the North Sea oil, there are plenty of jobs and economic growth in the sale of Danish health data - if we dare’ (Holm-Larsen, 2017, See Article V). Data for Good Foundation is among the select platforms that have managed to navigate the volatile space of entrepreneurialism in Denmark due to state investment from the Danish Innovation Fund.[5] In order to extract further economic value from the *New North Sea* (Holm-Larsen, 2017), Data for Good Foundation aspires to create ‘an ethically responsible and safe environment’ for intensified data re-use (‘About’, n.d.). As suggested by its title, the platform usually employs altruistic rhetoric to promote the idea of individual control where citizens are for instance encouraged to “donate” their personal data for third-party reuse. The idea of individual control is however promoted in a variety of ways that employ different metaphors and styles of rhetoric. One contrasting example is Cima Technologies, a Danish company founded around the same time as Data for Good Foundation but discontinued shortly after. In order to incentivise sharing of personal data, individuals were encouraged to *Cash in on your data gold mine* on the company website (Cima, 2017*)*. In this instance, Denmark was conceptualised as part of a *European* “gold

---

[5] Another startup is *HealthD360* that was pitched as a “public patient near data platform” at the aforementioned conference in Copenhagen by a representative of a co-hosting institution.

mine" worth €100 billion of which an average European citizen could earn €1000 a year by "selling" personal data to corporate and state actors alike. By following the work of Data for Good Foundation and its networks during my fieldwork, I sought to complement my focus on initiatives coming from within the health system, by those coming from without. Over the course of my study, startups within the health sector were given importance, not least by the European Data Strategy, published towards the end of my field research, that sees a 'regional response to the problem of 'balancing the flow and wide use of data, while preserving high privacy, security, safety and ethical standards' (EU commission, 2020: 3).

Whereas the former case represents a traditional model of 'state-centred' data-intensification in Denmark, the latter case proposes an alternative model that is by contrast centred on ideas of 'individual control'. Seen together, the two cases provide a mix of post-mortem and in-the-making examinations of infrastructures that either claim to provide a sustainable model for data-intensification or perished in their attempt to build one.

1.4 Significance

As I will seek to argue, as the health data ecosystem evolves, so does the relation between citizen, state and market. In Denmark, this is becoming visible where tech actors are looking to 'scientific goldmine' is being repurposed. And as the health data ecosystem is evolving new questions of societal and political importance emerge. This doctoral study aims to describe what these shifts look like and interrogate the questions they enkindle from a Danish perspective. Overall, I aim to make a contribution by studying the *life cycle of data within the emerging ecosystem*, from its collection to its use and deletion. Empirically, through my two case studies, I analyse the failure of one infrastructure on the grounds of ethical and re-use concerns, and how claims to ethical data sharing also have an infrastructural dimension that is under-exposed in the literature. Conceptually, I enrich conversations that pertain to how data is thought of and managed, from the deletion and retention of data to its 'call'. Finally, my methodological contribution, with colleagues, is to scholarship on industry conferences, reviewing different modes of participant observation as a means of studying the role of conferencing.

Framing these contributions is the idea of the health data ecosystem. While data sources, capabilities and stakeholders are key components, as I go on to discuss in my review of the literature, the concept of an 'ecosystem' is an important idea that appears both as an emic and etic notion - it is used by the actors I study, but also to describe the scope of my study. My aim is to show how, in the work of infrastructuring an ecosystem for health data, specific values become contested and resolved.

The kappa proceeds in the following order: I first introduce my literature review, which positions the work in relation to already published studies. I establish the contribution the papers make to this body of literature, at the intersection of STS and CDS. In the next two sections, I introduce the perspectives I have drawn upon in both collecting and analysing my data. First, my theoretical framework establishes some key thinkers behind the research questions, and my method and methodology explain what I have done and why. This is followed by a summary of the articles as they appear, with attention to their distinct contributions.

2. Literature Review

In this section, I position the thesis and establish its academic relevance in the intersection of CDS and digital health. This endeavour is conducted by focusing on the 'health data ecosystem' concept that I adopt as the central frame for the literature review and thesis at large. I begin by outlining how the concept has engaged in the existing literature before taking the reader through a history of the concept and its legacies. My aim is to illustrate with a fresh perspective the legacies and inter-relationalities between ecological and health data thought, shedding new light on how the literature I seek to engage across the articles sits within a field of rapid change and social importance.

2.1 The Evolving Health data ecosystem

In a recent study, computer scientists in the field of information systems conducted a review of the limited and yet growing literature on 'data ecosystems' across various academic disciplines (n=29), little more than half of the publications defined the notion and did so with great variation (Oliveira et al., 2019).[6] Based on the review the authors define data ecosystems as 'socio-technical complex networks in which actors interact and collaborate with each other to find, archive, publish, consume, or reuse data as well as to foster innovation, create value, and support new businesses' (ibid., p 1). While highlighting characteristics of networks, circular data flows and value creation this definition is also expansive, which may explain the multitude of its competing variations.

In policy and academia, we find subordinate classifications of data ecosystems where the prefix of 'data' is specified to 'big data', 'open data', 'personal data' and 'health data'.[7] The latter prefix is of obvious relevance to the present thesis due to its sectoral specificity to health (e.g. Vayena and Blasimme, 2018; Vayena, Dzenowagis, Brownstein et al., 2018). This does not however mean that all other prefixes are irrelevant to the thesis, since they in practice may - and often do - intersect and overlap in ways that can be difficult to predict. Naturally, data can be personal and tied to health at once. This is for instance the case for personal data from personal health records (Lehtiniemi and Ruckenstein, 2019) that via state authorities or data intermediaries can be repurposed to non-personal data for big data (e.g. Vayena and Glasser, 2016; Vayena and Blassime, 2017) or open data (e.g. Wessels et al., 2017; Heijen and Crompvoets, 2021) applications in health-related research or innovation.

The concept of a *health* data ecosystem is developed by Vayena through a series of articles with Glasser and Blasimme as co-authors. In the following, I engage these articles chronologically to outline the context from which the health data ecosystem emerged, along with its gradual solidification as a concept. An early variation of 'data ecosystems' that is relevant to the present thesis is discussed by Effy Vayena and Urs Gasser (2016) in *"Strictly Biomedical? Sketching the Ethics of the Big Data Ecosystem in Biomedicine".* Writing from

---

[6] The 29 studies included in the review are published between 2011 and 2016 where 25 are distributed in the latter three years.

[7] The term 'ecosystem' is also used to describe other data-intensive environments, such as 'digital ecosystems', 'software ecosystems', 'platform ecosystems', 'API ecosystems', 'innovation ecosystems', 'business ecosystems' and 'service ecosystems' (Oliveria et al., 2019).

the field of bioethics, the authors argue that the use of big data in biomedicine calls for a new ethical framework that 'has to be rooted in an *ecosystem perspective*' (Vayena and Gasser, 2016: 30, original emphasis). This condition responds to an observation that conventional categories and boundaries have become visibly stale and outdated given the shifts in motion in an age of big data. The main shift addressed in the article regards the category of biomedical data that is traditionally defined by source (i.e. the human body) and content (e.g. genetics sequence). This understanding however struggles to capture the emerging practices by which pregnancy, for instance, can be inferred from loyalty card data (i.e. shopping habits) as opposed to a urine sample or blood test. The growing trend of repurposing non-biomedical data for biomedical use is widely recognised in critical scholarship and reflected with the increasing use of terms such as 'health-*relevant* data' or 'health-*related* data'. What an ecosystem perspective is seen to add is a recognition of the 'larger seismic shifts' (Vayena and Glasser, 2016: 25) that problematises a *static* view of which actors are relevant, the responsibilities of private and public sector and the effectiveness of current mechanisms of consent and anonymisation. This implies that the framework must itself have the 'ability to evolve over time' in response to movements in a 'technologically *fluid* environment' but also 'social norms and values around data, privacy, and research' (Vayena and Gasser, 2016: 31, own emphasis). As a final remark, it is also noteworthy that the ecosystem perspective is taken to indicate an inclusion of 'the full lifecycle of biomedical research data, including creation, storage, sharing, aggregation, and re-use' (ibid.). What an ecosystem perspective entails for Vayena and Glasser is then a premise of evolution in which movements in the field, conceptual landscape and regulation should evolve hand in hand with a gaze that considers the 'full lifecycle' of data.

In a subsequent publication from 2017, the Vayena elaborates the notion of 'data ecosystems' with co-author Alessandro Blasimme in relation to 'biomedical big data'. For the authors, biomedical big data 'designates all health-relevant data that can be made interoperable and thus amenable to predictive data mining for health-related purposes' (2017: 502-3). As an analogy, 'data ecosystem' is stated to stress 'the interdependence of the actors and processes that rely on the production and circulation of data as a key resource for their respective activities' (Vayena and Blasimme, 2017: 503). As an idea, it is taken to highlight two points. Whereas the first point reiterates the blurred boundaries between (non)biomedical data the second point underscores a prominent theme in the present thesis: 'that data governance will likely need to draw on a wider array of relevant stakeholders that should encompass actors well beyond the biomedical community, including, primarily, data subjects'.

This second point is the primary focus of the article that engages *New Models of Control Over Access, Use and Governance* (2017) as an alternative to the conventional mechanisms (e.g. informed consent and medical confidentiality) that are challenged as medicine becomes more data-intensive. With a focus on the right to data portability, dynamic consent and participatory governance schemes (i.e. data cooperative model of MIDA.coop), the article is an early contribution to the burgeoning field on individual control (e.g. Hafen, Kossmann and Brand, 2014; Lehtiniemi T and Kortesniemi, 2017; Riso, Tupasela, Vears et al., 2017; Prainsack, 2019b) that I engage with in my work on data intermediaries. For Vayena and Blassime, the need to experiment with new instruments of individual control is demonstrated by data-sharing initiatives in UK healthcare - involving care.data (e.g.

Vezyridis and Timmons, 2017) and google deepmind (e.g. Powles and Hodson, 2017) - where its inadequate consideration is argued to be a central explanation to why they became public controversies. A similar tension between the public and individual interests is also rife throughout the present thesis, and is arguably most explicit in the *public* controversy of the DAMD that I analyse in Article IV *Refusing Retention: Health data and its deletion in the Danish welfare state.*

The theoretical framework of the present thesis is informed by the above ideas, which as I shall return to later resonate with some of the theoretical reflections made in STS. It is however noteworthy how the ideas presented by Vayena and Gasser are present with little - if any - reflection on (or reference to) the genealogical path by which the concept of ecosystem has made its way from ecology to digital health and its metaphorical baggage. This is an observation that applies to the majority of critical scholarship that engages the metaphor in the context of health, although Taylor and Purtova (2019) are a notable exception, and possibly also Bot, Wilbanks and Mangravite (2019) and Blasimme, Vayena and Van Hoyweghen (2019). Before I embark on a reflection as to the kind of work the ecosystem concept does, and where it comes from, I will continue the review of how the 'health data ecosystem' has been described and used in critical data scholarship, and how it has informed my own approaches to research.

The health data ecosystem in critical data scholarship

Sharon and Lucivero (2019) take up 'health data ecosystems' as a special theme in the journal *Big Data & Society*: *The expansion of the health data ecosystem – Rethinking data ethics and governance.* In terms of the framework, it is very similar to the solidified version from Vayena and Blasimme in that it focuses on the entrance of 'new types of data', 'technological tools' and 'stakeholders' in the domain of health (Sharon and Lucivero, 2019: 1). The research orientation is however different and more in line with how I engage the health data ecosystem in the present thesis. The foregoing engagement with the health data ecosystem was conducted from a perspective of law and philosophy with a significant (but not exclusive) focus on *individual* rights and values (e.g. privacy and autonomy) in the context of regulation. Using different perspectives, the contributors in the special theme represent a range of perspectives - law, sociology, STS, philosophy and CDS - that has a collective focus on the *socio*-political questions regarding the health data ecosystem.

The contributions are all premised upon 'the understanding that increased individual control of data subjects is insufficient for anticipating the far-reaching risks and preventing the societal, if not individual, harms associated with this expansion' (Sharon and Lucivero, 2019: 2). The social orientation of the research is also reflected in the argument they put forward, as they assert a need for 'new governance frameworks, technological infrastructures and narratives that are predicated on the *shared* responsibility of multiple stakeholders and *collective* decision-making and control' (Sharon and Lucivero, 2019: 2, own emphasis). Contributions to this argument generally engage questions of public value in relation to identified movements in the health data ecosystem. Some contributions focus on concurrent initiatives involving 'the googlization of health' (Sharon, 2018; see also Jacobs and Popma, 2019) and the lifted ban in Switzerland on the use of genetic information in private life insurance (Blasimme, Vayena and van Hoywegen, 2019). Other contributions adopt a more future-oriented focus on alternatives that provide a possible response to the contemporary power asymmetries exemplified by big tech conglomerates. These responses reorient the discussion of individual control to include collective agency, with studies ranging from engagements with information commons (Prainsack, 2019; Taylor and Purtova, 2019) to data activism in the MyData community (Lehtiniemi and Ruckenstein, 2019).

My research engages this question of individual control in the two primary ways. First, by addressing the DAMD controversy that enkindled the question of self-determination in response to contested ideas of public value in a data regime of presumed consent (Hoeyer, 2016; Nordfalk and Hoeyer, 2020). Second, through an ethnographic exploration of data intermediaries that promote competing models of individual control in the Nordic region. Before I address these strands of research, I first take a moment of critical reflection to account for the central concept that frames this dissertation, namely the ecosystem. In what context did the ecosystem concept originate? What projects did it support? How might its legacies relate to health and data practices in the 21 century? What politics arise as prudent in the demonstrated uptake in health? Reflecting upon these questions provides enriched understanding of the concept that improves the ground for a reflexive engagement.

2.2 A critical reflection on the heritage of the ecosystem

Although many have adopted the term ‘ecosystem’ in their academic vocabulary, it is remarkably seldom that critical scholars take the time to reflect upon the genealogy of its arrival. With its origin in ecology, how - and why - has the ‘ecosystem’ metaphor come to be an emic term among platform practitioners in the field of digital health since its coinage in the mid-1930s (Golley, 1993; Ayres, 2012)? What philosophical and epistemological baggage is cross-pollinated from ecology to digital health (or at least shared between the two)? I answer this question in two steps. In the first, I highlight how the ideas on the origin of ecosystems in ecology can be traced to the contemporary conception of health in terms. In the second, I shift the focus to the ‘data ecosystem’, foregrounding how its adoption in health is preceded with a lineage platform logics that has implications for the traditional role of the state.

From the ecology of nature to human health

As a term, ‘ecosystem’ was first coined in academic publication by British plant ecologist Arthur George Tansley (1871-1955) in 1935. Tansley’s conceptual innovation was published in the article *The Use and Abuse of Vegetational Terms and Concepts*, which today stands out among his seminal contributions to the field of ecology. The term itself would provide answers to what Tansley believed was the central question that should occupy ecological thought in the early decades of 20th century, namely: “What forces drive change in vegetation?” (Ayres 2012, p. 18). The formulation of this question brings out the physiological orientation of Tansley’s thinking that is present in his definition of ecosystem:

> Though the organisms may claim our primary interest, when we are trying to think fundamentally we cannot separate them from their special environment, with which they form one physical system. …. Our natural human prejudices force us to consider the organisms (in the sense of the biologist) as the most important parts of these systems, but certainly the inorganic "factors" are also parts-there could be no

> systems without them, and there is constant interchange of the most various kinds within each system, not only between the organisms but between the organic and the inorganic. These ecosystems, as we may call them, are of the most various kinds and sizes. They form one category of the multitudinous physical systems of the universe, which range from the universe as a whole down to the atom (1935, 299).

As suggested by the article title, Tansley's notion of ecosystem responded to competing theories on how things relate and add up. Tansley was especially discontent with Frederic Clements' 'superorganism', which extrapolated the life cycle of individual organisms (birth, growth and death) to the level of communities that was linked to ideas of 'holism' - the tendency in nature to produce wholes from the ordered grouping of unit structures (Golley, 1993). This (ab)use of the 'organism' metaphor, Tansley argued, failed to capture the dynamic interactions at play necessary to understand the 'whole' system seen through the lens of physiology. In the proposed theory of ecosystems, flora and fauna needed to be understood in relation to the climate, soils and other environmental factors that all interact in functional relationships to shape organic life.[8] Ecologist Frank B. Golley argues that the neological contraction of 'ecological' and 'system' was a timely move. As a metaphor, the latter part of the neologism eco-*system* was 'modern, technical and culturally appropriate' (Golley, 1991: 131). In scientific discourse, the notion of 'system' is noted to have been widely accepted as 'a complex, in which the parts interacted to create a recognizable distinct order or pattern' (Golley, 1991: 131). Contrary to the taxonomic convention of the discipline (e.g. 'biogenocenosis' and 'ecotope'), system is also noted to have been part of the contemporary parlance in the English language. In the subsequent decades, the ecosystem would gain an influential status within ecology and beyond that would most likely exceed the expectations of Tansley. One observation that may not be trivial in understanding this success pertains to the metaphor that the notion of system advanced in ecology: the machine metaphor.

In *A history of the ecosystem concept in ecology* (1993: 168), Golley highlights that the machine as a metaphor generally had a 'strong impact and appeal' in Western society. The decades that followed Tansley's (1935) original publication would provide the first examples of how the ecosystem - and thereby machine metaphor - could be applied in practice. Although Tansley was a terrestrial ecologist, it was an aquatic ecologist called Raymond L. Lindeman (1942) who began the first ecosystem study in 1938 with Cedar Bog Lake, Minnesota. The field of aquatic ecology had decades earlier foreshadowed similar ideas to Tansley in which the lake had been described as a 'microcosm' in which 'nothing can be fully understood until its relationship to the whole is clearly seen' (Golley, 1993: 37, citing Forbes, 1887). The study was partly used to understand ecological succession, which he studied by paying attention to the species composition, food cycles and more drawing on applied energy theory and mathematics to understand the lake as an ecosystem (Golley, 1991). Lindeman's study was eventually superseded by another one by Howard T. Odum in 1957 at Silver Springs, Florida. In this instance, the spring ecosystem was communicated with

---

[8] Tansley is not the first to propose environmental factors as explanatory determinants to plant formation. This suggestion was for instance proposed by German botanist Alexander von Humboldt in 1805 (Essai sur la Géographie des Plantes), highlighting the importance of temperature, rainfall, and soil.

diagrams of energy flows. What made this latter study remarkable, was the possibility to monitor the incoming and outgoing flow of water to calculate the balances of energy and material of the whole system (Golley, 1991) (See figure 3).

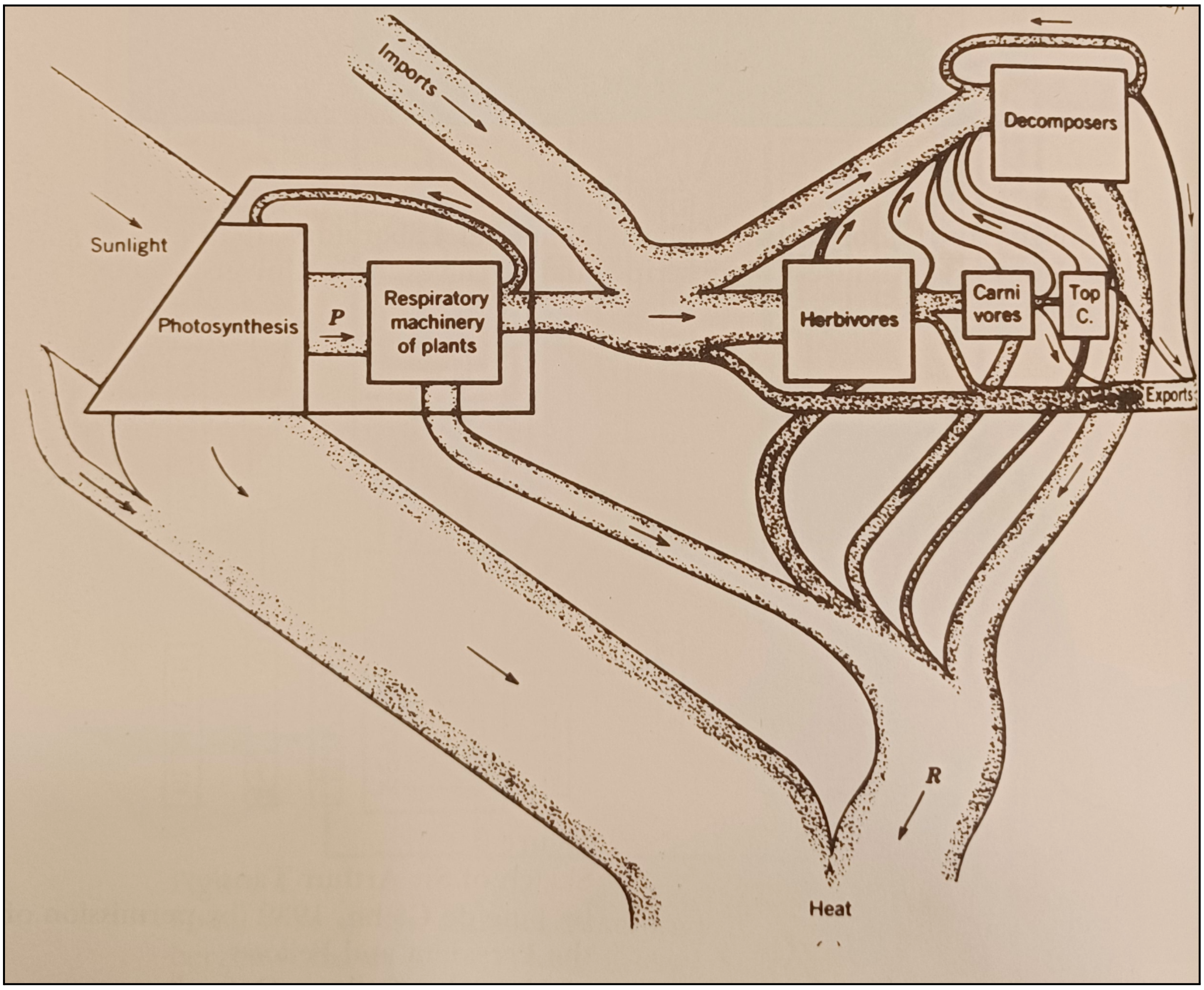
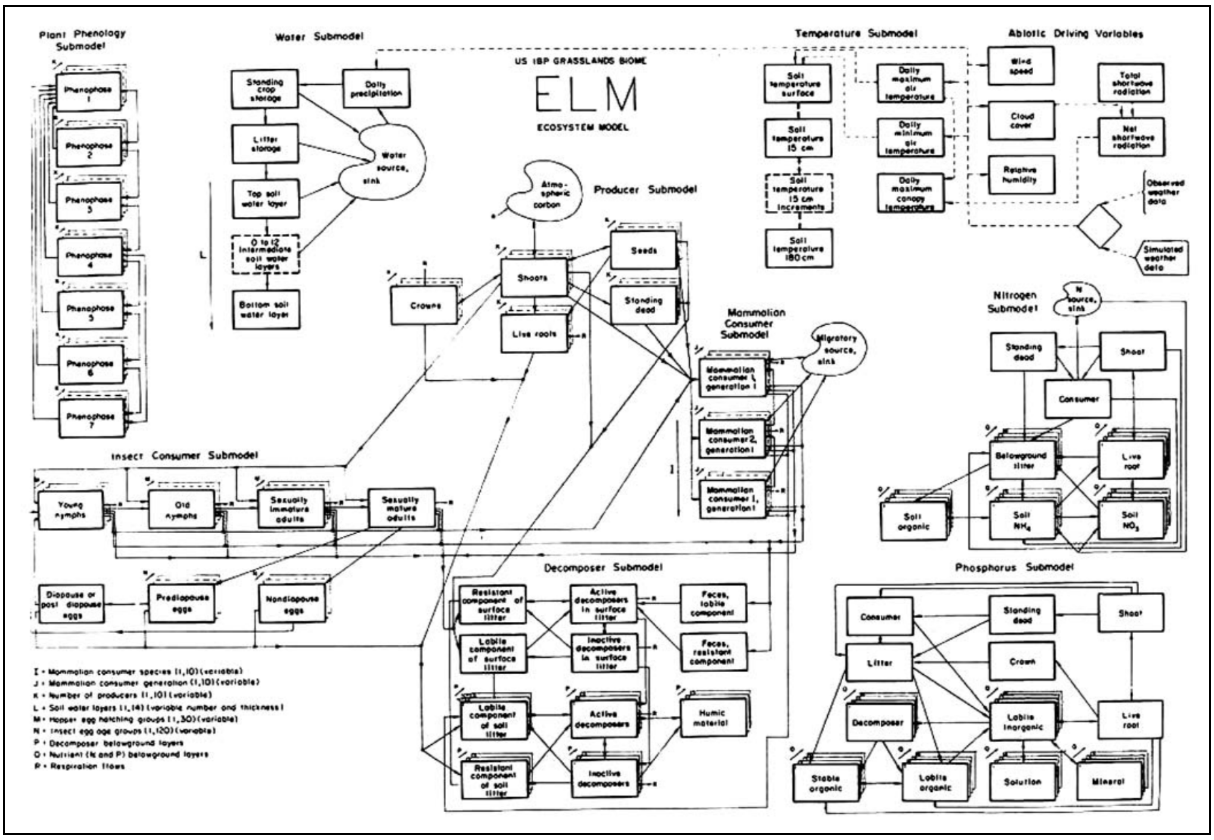


Figure 3. Simplified version of the diagram Howard T. Odum used to study energy flows through Silver Springs, Florida (Hagen, 1992: 88).

Studies in the 1960s led to a progressive advance of 'hard' science in ecology with the embrace of applied mathematics and computational analysis (See Willis, 1997: 269). This advance is indicative of a scientific paradigm of system thinking in which natural scientists attempted to understand the function and structure of nature; 'a social-cultural controlled search for natural order, in a world that seems increasingly chaotic' (Golley, 1991: 129). By 1970, the concept of an ecosystem had been taken up in different directions where it was in part used for the management of natural resources (Van Dyne, 1969), but also organisational life in human society (Pickering, 2010). This was a period characterised by great ambition and creativity at once.

In the early 1970s, system ecologists sought to build a computational model of the grasslands biome in North America as part of a big science project called the International Biological Programme. In retrospect, the objective has by some been described as 'somewhat vague' with a suggestion that the 'real objective' was to demonstrate that a *mathematical representation of a natural system* was feasible despite its complexity (Coleman, Swift and Mitchell, 2004: 11, emphasis added). A team of researchers and assistants collected data on the various ‘submodels’ (e.g. temperature, water and plant) where food cycles of grazing animals were recorded through shadow observations or digestion samples extracted from portholes carved in the stomachs of grazing animals. Built on a CDC 6400 mainframe computer the ‘Ecosystem Level Model’ eventually contained 4400 lines of code, 180 stable variables and 500 parameters (Coleman, Swift and Mitchell, 2004: 11). As the computer was fed more data, what emerged from the simulation was not a system of order, but one of chaos that challenged the theory of equilibrium (Curtis, 2011; see also Cameron and Early, 2015; Bigger and Rossi, 2012). Although segments of the model had its utility at a more modest level of scale, it provided an inadequate representation of the ‘whole-ecosystem’ to prove the underlying theory and manage livestock grazing (Golley, 1993: 134).[9]



---

[9] Looking beyond ecology, we can see how the idea of an ecosystem inspired projects elsewhere (See Cameron and Early, 2015). In *The Cybernetic Brain* (2010: 2), for instance, STS scholar Andrew Pickering details how British cybernetician Stafford Beer, ‘thought that ecosystems are smarter than we are—not in their representational cognitive abilities, which one might think are nonexistent, but in their performative ability to solve problems that exceed our cognitive ones. In biological computers, the hope was that “solutions to problems simply grow” (1962a, 211)’ (Pickering, 2010: 237). This ‘hope’ of problem solving was in the 1960s and 1970s explored through a project in which he used ecosystems to reimage how biological computing could automate factories and redesign how the ‘nervous system’ of the Chilean economy could be managed (Pickering, 2010: 2).

Figure 4. Schematic diagram of the Ecosystem Level Model developed in the Grassland Biome programme (Coleman, Swift and Mitchell, 2004: 12).

These histories speak to the way that computation, modelling and ecosystem have been intertwined. They also speak to the human project of technoscientific advancement that feeds from a persistent zeal to codify, know and regulate worlds, but also the practical limitations and tensions concerning systems thinking that denies a perfect collapse between data and natural worlds. In the following, I demonstrate how contemporary efforts in the 21 century are extending the human project from the grasslands of North America to the health of mankind.

Convergences in the field

Today, in the early decades of the 21st century, ideas that emerged with ecosystems thinking throughout the twentieth century have been adopted in digital health to help explain what forces drive change in *human* life. The contemporary conceptualisation of health has undergone an expansion that to an extent is similar to what the ecosystem entails for nature. If previously, health was something largely confined to the 'body' as the primary source of discovery, assessment and treatment. This view on health is however increasingly considered inadequate in tandem to the biomedicalisaiton of life itself (Clarke, 2010; see also Rose, 2009). Granted, advancements in technoscience continue to push the boundaries by which the medical gaze can pierce through the human body with increased granularity. Such advancements are however also deployed *beyond* the body to appease a holistic conception where health determinants include lifestyle, socioeconomic and environmental factors (e.g. Vogt, Hofmann, and Getz, 2016).

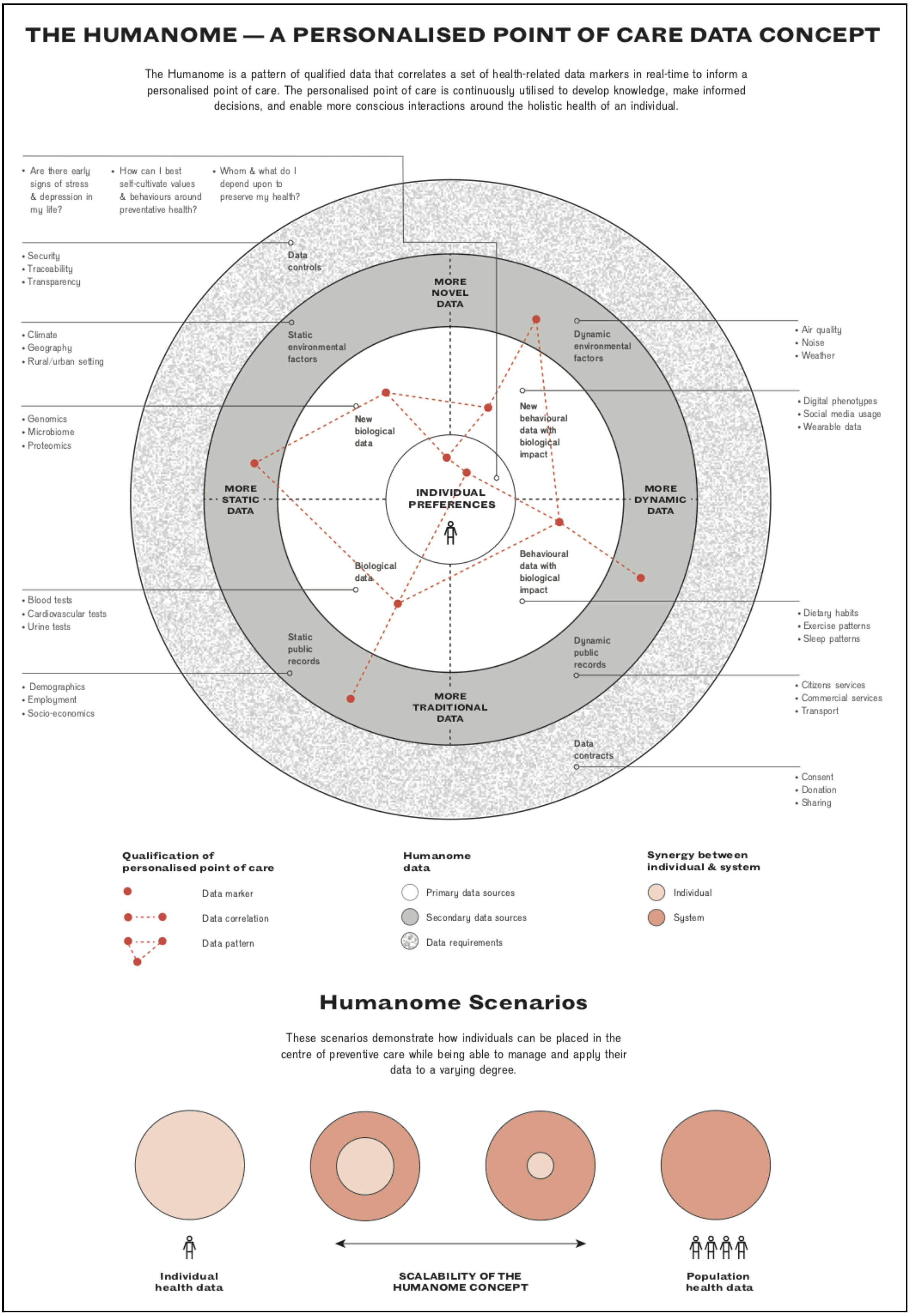
One clear example from my research where the convergence between ecosystem and health is remarkably evident is formulated in the magazine *Nordic Health 2030: Towards Preventive Health* authored by the Copenhagen Institute for Future Studies (2019) and funded by the Nordic Innovation[10]. The publication is based on a series of workshops on the future of health in the Nordics attended by regional healthcare experts from the private and public sector. The expansion of health to include what Tansley terms 'the habitat factors in the widest sense' is depicted in the 'humanone'[11] (see figure 4). What the authors mean by the humanone is a 'qualified data pattern that correlates a set of health-related data markers in real-time to constitute a personalised point of care' (Copenhagen Institute for Future Studies and Nordic Innovation, 2019: 37). What constitutes 'health-related data markers' includes climate, geography and rural/urban life as 'static environmental factors' and air quality, noise and weather as 'dynamic environmental factors' (Copenhagen Institute for Future Studies and Nordic Innovation, 2019: 38). These abiotic and biotic environmental factors are envisioned to supplement public records, biomedical data and behavioural data

---

[10] The Nordic Health 2030 publication is not the first time to forge an independence between *The Health of Ecology and the Ecology of Health* (Rapport, 2010). The humanome does however provide a remarkable illustration of how the idea of an ecosystem can be traced in what health is and how it should be studied. With an embrace of complexity and an ideal of holism, the methodological/ecological apparatus to monitor and intervene in health includes the potential of an endless array of data points that notably transcend the body.

[11] As the publication states, 'the suffix '-ome' refers to the totality of a subject' (Copenhagen Institute for Future Studies and Nordic Innovation, 2019: 37).

are envisioned to be ‘continuously utilised to develop knowledge, make informed decisions, and enable more conscious interactions around the holistic care of the individual’ (Copenhagen Institute for Future Studies and Nordic Innovation, 2019: 37). The link between humans and their natural surroundings is explicitly proposed as a possible means to promote preventive self-care to a new, environmentally aware generation: ‘The health of the individual is inseparable from the health of the ecosystem. To care for the environment and biodiversity, in all their complexity, is to care for ourselves and our loved ones. To take proactive responsibility for our health is to contribute towards planetary care’ (Copenhagen Institute for Future Studies and Nordic Innovation, 2019: 94).



THE HUMANOME — A PERSONALISED POINT OF CARE DATA CONCEPT

The Humanome is a pattern of qualified data that correlates a set of health-related data markers in real-time to inform a personalised point of care. The personalised point of care is continuously utilised to develop knowledge, make informed decisions, and enable more conscious interactions around the holistic health of an individual.

Humanome Scenarios

These scenarios demonstrate how individuals can be placed in the centre of preventive care while being able to manage and apply their data to a varying degree.

Figure 5. Illustration of how the ‘Humanome’ and the data points considered relevant for the holistic health of an individual (Copenhagen Institute for Future Studies, 2019: 38)

This newly conceived relationship between health and data is reflected in the two core case studies of the present thesis. The DAMD database was introduced to general practice to promote logics of prevention and self-care in diabetes care through a decision support tool called Sentinel. To this end, patients were requested to disclose information on their lifestyle to their general practitioner that would in part be used to set individual targets for self-improvement. This lifestyle data was largely limited to the somewhat conventional habits in disease prevention in Denmark (i.e. dietary, smoking, alcohol and exercise, See Vallgarda, 2007). The data intermediaries, by contrast, take the datafication of health to a new level (Ruckenstein and Scüll, 2017; see also Lucivero and Prainsack, 2015). As the founder of the Data for Good Foundation formulated it in 2016, they seek to create a '360 degree overview of the individual' to understand the individual's 'needs, motives, barriers and so on' (Article V). This description was given by the founder of the Data for Good Foundation, but it can be extended to the Digi.Me too since both startups believe that virtually all personal data can be relevant to health (e.g. retail data, social media data, bank record, spotify data etc.). Indeed, one could even claim that what the data intermediaries want to achieve is strikingly similar to the 'personalised point of care' described in the Nordic 2030 vision where personal data are combined from public and private services alike. The epistemological underpinnings exhibited in the vision of a '360 degree overview' entertains the idea that a complete 'data double' (Ruckenstein, 2014) can be obtained. This would however imply a total collapse between the world and data, which in the ambitious grassland biome project proved so elusive to be achieved in practice. While Sentinel was useful for the administration of diabetes care, the digital patient profiles were found inadequate for the envisioned benefits concerning remote clinical judgement (Article II, see also Lippert, Kousgaard and Bjerrum, 2014; Lippert, Reventlow and Kousgaard, 2017). Regardless of the amount of data points, what is illuminated is produced and interpreted in specific ways that in turn produce data shadows (Leonelli, Rappert and Davies, 2017).

Data ecosystems and the advance of platform logics

According to the authors of *Open Data and the Open Knowledge Society* (Wessels et al., 2017: 62), 'data ecosystem' was coined by Tim O'Reilly in the 2010 anthology *Open Government: Collaboration, Transparency, and Participation* (Lathrop and Ruma, 2010: xx). Dubbed 'The Oracle of Silicon Valley', O'Reilly has in the past decade been a vocal advocate of 'Gov 2.0', asserting that government must be more than a 'vending machine'; something 'we put in our taxes and we get out services [...] And if we don't get the services we want, we shake the vending machine. We get to protest. We write our congressmen. We have a tea party' (Chafkin, 2010). What O'Reilly (2013: 13) proposed was an alternative government modelled in the image of platforms from which he draws the connection to ecosystems:

> In the technology world, the equivalent of a thriving bazaar is a successful platform. If you look at the history of the computer industry, the innovations that define each era are frameworks that enabled a whole ecosystem of participation from companies large and small. [...] This is the right way to frame the question of Government 2.0. How does government become an open platform that allows people inside and outside government to innovate? How do you design a system in which all of the outcomes aren't specified beforehand, but instead evolve through interactions

> between government and its citizens, as a service provider enabling its user community?

The proposal O'Reilly makes in this quote on *Government As a Platform* (2013) speaks to a specific agenda in the open government data movement that focuses on innovation and services delivery more than democratic ideas of transparency and accountability. It is also one that calls into question the role of government as the central agent to solve the problems of the collectives, advocating for the adoption of Web 2.0 technologies and APIs to facilitate entrepreneurial participation.

The term ecosystems is also widely used in platforms studies, which is an interdisciplinary field that has provided important contributions on the logics of platforms (e.g. Pascquale, 2015; Guyer, 2016; Srnicek, 2016; 2017; van Dijck, 2018; Gillespie, 2018). As the titles of *Platform society* (van Dijck, 2018) and *Platform Capitalism* (Srnicek, 2016) suggest, the contemporary platformisation has profound implications on social and economic life. Drawing on such literature, scholars in STS have recently turned their attention to explore how *ecosystems* are enrolled in the advance of platform logics in the traditional role of states in national health systems. In *Platform NHS: Reconfiguring a Public Service in the Age of Digital Capitalism* (Faulkner-Gurstein and Wyatt, 2021), an early appearance of the ecosystem from 2011 is traced to the gradual process of platformisation in the UK's National Health Service. Ecosystem is in the case study engaged as an emic notion in relation to a question of 'how to connect this research infrastructure to both public and private investment' (Faulkner-Gurstein and Wyatt, 2021: 11). The timing of this appearance is by the authors understood to mark a discursive shift toward 'a more entrepreneurial approach to patient data' in a gradual embrace of platforms in the delivery of public services in the UK (Faulkner-Gurstein and Wyatt, 2021:11). The logics salient in this instance of platformisation regard the logic of catalysis, delegated governance, and data entrepreneurialism. Moving across the North Sea, Aaro Tupasela, Karoliina Snell and Heta Tarkkala articulate similar logics in relation to the *Nordic Data Imaginary* (2020)*.* In this instance, ecosystems are traced as part of the response by which the Danish and Finnish welfare states 'try to adjust and benefit from, new pressures and opportunities to utilise their data resources in data economies' (Tupasela, Snell and Tarkkala, 2020: 2). Analysing the data imaginaries in the policies and strategies, the authors describe the creation of national data ecosystems as something 'that allows for the platforms to function' (Tupasela, Snell and Tarkkala, 2020: 5). In doing so, states become producers, consumers and notably also *enablers* of data, which echoes aforementioned logics in Platform NHS.

My study of data intermediaries in the present thesis is an extension of this trajectory of research. Commenting on the Nordic data imaginary, the latter study highlights how the 'development of national data ecosystems and platforms which increasingly support public–private collaboration can also be seen as a type of legitimation exercise for both public and private actors' (Tupasela, Snell and Tarkkala, 2020: 11). As discussed in Article V, the Data for Good Foundation startup is predominantly financed by national funds and operates with an explicit objective to exploit the economic opportunities in Danish health data. This is for instance reflected in the formulated 'societal' benefit of one of the funded project in which cancers patients are invited to donate personal health data to supplement registry data (HedaX, n.d.): ‘New insights on the basis of Danish health data can attract companies and clinical trials to Denmark. This will strengthen the Danish ecosystem within life science even more and create more jobs and economic growth’. With this funding, the state takes on the role of an enabler or catalyst for data entrepreneurialism with DfG as a delegated instance of governance. While Tupasela, Snell and Tarkkala correctly suggest that such collaborations as ‘as a type of legitimation exercise for both public and private actors’, I suggest that what this looks like beyond the policy landscape requires further elucidation. The data intermediary study responds to this shortcoming by exploring how the data startups promote themselves as ‘the ethical choice’ to different audiences based on event ethnography, interviews and document analysis.

The politics of ecosystems

In sum, despite its biological heritage, the creation of data ecosystems is not a ‘natural’ course of action but a social and political one. The notion of ‘ecosystems’ has in recent years gained currency among regulators, practitioners and scholars as a metaphorical device to communicate ideas about how the circulation and value of data are imagined (e.g. European Commission 2017; for ‘data ecosystems’ see OECD, 2015; United Nations Development Programme, 201x). At an OECD conference that took place in 2019 on competition and the digital economy, the EU competition commissioner Margrethe Vestager stated that ‘The biggest threat to competition and innovation [...]comes from platforms that are not just a single business, but the centre of large empires’ (Stolton, 2019). Asserting the stance of the EU, Vestager invoked an analogy from biological life: ‘As ecosystems grow, the Commission will need to keep a close eye on strategies that undermine interoperability[...] So the expansion of platforms into new markets doesn’t undermine competition as it goes, like a bloom of algae that kills off every other form of life as it expands’ (Stolton, 2019).

The European data protection landscape, as a means of managing the ecosystem, comprises an important backdrop to the present thesis. The newly introduced right to data portability has been heralded by regulators and practitioners operating in the volatile innovation space for data intermediaries. I address this directly in Article I, but Article V establishes the relevance of the right to data portability while also attending to the broader enrollment of the GDPR in the promotion of ethics through infrastructures. While the data intermediaries I study offer the assurance of GDPR compliance, the promise they put forward is one of data-intensification. Article III and Article IV by contrast illustrate the perils of data-intensification by examining the downfall of an otherwise promising national database (Article I). The former article scrutinises how the DAMD grew into a data hoard between 2007 and 2014, which exceeded the remit of Danish law that employs the principles of data protection currently championed by the GDPR (e.g. data minimisation, storage minimisation and purpose limitation).[12] Shifting the attention to the public controversy, Article IV explores the political debate that ensued between 2014 and 2015 on what to do with the data that was illegal and yet potentially invaluable. The article challenges the contemporary positivity of data retention as a default mode for valuing data by shifting the attention to *deletion* that, with the GDPR introduced sanction for non-compliant data conduct, has become an enhanced priority for organisations.

---

[12] The regulatory and infrastructural reform on the secondary use of health data have more recently been proposed in Finland (Aula, 2019).

My aim of paying attention to the movements in the health data ecosystem is then not to build a generalised theory, but to stimulate reflection about the broader relevance of the situated accounts analysed in the case studies. In sum, I have traced the origin of the ecosystem in ecology, the project it supported in natural worlds, its extension to human life and the politics it engenders.

3 Theoretical Framework

The analysis I have undertaken is founded in STS but informed by work in infrastructure studies, CDS and new materialism. Building on the review of the ecosystem concept established above, in this introduction to my theoretical framework, I use the ecological metaphor of ecosystems as a heuristic to think through the social theory included in the theoretical framework. As I have proposed to consider the building of health data ecosystems, I first consider how my work has been informed by work on infrastructure from STS and anthropology. I then draw on the life cycle concept that has been embedded within ecosystem literatures and within CDS, to explain how I have made sense of the different phases of data's life appearing in my papers. Finally, I turn to the way the life of data, in particular its liveliness, is made animate in my writing. Building on these three sets of theoretical commitments, I am able to pay attention to the particular sociomaterial coimbrication of personal health data ecosystems. In other words, my theoretical framework enables me to understand and analyse personal health data ecosystems as material and social constructs that are laden with meaning, surprising dynamics, countervailing tensions, particular histories and consequential choices. Rather than simple or neutral "technical" objects or standards, this framework allows me to show that the promises, hopes and politics associated with personal health data are not simply related to human actors, but also stem from and interwoven with the non-human infrastructure and data itself.

In the following section I will outline this theoretical framework in greater detail, describing the concepts of infrastructure, life cycle and liveliness in turn. The sections both introduce the concepts, the key sources from which they are derived and discussions they engage in, and where relevant, connect them to the various articles of the dissertation. A short paragraph concludes the section, describing the contributions I believe the dissertation makes to the fields/understandings comprising the framework.

3.1 Infrastructure

In Section 1, I posited that if infrastructures are part of the ecosystemic shifts then they can also be a means to study an ecosystem itself. While this explains the function of infrastructures in this thesis, it does not explain its selection over alternative possibilities to make sense of the evolving ecosystem. Replacing infrastructures with policies in the research design would for instance have taken my fieldwork to different sites, actors and tensions and my thesis in another direction. Policies can also be viewed to shape the conditions of possibility for data through regulation, why a study of their making or discourse could also be relevant. Digital infrastructures however offer something different. As Edwards points out, infrastructure is analytically generous, both as concept and practice: it provides a construct that 'not only bridges the scales but also offers a way of comprehending their relations' (Edwards, 2003: 186). This is particularly important to the research design of this thesis considering that the task is to use *situated* case studies to enhance comprehension about the ecosystem they inhabit and its moving parts.

The fertility of infrastructures as a concept or analytical device is however not limited to the bridging of *scales* and comprehension of *relations*; they can also aid the task of finding *tensions*. Drawing on Larkin, infrastructures can be defined as 'matter that enable the movement of other matter' (2013, 329). This is a relational construct constituted from whatever is enrolled to move people and things from one place to another. Much of this is often taken for granted in everyday life (Bowker, 1995). When surfing the internet, as it was once called, few stop to think about the sociomaterial practices of maintenance (Bowker, 2005; Graham and Thrift, 2007; Strebel. Bovet and Sormani, 2019), content moderation (Neyland, 2018; Skinner and Wienroth, 2019), storage (Hogan Vonderau, 2019; Maguire and Winthereik, 2019), automation (Torenholt and Langstrup, 2021), transmission (Starosielski, 2015), classification and standards (Bowker and Star, 1999; Krogness, 2011) as long as the requested bytes appear on the screen. They become 'black boxed', deceptively dormant in the background. As experienced in Danish healthcare, infrastructuring a seamless flow for the movement of data requires work that in practice will often encounter 'friction' (Bonde, Bossen and Danholt, 2019; Langstrup, 2019). As Wadmann and Hoeyer (2018) point out, the DAMD case showcases remarkable infrastructural work in this respect that I engage more closely in Article II and III.

For years, the DAMD operated as a black box, growing in size and importance without public concern nor awareness. That is until its contentious life began, after which it became the main object of inquiry for journalists, officials and civil society (see Article IV). Paying closer and more critical attention to the infrastructure were scholars who also gained an interest in the case (Langhoff et al., 2016 and Wadmann and Hoeyer, 2018) of whom I follow with my contributions (Articles II, III and IV). Infrastructures can however also provide an entrance to understand tensions in settings that are better characterised as mundane than controversial. Article II for instance unboxes the logics accompanying the algorithmic design of the decision support tool for improving diabetes care in general practice (see also Lippert, Kousgaard and Bjerrum, 2014; Lippert, Reventlow and Kousgaard, 2017). What the DAMD case illustrates with regards to infrastructures is that while they are material, they are most certainly also social and in ways that are deeply political (Knox and Nafus 2018). To borrow a phrase, infrastructures can be viewed as *political machines* (Barry, 2001) that organise social complexity, manifested in ways that paves way for some and builds walls for others (Star, 2002; Reeves, 2017).

Infrastructures can also provide a means to understand the political economy of data. Writing from the US health sector, STS scholar Linda F. Hogle highlights that 'Scrutinizing the points in development where infrastructures were constrained or made to flow freely can reveal a good deal about political-economic contexts in which technologies arise or stumble, and tensions between prospective stakeholders and users' (2016: 402). Examples of the deployment of this tactic can be observed throughout the article collection of this thesis. The question of political economy is however most saliently articulated in relation to the case study on data intermediaries. As Hannah Appel and her co-authors in anthropology write, 'New infrastructures are promises made in the present about our future' (Appel, Anand and Gupta, 2018: 27). The promissory future entertained by data intermediaries is an *ethical* and *sustainable* way to unleash new opportunities of extracting personal, commercial and public value from the 'scientific goldmine' of Denmark and personal data more broadly. This is detailed in Article V, which also provides a reminder that the studied data intermediaries exist amidst the ruins of legacy infrastructures that were once a placeholder for a similar promise Appel, Anand and Gupta, 2018).

Thus seen, infrastructures never exist in isolation and are always in the making, which alludes to the relevance of also positioning the DAMD case as a backdrop for emerging data intermediaries in Denmark (see Article III for long-term implication of the DAMD controversy).

Platforms as infrastructure

State institutions have traditionally held a central role in the registration of health data. This is particularly true for Nordic welfare states where health registries have in scientific discourse been heralded as 'goldmines' for their comprehensive, longitudinal and person-identifiable qualities (Bauer, 2014; Tupasela, 2020). The authority of such institutions as gatekeepers to health information is however in decline as societies grow more digital. Wearables, social media and retail are among the emerging consumer services that offer a plethora of data points that can be used or repurposed to gain health-related insights (Vayena and Blasimme, 2016). With this in mind it is less surprising that the big tech oligopoly of Google, Apple, Facebook and Amazon (GAFA) have moved to assert their mark on the lucrative market of digital health - bringing questions of power concentration with them (e.g. Sharon, 2016; Zuboff, 2019; Prainsack, 2019a) - along with less established players. Not only does the consumer market for digital health appeal to the elimination of illness, it also encompasses governing the risk of illness and indeed life itself (Rose, 2009). To enter this expansive market it is somewhat insignificant whether data are defined according to the conventional criteria of 'health data' (source, content or declared purpose upon collection) as long as they can be *made* relevant to health (Vayena and Blasimme, 2016). This is in part facilitated by technological advancements in computational capabilities of analysis, that in turn create a demand for more data. The reciprocity of this relation is for instance articulated by Ian Hacking in his work on the history of statistics that Slota et al. extend to contemporary data science. Similarly to how Hacking (1982: 280) describes counting as 'hungry for categories', Slota et al. (2020: 4) characterise data science as 'endlessly hungry' as 'in the ongoing discovery of new analytic tools, techniques, and applications more unanalyzed data, domains, and opportunities for development are revealed'. Given that the same observation is increasingly true regarding the sectoral origin of data, new platforms are in this particular 'data moment' (Maguire et al., 2020) being built to facilitate combined access to health data from state institutions and industry alike. The moment of uncertainty described in the introduction of this thesis then also extends to the roles, responsibilities between state, market and citizen and the relation between (Article V, see also Tupasela, Snell and Tarkkala, 2020; Faulkner-Gurstein and Wyatt, 2021).

Infrastructures for ownership

Platforms tend to advance totalising rights claims to 'ownership' over personal data about Danish citizens (Article V). This can be seen as problematic in a society where data about individuals are financed by taxpayers, used in the interest of society (e.g. public health) and can contain information about others (for more see Purtova, 2012; 2015). Yet, ideas of ownership remain central to contemporary platformisation, which has previously led to scholarly debates around property rights in personal data. Proponents of propertisation have in this context argued for the legal acknowledgement of commodification, while antagonists argue that the market exchange of personal data is ill-suited for the protection of information privacy as a public good (Purtova, 2015). The Swedish Data Protection Agency for instance voiced their concerns about the unfathomable consequences of sharing sensitive personal data to a potentially endless array of third-party services via the Hälsa for Mig platform (Ny Teknik, 2018). This model assumes that individuals can manage their own privacy, which includes anticipating future use through techniques of inference in combination with additional data sources (Lehtiniemi, 2017; Solove, 2013). Even if feasibility issues pertaining privacy self-management are bracketed, a precursory question is to whom are the prescribed benefits of active citizenship or patientship desirable?

Figure 6. Screenshot from Taltioni's promotional video where the full sentence reads: "It's a new form of empowerment! It encourages taking responsibility for your own and your loved ones' well-being. It's what you make it!" (SitraFund, 2010).

The platforms all promote a promise of 'empowerment' to enable individuals to control their own data and thereby also health. We have just seen how a step towards greater control over personal data is envisioned to cast individuals to take a more proactive role in the self-management of privacy. This suggests a responsibilisation of the individual, which extends into the domain of health in support logics of self-care. In a promotion video for the Taltioni, access to a health profile and derivative services allegedly "makes you your best wellness expert and coach" (SitraFund, 2010). Aside from issues of digital and health literacy, this presupposes access to valid data and responsible interpretation of data produced through fragmented practices. For obvious reasons, the difficult side of data sharing is hard to find in the promotional material. Indeed, it is typically a utopian discourse of data sharing with statements such as "It will make you feel better" or the jubilant portrayal of 'empowerment' and 'responsibility' depicted in Taltioni's promotion (ibid., see figure 5). The Data for Good foundation is another example of such discursive asymmetry as data sharing is embellished in an altruistic language enticing individuals to "donate" their data for purposes that are suggestively “good”. This includes combining health, behavioural and demographic data for the exploration of new insurance models based on predictive analytics to “nudge individuals to take proactive care of their health” (Data For Good Foundation, n.d.). Whether the exercise of power through nudging technology is equally welcomed by insurance providers and takers has its sceptics (Prainsack, 2020), but it is ultimately an empirical question. While Data for Good Foundation offers infrastructural precautions of data protection for individuals (see Article V), the insurance example of risk stratification makes the question of ‘group privacy’ more prudent (Taylor, Floridi, & van der Sloot, 2016).

3.2 Life cycles (of data)

The notion of the health data ecosystem is however not merely a tool to abstract the level of perspective in the case studies. Vayena and Gasser have noted that an ecosystem perspective calls for consideration of ‘the full lifecycle of biomedical research data, including creation, storage, sharing, aggregation, and re-use’ (2016: 31). Although the present thesis does not focus on biomedical research, the idea of attending to the full life cycle of data is instructive for the way the articles are organised. Bracketing Article I (which provides a contribution on methods in the study of ecosystems), the four other articles of this thesis attend to specific stages in the life cycle of data (See table 1).

Article	**Data life cycle stages**
Sentinel Vision (II)	Production and reuse
The Call of Data (III)	Collection and retention
Refusing Retention (IV)	Retention and deletion
Infrastructures in the Promotion of Ethical Tech (V)	Collection and reuse

Table 1. Overview of salient stages in the life cycle of data addressed in each article.

Using the life cycle to organise a collection of disparate contributions is common in data studies (e.g. Leonelli and Tempini, 2020). The idea that data goes through *stages* is familiar to most readers, but what it means to view these stages as *cyclical* is intuitively perhaps less legible. The *cycle* is a notion that signifies the structuring of the life of data in a particular way. In a research article published three decades ago, Angela O'Rand and Margerat Krecker (1990) are among the scholars who discussed the import of biological concepts into social science with a focus on the life cycle. At the time of writing, the authors placed the life cycle ‘among the most widely used concepts in the social sciences’ (O’Rand and Krecker, 1990: 241) with noteworthy histories in anthropology, psychology and economics too. One key explanation for the cross-pollination of concepts from biology to other scientific disciplines is attributed to their metaphorical capacity ‘to simplify complex social phenomena by invoking familiar schema in scientific thought’ (O’Rand and Krecker, 1990: 242)

However, as I suggested earlier, ecological thinking has to a degree at least become naturalised in certains branches of social sciences. Whether this lack of friction might explain why Vayena and her co-authors - and most others who draw their work - do not interrogate the genealogical backdrop of the ecological terms they use can only be speculated. Regardless of the reason, concepts always come with baggage and are never neutral, which is important since they are part and parcel of how humans make sense of the world. Among those who subscribe to this claim is Donna Haraway (2016: 12), who drawing on anthropologist Marilyn Strathern writes:

> It matters what matters we use to think other matters with; it matters what stories we tell to tell other stories with; it matters what knots knot knots, what thoughts think thoughts, what descriptions describe descriptions, what ties tie ties. It matters what stories make worlds, what worlds make stories.

The idea that concepts matter was also recognised by the man who coined 'ecosystem' in the academic discourse of ecology about a century ago in 1920. 'We must never conceal from ourselves that our concepts', Tansley wrote, 'are the creation of the human mind which we impose on the facts of nature' (Ayres, 2012, xi, citing Tansley, 1920). Considering the time of writing, it can perhaps be debated whether the concern of Tansley was more epistemological and less political than Haraway's poststructural critique. It has however been suggested that Tansley was wary of the societal and political adoption of biological concepts. This has for instance been suggested in *Imperial Empire* (Anker, 2001) in relation to a particular ideation of holism advanced by scholarly peers 'because it could be used to condone the placement of non-whites in separate "biotic communities"' (Cameron and Early, 2015: 476).

Regardless of his political orientation, the fact remains that Tansley recognised that concepts had implications for how the object of study is studied and understood, which he also voiced when he introduced 'ecosystems' in *The Use and Abuse of Vegetational Terms and Concepts* (Tansley, 1935). The political dimension highlighted by Haraway refers to the politics of how thoughts, descriptors and stories 'make worlds', which with the Foucauldian amalgamation of knowledge/power also extends to scientific knowledge production (Gordon, 1980). For example, returning to the field of health data, Vayena and Gasser draw on such thinking to call for a consideration of the full life cycle of data with a view to recalibrate the conceptual and regulatory landscape for an ethical framework. The metaphorical application of life cycles to data aptly highlights the non-linearity of contemporary practices in which the beginnings and ends of data have become increasingly difficult to demarcate. The linear model that prescribes that data are produced once for one purpose falls painfully short of the current reality where it has become the ideal, and to an extent norm, to recycle, repurpose and recontextualise data from a living source (Custers and Ursic, 2016). The metaphorical application of the life cycle to data is in this respect a *cyclical* depiction of data that introduces a necessary complication that should be reflected in the current regulation.

Moving beyond an assumed cyclical nature, and drawing on critical scholarship, I reorient the focus to take a closer look at the *life* of data through a social lens informed by STS and CDS. Viewing data as something with a *life* is even more challenging and an exercise that some might derogatorily write off as fictitious or speculative, a troubling notion which disturbs boundaries of what is conventionally considered 'living'. Nevertheless, feminist scholar Donna Haraway argues that "*staying with the trouble"* (2016) can in such exercises still be analytically rewarding. In my work, I bring the life cycle into conversation with CDS and new materialism as a means to cultivate an understanding of data that is even more appreciative of the complexities of data's life. The theoretical ground on which the thesis is based draws inspiration from the idea of taking the *life* of data seriously, as something living with the capacity to evolve. Phrased differently, the following establishes the conceptual foundation on which the present thesis has attended to the *life* of data as something that is simultaneously social and lively.

3.3 The Liveliness of Data

The analytical move of conceptualising data as something lively with the capacity to evolve will not be entirely alien to scholars in CDS and STS. Framing something with the capacity to 'evolve' signifies a *process* that entails a development over time. This is for instance recognised in the health data ecosystem concept with the focus on territorial *movements*. When health data are described to evolve it is then an attempt to make a break with the unproductive tendency to view data through a lens of inertia. This is not to say that data undergo a 'natural' development. The point is rather that data are social by nature. It is for this reason that Geoffrey Bowker in *Memory Practices in the Sciences* criticises the construct of 'raw data' as 'an oxymoron and a bad idea' (Bowker, 2005: 184). Instead, Bowker argues that data are better understood as always 'cooked' (2005: 184; see also Gitelman, 2013), which is a position of broad acceptance among critical data scholars. As Christine L. Borgman writes, 'recognizing that some phenomena could be treated as data is itself a scholarly act' (Borgman, 2015: 5). Characterising data as 'ubiquitous, yet often ephemeral', Borgman arrives at the conclusion that it is often more fitting to ask "*when* are data?" - as opposed to "*What* are data?" - if one wants to understand the ontology of data (Borgman, 2015: 4).

The conceptual exercise of relating liveliness to data has itself occupied the attention of other scholars writing in the intersection of digital health and CDS. In *Living Data: Making Sense of Health Biosensing* (Roberts, Mackenzie and Mort, 2019: xi), the authors of this book publication note that two meanings are embedded in the first part of the title: 'that data is lived with, and in, by individuals, groups and constituencies; and that it is living, changing and proliferating in sometimes unpredictable ways'. A similar characterisation of data as social and ever-changing are also articulated by Deborah Lupton in *Lively Data, Social Fitness and Biovalue: The Intersections of Health Self-Tracking and Social Media* (Lupton, 2017). In her chapter contribution, Lupton highlights that a conceptualisation of 'lively data' can be understood in at least four ways. The first conceptual take draws attention to how 'data are generated from life itself by documenting humans' bodies and selves' (2017: 563). The second take speaks to the aforementioned characterisation of data as ever-changing or evolving, as they are described as 'labile and fluid, open to constant repurposing' (ibid.). The third take reiterates the sociality of data by pointing to their 'potential effects on the conduct of life and life opportunities' (ibid.). The fourth and final understanding also highlights how personal data 'by virtue of their growth as commodities or research sources[...] have significant implications for livelihoods' (ibid.). With livelihood, Lupton refers to those using the personal data industries such as insurance and data science. In summary, we can provisionally understand the relation between liveliness and data in terms of the (1) datafication of life, (2) the fluid ontology of data, (3) the use of data in the governance of life and (4) the implication for livelihoods caused by the growing value of personal data.

All of the articles in this thesis speak to the life and liveliness of data across these various conceptions. However, liveliness is most evident in Article III, *The Call of Data,* where I ask “How could personally identifiable data on three-quarters of the Danish population end up in the Danish General Practice Database when its authorisation was limited to four chronic conditions?”. Drawing on work on vital materialism, in particular Jane Bennett’s *Powers of the Hoard* (2012)*,* I follow her in engaging the notion of hoarding from said perspective. I devise an analytic in which data are conceptualised with the capacity to ‘call out’ and command both action and inaction (Article III: 16). The ‘call of data’, as I call the analytic, brings a vital materialist understanding of hoarding into conversation with CDS that more precisely brings attention to ‘how data, in their embedded materiality, exert an agential and aesthetic imprint on social phenomena’ (ibid.: 2). With this lens, I conduct a *post-mortem* of a now defunct database to find the answer to the conundrum of how three-quarters of the Danish population ended in the DAMD when its authorisation was limited to four chronic disease areas. What unfolds in the analysis is a story wherein it at times ‘was almost as if the DAMD was seen to have “its own momentum or drive to persist and grow” (Bennett, 2012: 252), damned with a curse of being “endlessly hungry” (Slota et al., 2020: 4)’ (p. 20). As I describe in the article, ‘in contrast with a psychopathological focus on subject behaviour, Bennett problematises the ontological bias of depriving objects the possibility to speak before the study even begins. Writing against anthropocentrism Bennett asks: what if the objects that come to form hoards also play an active role in the phenomenon of hoarding?’. From here, I develop an analysis of how data accumulation occurred within the Danish database, asking how Bennett’s questions about the agency and aesthetics of material objects might illuminate the phenomenon of digital hoarding. In this respect, my approach is similar to what anthropologist Antonia Walford describes in *Data Aesthetics*, namely ‘the ways in which data’s forms are charismatic, how they capture attention, or command action, in ways other than because of their objectivity or accuracy’ (Walford, 2021: 206). Instead of imposing philosophical ‘judgements about the beautiful and good’, Walford (2021: 205) understands aesthetics to be ‘about what constitutes a 'persuasiveness of form' in any cultural or social context (Strathern 1991: 10; cf. Riles 1998)’.

The theoretical framework of the present thesis thus consists of three overall streams of work, dealing respectively with the infrastructures, life cycles and liveliness of data. By way of the articles making up the thesis, I both draw from and contribute to these three topics in a number of different ways. The main contributions consist of 1) foregrounding the role of specific sociomaterial infrastructures and the scales, relations and tensions they produce within health data ecosystems, 2) exemplifying how health data ecosystems can be understood through the concept of “lifecycles” whilst simultaneously denaturalising and interrogating the concept itself, and finally 3) I show the potential and importance of making new materialist readings of data, underscoring how the penchant for accumulating data also resides within the data itself. In particular, the dissertation provides a contribution via its new materialist reading of data, which can be thought of as a problematisation of the tendency towards schematism inherent in the approach of designating abstract life cycles. Rather than think of data as reducible to predetermined patterns or cycles, the dissertation shows how the life of data is unpredictable and must itself be accounted for in our analyses. Taking the above into account, it is not easy to neatly delineate between particular contributions in terms of STS, CDS, infrastructure studies or new materialism, but I believe that all of these disciplines and sub-fields benefit from the cross-pollination which the dissertations’ articles represent.

4. Methods and Methodology

As presented in the introduction, this thesis sets out to answer the central research question of how data-intensification is sought by infrastructural initiatives in the Danish ‘health data ecosystem’, and what tensions arise in the pursuit of valuations assigned to personal health data? In order to arrive at answers, I conducted fieldwork between November 2016 and June 2021 and during that period, have deployed a range of methods that are both traditional and emerging in ethnographic inquiry. With traditional methods, I am referring to participatory observation, interview and document analysis while emerging methods relate to the use of digital methods for the purpose of collecting and visualising Twitter data. While each method contributes to the thesis as a whole, they do so in distinct ways and with different weight. This is for instance visible in the two phases of fieldwork conducted in the above-mentioned period. The first phase is *preliminary fieldmapping*, which was carried out within the first year of my study to explore the contemporary unfolding of the health data ecosystem in Denmark. The mode of inquiry was in this stage *explorative* to build an empirically informed research design through the identification of tensions, themes and potential case studies encountered in the field. The preliminary fieldmapping was conducted through 10 participatory observations (between November 2016 and November 2017), 10 interviews (between February and September, 2017) and document analysis.

What emerged through this preliminary fieldwork was a specific interest in two infrastructural initiatives that both engendered the same question: How can the use of health data be infrastructured in a sustainable manner? While the DAMD managed to intensify the use of health data, this achievement was ultimately short-lived as the infrastructure came to a sudden halt in 2015. Based on my preliminary field encounters, I looked to explore why the case was still observable - and observably sensitive - years after its controversial life and what insights the case could provide for sustainable infrastructuring. This question is most directly addressed in Article III that interrogates the ‘call’ of the DAMD, the stakes involved and lessons for the social sustainability of future data infrastructures. Whereas much of the analytical use of the DAMD case was due to its failure, the platform startups provided a contrasting case to explore how ongoing efforts are trying to tackle this question of sustainable infrastructuring. As part of a broader movement in the Nordic region, the platform startups proposed emerging ideas of individual control that raise important questions about how, and why, claims to ethics are enrolled in the making of infrastructures.

Following the phase of preliminary mapping, I began *case-specific fieldwork.* The fieldwork activities in this phase were carried out in a *targeted* mode of inquiry beginning in July 2017 for the DAMD case and August 2017 for the platform case study. The choice of methods is reflective of the temporal and spatial differences posed by each case study. For instance, participatory observation is useful for studying infrastructures in the making (e.g. Karasti and Blomberg, 2017; Anand, 2018; Ratner and Gad, 2019) - or un-making even (e.g. Yarrow, 2017) - but less so for those that are already unmade or no longer with us. Given that the controversial life of the DAMD ended in 2015, I chose a selection of methods more suitable for the historical research, such as interview, document analysis (e.g. official documents, news media, professional journals) and Twitter data. The targeted study of the platform cases took place at various sites across the Nordic region through participatory observations, interviews and document analysis.

In this section, I overview the methods I used, including those that are specific to some papers. I describe how I approached the field, and offer reflections on my positionality.

No.	Start date	Days	Research focus	Event title	Host	Location
1	2016-11-24	1	National debate: DAMD	Data og Registerforskning	Dagens Medicin	Copenhagen
2	2016-11-05	1	Health data innovation	Datathon for Diabetes	Novo Nordisk, Deloitte, IBM, and the IT University of Copenhagen	Copenhagen
3	2017-01-19	1	Health data innovation	Disruption på Vej i Sundhedsvæsenet	MedTech Innovation Consortium	Aarhus
4	2017-02-09	1	Platform building: Sundhedsdatabank	Kunstig Intelligens i Sundhedssektoren	Enversion	Aarhus
5	2017-03-14	1	Health data innovation	Bring Sundhedsdata i Spil	WelfareTech Innovation Network	Aarhus
6	2017-03-23	1	Health data innovation	CHC Innovationsdag	Copenhagen HealthTech Cluster and Danish Diabetes Academy	Copenhagen
7	2017-03-28	1	National debate	Sundhedsdata - Livredder eller Livsfarligt?	Sundhedspolitisk netværk, Samuelsen	Copenhagen
8	2017-04-05	1	National translation of GDPR	GDPR - Hvad med de ustrukturerede data?	CSC and Hitachi	Copenhagen
9	2017-05-22	1	Preliminary fieldwork: DfG	Sundhedsdata og Kunstig Intelligens til Gavn for Sundhedsvæsenet og den Enkelte	WelfareTech	Copenhagen
10	2017-06-14	1	Health data innovation	The Data-driven (R)evolution	The Confederation of Danish Industry	Copenhagen
11	2017-08-30	3	Platform building: Dig.Me	MyData 2017: Advancing Human-Centric Personal Data	MyData	Talin and Helsinki
12	2017-11-15	1	Platform building: DfG	Hvem Ejer Borgernes Data?	Data for Good	Copenhagen
13	2017-11-21	1	National translation of GDPR	Bliv Klar til Persondataforordningen	Innovation Network for IT	Aarhus
14	2017-11-23	2	Platform building: Cima Technologies	MyData Aarhus Meetup	MyData	Aarhus
15	2018-01-26	1	Professional debate: DAMD	Den Postfaktuelle Læge	Organization of Danish Medical Societies	Copenhagen
16	2018-03-19	1	Health data innovation	Health Data Initiatives - Learnings from Finland	Copenhagen HealthTech Cluster/Data Saves Lives and the University of Copenhagen	Copenhagen
17	2018-04-11	1	Professional debate: DAMD	Deling af Patientdata	Dagens Medicin	Copenhagen
18	2018-08-29	3	Platform building: incl. Digi.Me and DfG	MyData 2018: From Vision to Action - the Future is Now	MyData	Helsinki

19	2018-10-11	3	Platform building: DfG, HealthD360	Week of Health Innovation 2018	Welfare Tech, Odense Universitetshospital OUH, Odense Kommune, Region Syddanmark, Healthcare DENMARK og Syddansk Universitet	Odense
20	2018-11-13	1	Health tech innovation	Sundhedsinnovation i Fællesskab	Copenhagen Healthtech Cluster, Copenhagen Health Innovation and CACHET	Copenhagen
21	2019-01-17	1	Civil society debate: DAMD	Når algoritmer styrer dine sundhedsdata	Danish IT Society	Copenhagen
22	2019-03-22	3	Platform building: Digi.Me	Nordic Health hackathon	Nordic Health hackathon	Reykjavik
23	2019-03-29	3	Platform building: Digi.Me	Nordic Health hackathon	Nordic Health hackathon	Helsinki
24	2019-09-25	3	Platform building: incl. Digi.Me and DfG	MyData 2019: Rebuilding trust - for human-centered data economy	MyData	Helsinki
25	2020-09-21	1	Platform building: DfG	Data for Good Summit	Data for Good Foundation, The Confederation of Danish Industry and Tech der Tæller	Copenhagen

Notes

Infrastructure actors of primary ethnographic interest are stated in extension of the empirical focus where relevant.
DAMD relevant events (n=4).
Data-intermediary relevant events (n=11).
Other events, used to map the field and relevant themes (n=9).

Table 2. Events researched through participatory observation, ordered by date.

No.	Start date	Days	Research focus	Event title	Host	Location
1	2015-01-23	1	National debate: DAMD	Kampen om fremtidens sundhedsdata*	The Organization of Danish Medical Societies	Copenhagen
2	2016-09-16	1	Platform building: DfG	Data Control - monopoly or individual control	Data Ethics	Copenhagen
3	2016-10-04	1	Health data innovation	Week of Health Innovation 2016*		Odense
4	2017-06-15	1	National debate	Folkemødet 2017: Sundhedsdata - vores fælles skattekiste?	Copenhagen Health Cluster, IBM and Region Denmark	Bornholm
5	2017-09-29	1	Platform building: DfG	European DataEthics Forum 2017	DataEthics and Danish IT Society	Copenhagen
6	2017-11-29	1	Platform building: HealthD360	Mere Borgernær Velfærd med Sundhedsdata	The Liberal Party of Denmark (Venstre) and Aarhus University	Copenhagen
7	2018-02-08	1	Platform building: Sundhedsdatabank	Sundhed og Kunstig Intelligens '18	Enversion	Aarhus
8	2019-02-18	1	Platform building	Kick-off arrangement for HealthD360*	Project partners including Aarhus University	Aarhus
9	2019-05-16	1	Health data innovation	Data Redder Liv konference 2019*		Copenhagen
10	2019-11-04	1	Health data innovation	Nordic Health Data: How do we collaborate on mapping metadata?*	Nordic Health Data Collaboration/Confederation of Danish Industry	Copenhagen
11	2020-06-24	1	Health data innovation	Data sharing for a more self-sufficient healthcare system.	Nordic Innovation	Copenhagen
12	2020-09-28	1	Platform building: DfG	Åbenhed og privatliv kan kombineres	The Confederation of Danish Industry	Copenhagen
13	2020-10-06	1	Platform building: DfG	Borgeren kontrollerer selv sine data	The Confederation of Danish Industry	Copenhagen
14	2020-10-20	1	Platform building: DfG	Privacy og blockchain	The Confederation of Danish Industry	Copenhagen
15	2020-10-27	1	Platform building: DfG	GDPR samtykke - udfordringer og alternativer	The Confederation of Danish Industry	Copenhagen
16	2020-11-03	1	Platform building: DfG	Forretningsudvikling ved hjælpe af sundhedsdata	The Confederation of Danish Industry	Copenhagen
17	2021-02-03	1	Platform building: DfG	Online Morgenbooster: The new data economy: Own you data	1508	Copenhagen and Bangkok
18	2021-06-08	1	Platform building: DfG	Meet & Inspire: Etiske data i arbejde	Dansk Metal	Copenhagen

Notes

Event titles concluding with an asterisk symbol (*) indicate that the event was not researched through video recordings (n=4).
DAMD relevant events (n=1).
Data intermediary relevant events (n=12).
Other events, used to map the field and relevant themes (n=5).

Table 3. Events researched online through documents and video recording, ordered by date.

No.	Date	Interview type	Organisation	Position	Location
1	2017-02-09	Informal	Enversion	CEO & Founder	Aarhus
2	2017-02-10	Semi-structured	The Danish Health Data Authority	Director General	Copenhagen
3	2017-03-28	Semi-structured	Kite Invent*	Founder	Copenhagen/Aarhus
4	2017-04-04	Semi-structured	Copenhagen HealthTech Cluster/Copenhagen Capacity	Business Development Manager and Project Manager	Copenhagen
5	2017-06-09	Semi-structured	Data for Good Foundation	Founder	North Zealand
6	2017-04-09	Semi-structured	Danish General Medicine Quality Unit (DAK-E)	Former director	Odense
7	2017-09-30	Informal	Lynkeus	Project manager	Tallinn
8	2017-09-30	Informal	Cozy.IO	Online marketer	Helsinki
9	2017-09-30	Informal	Digi.Me	CEO	Helsinki
10	2017-11-24	Informal	Datafund	Chief Operating Officer	Helsinki
11	2018-08-29/31	Informal	The Digital health Revolution	Professor and MyData activist	Helsinki
12	2018-08-29/31	Informal	Data for Good Foundation	DfG affiliate	Helsinki
13	2018-08-29/31	Informal	Nordic Innovation	Senior innovation advisor	Helsinki
14	2018-08-29/31	Informal	Open Humans Foundation	Executive director	Helsinki
15	2018-08-29/31	Informal	HealthBank	Director of Research Partnership	Helsinki
16	2018-04-29/31	Informal	HealthD360	Project Manager	Copenhagen
17	2018-10-10	Informal	Sundhed.dk		Odense
18	2018-10-10	Informal	Data Donor	Representative	Odense
19	2019-03-22/24	Informal	Dattaca Labs	CEO	Reykjavik
20	2019-03-22/24	Informal	Students	Hackathon participants	Reykjavik
21	2019-03-22/24	Informal	Students	Hackathon participants	Reykjavik
22	2019-03-22/24	Informal	Students	Hackathon participants	Reykjavik
23	2019-03-22/24	Semi-structured	Digi.Me	Senior officer	Reykjavik
24	2019-03-22/24	Informal	Digi.Me	Software engineer	Reykjavik
25	2019-03-22/24	Semi-structured	Arion Bank	Representatives	Reykjavik
26	2019-03-22/24	Semi-structured	Nordic Innovation	Representative	Reykjavik

27	2019-03-22/24	Informal	Startup Iceland	Founder	Reykjavik
28	2019-03-29/31	Informal	Digi.Me	Software developer	Helsinki
29	2019-03-29/31	Semi-structured	Nordic Innovation	Senior Innovation Adviser	Helsinki
30	2019-03-29/31	Informal	Digi.Me	Software developer and former entrepreneur	Helsinki
31	2019-04-03	Semi-structured	Taltioni	CEO	Helsinki
32	2019-05-10	Semi-structured	Kanta Personal Health Record*	Product owner	Copenhagen/Helsinki
33	2019-09-25/27	Informal	Data for Good Foundation	Founder	Helsinki
34	2019-09-25/27	Informal	Universal Basic Data Income App	Co-founder	Helsinki
35	2019-09-25/27	Informal	Diabetes Services	Founder	Helsinki
36	2019-09-25/27	Informal	MyData	Team members	Helsinki
37	2020-09-21	Informal	Data for Good Foundation	Booth representatives	Copenhagen

Notes

Organisation titles concluding with an asterisk symbol (*) indicate that the interview was conducted online.
All semi-structured interviews have been recorded of which primary interviews have been transcribed.
Informal interviews have been documented in note form.
DAMD relevant interviews (n=2)
Data intermediary relevant interviews (N=33)
Other interviews, used to map the field and relevant themes (n=2).

Table 4. Interview conducted during fieldwork[13]

---

[13] The informant's position in interview number 23 has been revised following the examination committee's assessment due to GDPR considerations.

4.1 Entering the field

From the very beginning of the project, I was privileged to be handed ample leeway to choose my empirical and analytical focus based on my ethnographic encounters in the field. In order to understand the explorative approach used for this thesis, it is first useful to establish how I understand the term ‘field’ since its meaning can differ from one research tradition to another. In natural sciences, for instance, the field is often conceptualised as something ‘out there’ where objective truth can be ‘discovered’ through rigorous use of scientific method. While the ideal of scientific rigour remains valued, the epistemological underpinnings for this study are informed by a constructivist tradition of research. As the late Renée Claire Fox puts it, a prominent sociologist of medicine, ‘Aspiring to enter the field devoid of any preconceived ideas of what one may be looking for, or might expect to find, in order to fulfill a purist conception of scientific objectivity and impartiality is neither warranted nor realistic’ (Fox, 2004: 311).

The intellectual critique of scientific objectivity was arguably most vocal during the so-called “science wars" of the 1990s when scientific realists and social constructivists disputed the basis of scientific authority and legitimacy. Although the controversy had a tendency to seek polarisation, it is simplistic to reduce the beliefs of either camp to one uniform position. In truth, there were internal disputes too - as those familiar with the genealogy will know (see e.g. Heaphy, 2020) - which demonstrates that the picture is much more nuanced. As such, constructivism is better understood as a spectrum along which various positions are available. The ethnographer is to an extent always implicated in the construction of the field itself, with Donna Haraway (1988) proposing through ‘situational knowledges’ (1988) a demarcation of one’s own position: impartiality, "view from above, from nowhere" (589), ‘feminist empiricism’ as an alternative to radical feminism. Returning to Fox,

> A complex, two-way process is called for, which is both inductive and deductive and through which pertinent sociological concepts, ideas, and information are carried into the field and are activated both by the observational, interview, and documentary data being gathered and by the process of sociological reasoning in which one is trained….a deeper internalization of sociological thinking than usual is a requisite for doing the kind of "on the hoof," running analysis and interpretation of continuously inflowing data that fieldwork demands. (Fox, 2004: 311)

The articles resulting in this thesis reflect this process, responding to the way the field formed through my encounters with it and pursuit of valuations, tensions and themes pertaining to the quest of data-intensification. Although Actor-Network Theory (ANT) does not use the ecological terminology of ‘systems’, it does employ ‘networks’ that similarly stresses the notion of interconnectivity. The mode of wordly engagement is based on the principle of generalised symmetry in which the world is made up of heterogeneous networks of human and non-human actors. This is an anti-essentialist position that for instance rejects the apriori separation between abstractions - nature, society, culture and technology - and thereby their deterministic use to explain complex phenomena. Instead, ANT provides a co-constructivist approach that strives to understand complex phenomena by attending to *how* things relate on the basis of inductive empiricism without taking for granted the deductive powers of abstract entities (Latour, 2005).

Positionality

While Danish society is ‘native’ to me, since I have been a resident in the country since early adolescence, in some ways, I have been a ‘stranger’ to this ethnographic study of health data. From previous research, I had already conducted quasi-ethnographic research on the practices and politics involved with national data initiatives in Denmark. The knowledge I acquired through this research was however predominantly acquired in relation to the formulation and implementation of *open data* policies in the public sector. There are continuities between my past and doctoral research, in particular my ongoing focus on the political ambition in Denmark to make public sector data accessible for reuse to more actors and new purposes.

This observation is not exactly a remarkable finding considering that health is one of many sectors that open data initiatives. This was however not the case with my prior research, which focused on the reuse of open data to strengthen democratic ideals of good governance and business innovation. This background informed my choice of topic, and through early observations I was not only about getting a feel of the field; I was also finding a field which to me included case studies of “the next big thing” in regards to the deployment of big data in Danish healthcare.

4.2 Ethnographic Approaches in Anthropology and STS

What I borrow from the disciplines of anthropology and STS is a tradition of being ethnographic toward technologies (e.g. Bruun, Wahlberg, 2022). This thesis is not built on long-term ethnography in a single place – as commonly associated with the classical model of anthropology – but a methodological assemblage required to answer the question I set out to study. In this sense it builds on what Wahlberg calls Assemblage Ethnography to make sense of an evolving ecosystem based on fieldwork that engages different scales, sites and practices (Wahlberg, 2022). Entering the field, this fieldwork was empirically driven by an explorative quest to identify actors, sites, tensions and themes of salient importance.

The central research approach of this dissertation is ethnography. With its origins in anthropology, ethnography has deep roots in a discipline that held the scientific study of mankind as its primary objective. Ethnography is perhaps most famously known for its ‘thick descriptions’ (Geerts, 1973), which is valorised in anthropological research as the primary means of knowledge production to understand social and cultural life. Classical works in the anthropological canon are usually authored by European/Western explorers who travel to a remote location - e.g. the Trobriand Islands (Malinowski, 2001), the Sudan (Evans-Prichard, 1928) and Bali (Clifford Geertz, 2000) - to conduct descriptive, analytical and reflective studies of the people who live there. Although participatory observation is known as ‘the core modality of ethnographic inquiry’ (Fox, 2004: 310), the ideal of immersive fieldwork was also sought through other classical techniques ranging from interview to the collection of cultural artefacts.

However, with the ongoing datafication of everyday life, the method of ethnography has travelled along as we look to understand the practical implications of data in qualitative terms. In the reflections on *Ethnography for a data-saturated world* Hannah Knox and Dawn Nafus argue that the relevance of ethnography ‘is more than just empirical detail that can provide a reality-check on otherwise hyped phenomena’ (2018: 3). Done the right way, the authors continue, ethnography ‘also holds the promise of generating a new way of theorising and understanding digital data by building novel analytical concepts that are appropriate to the kinds of relations of knowledge production that digital data itself entails’ (ibid., see also Hoeyer and Winthereik, 2019; Douglas-Jones, Walford and Seaver, 2021; Lippert and Douglas-Jones, 2019 for other reflections on the utility of ethnography for data studies). Today, the merit of ethnography is recognised by a variety of research fields beyond anthropology.

One of the many fields to embrace ethnography is STS, where it has over the past two decades become a central research method (Lippert and Mewes 2021). A key contribution from this perspective is its capacity to bring the role of technology into view in social inquiry, which is mainly accredited to ANT (Callon, 1986; Latour, 1987; Law, 1987). Along with other non-human actors, technology and humans are analytically treated on equal terms based on the philosophical assertion that both actors shape socio-political change (Latour, 2005). The point is not that technology has agency per say, it is rather to frame agency as an empirical question to be examined in practice. Embracing that technology is inherently social, scholars in STS and beyond have for decades been attentive to the ways that values are – intentionally or not - inscribed in technology and its implications for the delivery of health and lives lived (Mackenzie et al., 2013; e.g. Bowker and Star, 1999; Winthereik, van der Ploeg, & Berg, 2007). This sociotechnical approach has in a similar fashion furthermore been employed to advance inquiry into data and platforms in various domains of digital health to engage questions of political economy (Guyer, 2016; Lucivero and Prainsack, 2015; Lupton, 2014; Sharon and Zandbergen, 2017; Vezyridis and Timmons, 2017). In this thesis I have drawn on this tradition and its questions to centre emerging ecosystems for digital health innovation, showing that they are not insignificantly driven by ideas of commodification, platformisation and commercialisation (see Article V).

In Denmark alone, there are numerous case studies that draw on ethnography to do empirical and conceptual work to make sense of the contemporary hype that on occasions seems to stick to data. This is, for instance, illustrated in Hoeyer’s (2019) careful unpicking of the epistemological claims embedded in public health policy, which additionally demonstrates how ‘promissory data’ can become valuable for the unexpected purpose of postponing action. Michael Hockenhull and Marisa Cohn (2021) provide another example from the Danish tech scene based on participatory observations that in part overlap with the doctoral study at hand. The authors propose the notion of ‘hot air’ as an analytical device to consider in serious terms how hype plays into imaginaries of futures in ways that matter (Hockenhull and Cohn, 2021). Both studies provide descriptive and analytical contributions on practices pertaining to data-intensification in Denmark. The commonality between the two studies is however not limited to the contribution as they both demonstrate the utility of what Delgado et al. (2014) describe as ‘multi-event ethnography’ that has informed the research for both publications, as it has done with the thesis at hand.

Crudely put, multi-event ethnography brings traditional tools of ethnography to pluralistic settings in an attempt to facilitate a constructive engagement with the geographical and temporal challenges produced by this move. This ties to the explorative approach of the thesis in that it is 'the actors themselves [who] define the boundaries of the empirical field. Actors living in the phenomenon are the ones putting together these events, which reduces the problem of defining where the field is and who the "natives" are' (Delgado, 2014: 6).[14] Following the infrastructuring of data intermediaries has for instance taken my fieldwork to multiple events across the Nordic region with a brief detour to Estonia. This provides a chance to observe how actors conduct and present themselves in plural settings composed of social arrangements that change from one event to another. The output of this fieldwork is illustrated in Article V that explores the role of infrastructures as Digi.Me and Data for Good Foundation promote themselves as 'the ethical choice' to different audiences. The analysis describes how engagement with the idea of data ownership develops for the two startups *and* their in MyData community with the progression of annual gatherings and other events. The longitudinal changes in the MyData community are also described in the Article I that furthermore details some of the ethnographic modes, techniques, practices and reflections concerning the type of events engaged in the present thesis.

With the case studies chosen for this thesis, and the structure of their social worlds, my research engages the promise of data infrastructure in the intersection of public health and tech industry. What ethnography offers to this thesis is then not only a method of data collection, but a methodology that extends beyond fieldwork to guide the broader process of knowledge production. Among the ethnographic studies that have inspired work done in this thesis is research regarding how data are sourced (Hoeyer, 2016) stored (Maguire and Winthereik, 2019; Hogan and Vonderau, 2019), imagined and valued (Lehtiniemi and Ruckenstein 2019; Hoeyer, 2019), disputed (Wadmann and Hoeyer, 2018) and clinically engaged (Langstrup, 2018). These are all questions that I engage in the article collection in this thesis where the approaches offered in aforementioned studies have fostered methodological reflection on how to approach my own study.

Interviews and Participant Observation

During the study, I conducted twelve interviews in adherence to a *semi-structured* approach (Kvale, 1999; Hastrup, 2003). The first two interviews with the Danish Health Data Authority and a Health Tech Cluster were conducted as part of a preliminary field mapping to help corroborate the focus of this dissertation conceptually and empirically. In contrast, the remaining interviews targeted informants in leading positions on selected platform initiatives included in the research design of this study.

Aside from interviews, I have also conducted participatory observations at 18 events across Denmark, Finland (n=4), Iceland (n=1) and Estonia (n=1). At these events, I had a range of roles, from attendee through shadow ethnographer to workshop facilitator. At all, I introduced myself as a researcher from Denmark, writing a PhD on data infrastructures in the health sector.

---

[14] As mentioned, the ethnographer is however also implicated in the field and must as Strathern famously phrased it make decisions on *Cutting the Network* (1996).

Approximately one-third of the events were attended in an explorative mode to inform my preliminary field mapping of the Danish health data scene. 4 events are related to the DAMD case that all took place in Denmark. These events were generally attended by established players from the Danish biomedical sector who debate the future of health data mainly in professional and political settings. 11 events are linked to the empirical study of infrastructural platforms in Denmark (n=6), Finland (n=4), Iceland (n=1) and Estonia (n=1). These events were attended by platform builders and related stakeholders that usually took place in innovation networks but also included hackathons and political meetings. Last, the thesis is also informed by a Twitter dataset as an online record of the public controversy of DAMD. Collected in the Summer of 2017, the dataset includes 1363 tweets dating back to late September 2014 when the DAMD became a matter of public concern.

4.3 Encountering the DAMD case

My original encounter with the case that formed my first case study was at a conference entitled *Data and Registry Research* in the capital of Denmark. The conference was attended by professionals and decision makers across the Danish biomedical sector and framed with an objective of creating clarity about the use of health data for research and its contribution. Being new to the field, I took notes of narratives and themes that I today have become accustomed to expect at similar events, such as how “unique” Danish health registries are, how this national advantage is insufficiently leveraged and how other nations and the private sector are ”catching up”.

One narrative in particular stood out, as it increasingly seemed to be the premise of the conference itself: Denmark was at risk of entering a crisis of trust in public institutions, which could impinge negatively on public attitudes to data sharing (which followed to its logical conclusion would undermine the very foundation upon which the “unique” landscape came into being). At stake was the preservation and potential expansion of the Danish health data ecosystem in which the present actors held vested interests in the dominant political line of data intensification.

It was in this context that I encountered “the DAMD case", as practitioners referred to it, which as far as I could recall was a public controversy about patient data that had reached settlement two years earlier. Yet, the reverberations of the controversy were clearly still observable as panellists operationalised the case as an explanatory device to make sense of the status quo, but interestingly in contrasting ways. Some argued that the DAMD case demonstrated yet another case of institutional incompetence or misconduct regarding the handling of sensitive health data. Others suggested that the locus of distrust was not necessarily founded in reality but in a *perception* of reality due to the “very, very critical approach” adopted by certain public opinion makers. Perhaps what surprised me the most was the tense and, at times hostile, atmosphere at the event. During the panel debate certain “critical” voices (who became vocal in response to the DAMD case) in advocacy of greater self-determination,[15] transparency and security for the handling of health data seemed to be ridiculed on stage by another panel member, which was met with chuckles from a considerable proportion of the audience. Puzzled and somewhat annoyed by the

---

[15] Which in itself can challenge benefits currently gained from the solidarity-based welfare model underpinning the collection and use of health data (Hoeyer, 2018: 145)

unconstructive tone of debate, I left the conference convinced that in order to advance my research I had to take a step backwards to understand the controversy of DAMD and its significance for the field.

Tracing the DAMD controversy through Twitter

This new research trajectory however invoked a question of methodological importance: How do you ethnographically research phenomena that still inform your field but happened in the past? In search of answers I took to Twitter based on the following specificities. First, the utility of Twitter for researching sociality and politics - the formation of publics, issues and political expression - is well documented (Rogers 2014: xix) and often organised through hashtags which exists for the DAMD case ("#DAMD") (Bruns and Burgess 2011: 3). Despite having a significantly weaker market penetration in Denmark compared to Facebook (StatCounter Global Stats 2018; DR Medieforskning 2015: 21), the social interaction between users on Twitter is, secondly, for the most publicly accessible. Third, the Twitter users include politicians, journalists and other influential actors whose online views could prove useful to analyse in the given case (DR Medieforskning 2015: 38). Last but by no means least, the digital disposition Twitter potentially offers an environmentally friendly habitat for 'traces' to emerge, thrive and ultimately endure (Rogers 2009: 10).

In recent work, Albris et al. (2021) build a case for the ethnographic utility of digital methods to the target audience of anthropology. The authors propose three ways by which anthropologists can benefit from venturing beyond the comfort of their traditional methods of which two are demonstrated in the thesis. The first proposes that 'scraping social media complement ethnographic research by providing a broader contextualization' while the second suggests that 'quali-quantitative methods can create 'thick' datasets that are non-traditional yet ethnographically salient' (Albris et al., 2021: 2; see also e.g. Latour and Venturini, 2010). Before I demonstrate the ethnographic utility of quali-quantitative methods, I will first attend to the method by which I acquired a case specific Twitter dataset on the public controversy of the DAMD.

While it was possible to conduct a qualitative reading of tweets included in the #DAMD debate through Twitter's Search User Interface (Search UI), I found the format by which they were presented suboptimal for the purpose of analysis. I was already acquainted with digital methods and therefore knew of its opportunities for assisting ethnographic inquiry (Rogers, 2009; Ruppert, Law, and Savage 2013; Latour and Venturini, 2010; Venturini, 2010; Moats and Moats, 2015; Wang, 2013). This would however require access to Twitter data, meaning tweets and related metadata in a machine-readable format, which is typically acquired through Twitter's Application Programming Interfaces (APIs). I had previously captured Twitter data through institutional access to DMI-TCAT, a non-commercial web application developed by Erik Borra and Bernhard Rieder (2014: 262). Among other things, DMI-TCAT enables researchers to capture tweets based on key terms (including hashtags) from Twitter's public APIs without the need for programming languages such as 'R' or 'Python'.

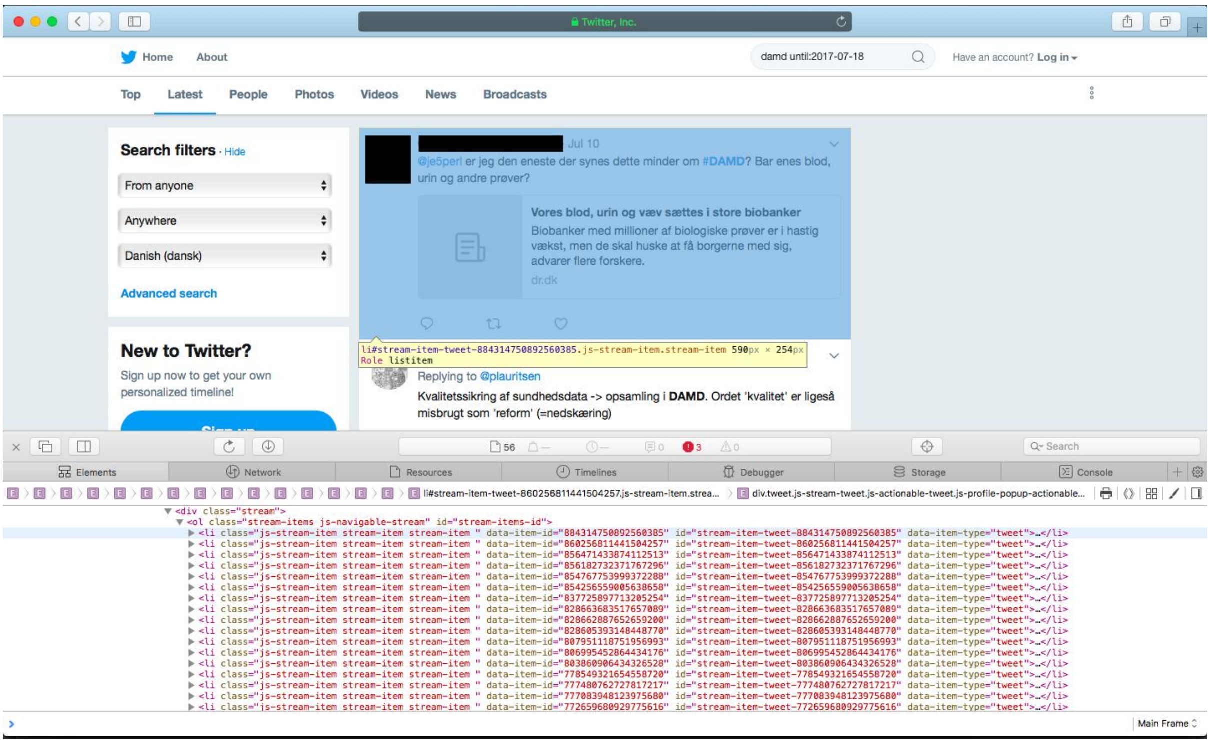
The difference this time however was that I needed to capture tweets engaging a controversy that took place two or three years in the past rather than an ongoing debate or future event. As experienced by Borra and Rieder (ibid.: 268), the technical configuration of Twitter's public APIs only afforded the capture of Twitter data a week old and forward when queried as key terms, omitting an obfuscatory percentage of tweets in the process (‘Search Tweets’, 2018: para 1). Tweets exceeding the one-week margin needed to be purchased through a commercial API from Twitter directly or indirectly through a partner. The access regimes of social media APIs however also change over time, which in January 2021 exempted academic researchers from the premium fee previously required for longitudinal access to historical twitter data (@atornes and @leanne_tru, 2021). While this access reform is good news for contemporary internet scholars, I was in 2017 forced to seek another solution.

Initially I contacted a commercial service about acquiring twitter data for the search term 'damd' from early 2014 to late 2017 and received an estimation of USD 24790. In hope of a more affordable price, I inserted a language criterion in my 'damd' query to include Danish tweets only. This was however in vain, as the new estimate was USD 0 recording 0 "activities", which I found hard to make commensurable with the +1000 tweets I would receive with the same query using Twitter's Search UI. Outpriced by the first option and puzzled by the second, I found myself in a situation "far from ideal", as the two tool developers put it (Borra and Rieder 2014: 268). To make matters worse, I found little solace - for myself and internet scholars more generally - in the authors’ continued assessment that the forward collecting method "[…]might be the only feasible way to record traces of an unanticipated event" (ibid.: 268). This prompts the questions about the emerging politics of platforms, including their capacity to shape which questions internet researchers interrogate (see Puschmann and Burgess, 2014: 45). Are scholars with limited financial support expected to anticipate relevant events within a one-week margin, including the adequate search terms or folksonomies required to capture it? Despite these obstructions, I managed to capture historical Twitter data related to the controversy without the use of commercial products, ultimately making way for a digital post-mortem analysis of the DAMD case. How this historical dataset was captured follows next.

Digital methods scholars have long promoted methodological creativity as defining characteristics of their community, as digital devices and traces for the mediation of sociality need to be repurposed for research (Rogers, 2013: 1; Weltevrede, 2016: 15). In response to Twitter’s policy restrictions against sharing tweets and metadata but not tweet IDs, DMI-TCAT enables researchers to reconstruct data sets from tweet IDs - old and new - as an alternative way of sharing twitter corpora (Borra and Rieder 2014: 268). By the same token, if a list of relevant tweet IDs related to DAMD can be identified, it should then in principle also be possible to reconstruct a data set from Twitter’s public APIs via DMI-TCAT or otherwise. The crucial question is then how to locate and compile a list of tweet IDs of relevance to the empirical case at hand?

Responding to the relevance criterion above, I had identified three key terms by probing Twitter’s Search UI, where the latter two were included to capture tweets by users who do not conform to hashtag practices:

- "#DAMD" (n=1956)
- "Dansk AlmenMedicinsk Database" (n=8)
- "DAMD", delimited to tweets in Danish (n=1319)



[redacted] [redacted] Jul 10

@je5perl er jeg den eneste der synes dette minder om #DAMD? Bar enes blod, urin og andre prøver?

Vores blod, urin og væv sættes i store biobanker
Biobanker med millioner af biologiske prøver er i hastig vækst, men de skal huske at få borgerne med sig, advarer flere forskere.
dr.dk

```
li#stream-item-tweet-884314750892560385.js-stream-item.stream-item 590px × 254px
Role listitem
```

Replying to @plauritsen

Kvalitetssikring af sundhedsdata -> opsamling i **DAMD**. Ordet 'kvalitet' er ligeså misbrugt som 'reform' (=nedskæring)

```
li#stream-item-tweet-860256811441504257.js-stream-item.strea…   div.tweet.js-stream-tweet.js-actionable-tweet.js-profile-popup-actionable…
```

```
▼<div class="stream">
  ▼<ol class="stream-items js-navigable-stream" id="stream-items-id">
    ▶<li class="js-stream-item stream-item stream-item " data-item-id="884314750892560385" id="stream-item-tweet-884314750892560385" data-item-type="tweet">…</li>
    ▶<li class="js-stream-item stream-item stream-item " data-item-id="860256811441504257" id="stream-item-tweet-860256811441504257" data-item-type="tweet">…</li>
    ▶<li class="js-stream-item stream-item stream-item " data-item-id="856471433874112513" id="stream-item-tweet-856471433874112513" data-item-type="tweet">…</li>
    ▶<li class="js-stream-item stream-item stream-item " data-item-id="856182732371767296" id="stream-item-tweet-856182732371767296" data-item-type="tweet">…</li>
    ▶<li class="js-stream-item stream-item stream-item " data-item-id="854767753999372288" id="stream-item-tweet-854767753999372288" data-item-type="tweet">…</li>
    ▶<li class="js-stream-item stream-item stream-item " data-item-id="854256559005638658" id="stream-item-tweet-854256559005638658" data-item-type="tweet">…</li>
    ▶<li class="js-stream-item stream-item stream-item " data-item-id="837725897713205254" id="stream-item-tweet-837725897713205254" data-item-type="tweet">…</li>
    ▶<li class="js-stream-item stream-item stream-item " data-item-id="828663683517657089" id="stream-item-tweet-828663683517657089" data-item-type="tweet">…</li>
    ▶<li class="js-stream-item stream-item stream-item " data-item-id="828662887652659200" id="stream-item-tweet-828662887652659200" data-item-type="tweet">…</li>
    ▶<li class="js-stream-item stream-item stream-item " data-item-id="828605393148448770" id="stream-item-tweet-828605393148448770" data-item-type="tweet">…</li>
    ▶<li class="js-stream-item stream-item stream-item " data-item-id="807951118751956993" id="stream-item-tweet-807951118751956993" data-item-type="tweet">…</li>
    ▶<li class="js-stream-item stream-item stream-item " data-item-id="806995452864434176" id="stream-item-tweet-806995452864434176" data-item-type="tweet">…</li>
    ▶<li class="js-stream-item stream-item stream-item " data-item-id="803860906434326528" id="stream-item-tweet-803860906434326528" data-item-type="tweet">…</li>
    ▶<li class="js-stream-item stream-item stream-item " data-item-id="778549321654558720" id="stream-item-tweet-778549321654558720" data-item-type="tweet">…</li>
    ▶<li class="js-stream-item stream-item stream-item " data-item-id="777480762727817217" id="stream-item-tweet-777480762727817217" data-item-type="tweet">…</li>
    ▶<li class="js-stream-item stream-item stream-item " data-item-id="777083948123975680" id="stream-item-tweet-777083948123975680" data-item-type="tweet">…</li>
    ▶<li class="js-stream-item stream-item stream-item " data-item-id="772659680929775616" id="stream-item-tweet-772659680929775616" data-item-type="tweet">…</li>
```

Figure 7. Screenshot of how Safari's Web Inspector can provide access to tweet IDs ("data-item-id") from the query 'damd', delimited to Danish

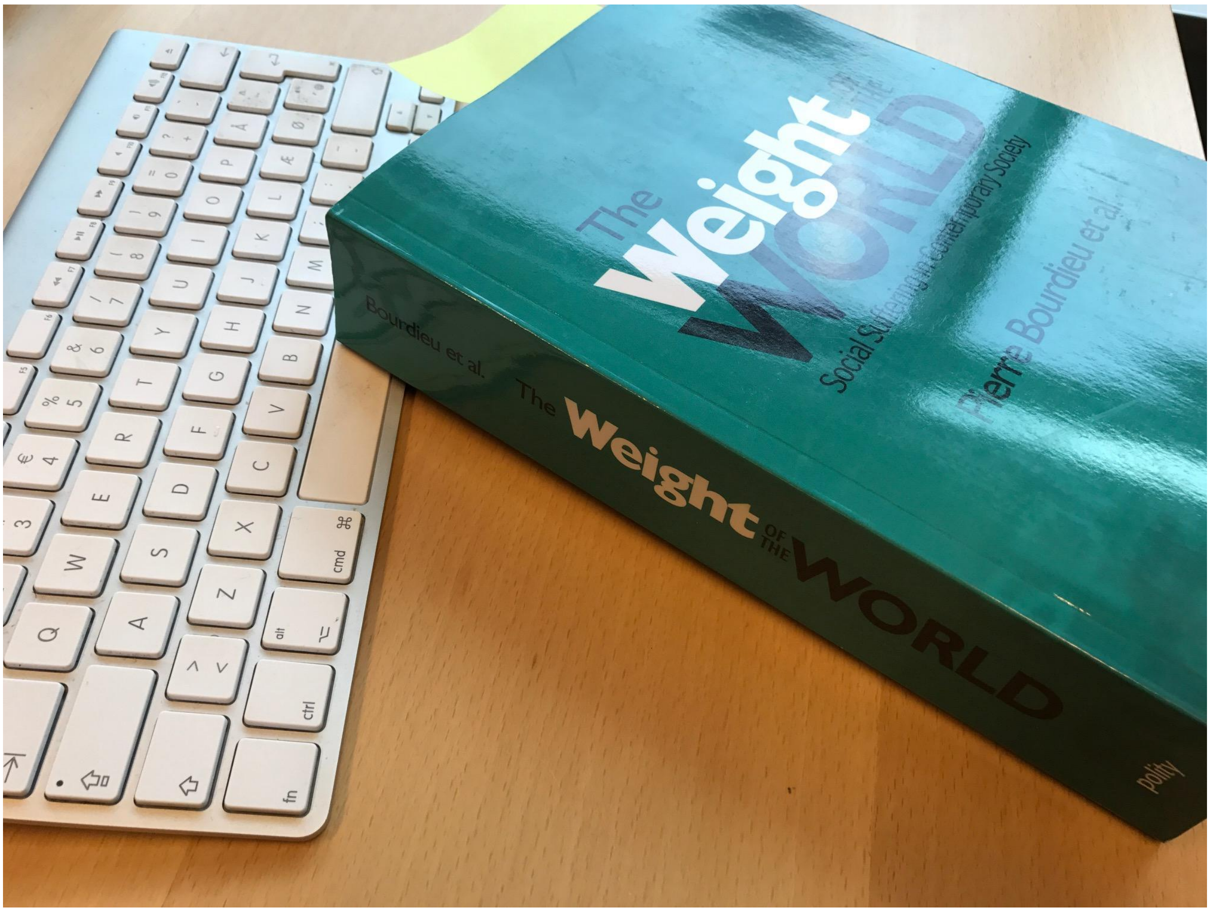
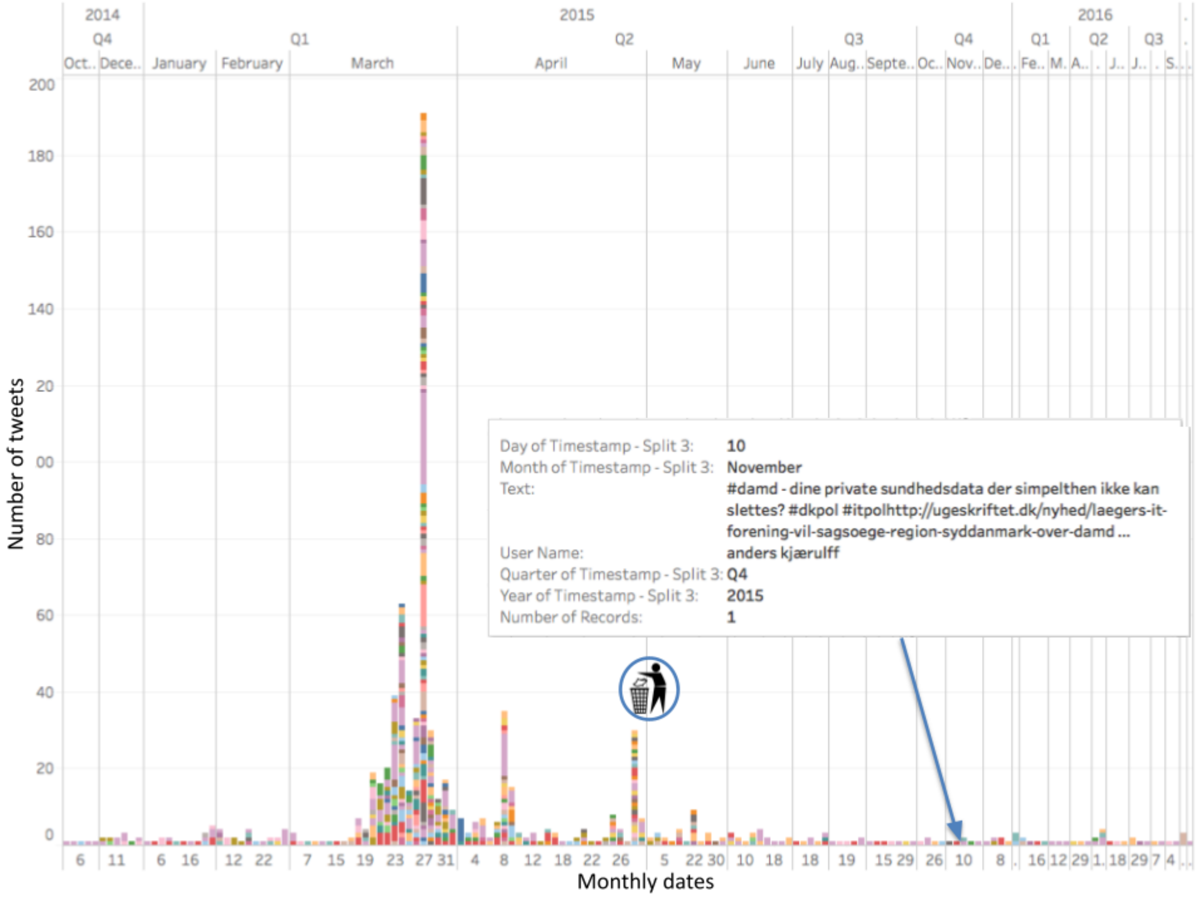


Figure 8. Documentation of method used to load query results in Twitter's Search UI



Day of Timestamp - Split 3: 10
Month of Timestamp - Split 3: November
Text: #damd - dine private sundhedsdata der simpelthen ikke kan slettes? #dkpol #itpolhttp://ugeskriftet.dk/nyhed/laegers-it-forening-vil-sagsoege-region-syddanmark-over-damd ...
User Name: anders kjærulff
Quarter of Timestamp - Split 3: Q4
Year of Timestamp - Split 3: 2015
Number of Records: 1

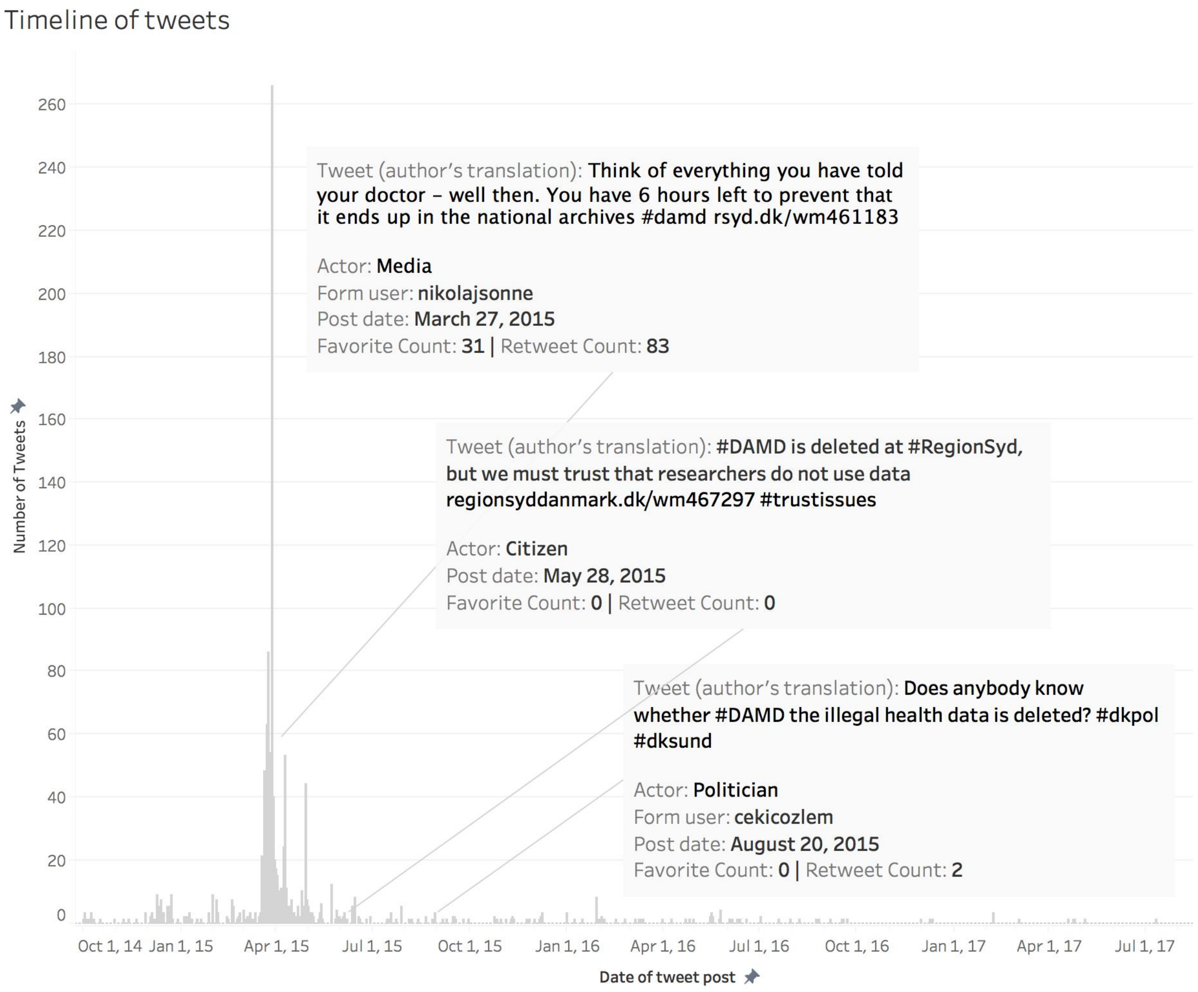
Figure 9. Presentation slide from an academic symposium in the UK demonstrating the questions identified with the quali-qualitative method through an early twitter data set. The data visualisation was produced with the software Tableau.

Having accounted for the collection of the twitter dataset, it is time to address its utility for ethnographic research. For this, I return to the quali-quantitative methods that allowed me to amass a quantitative overview of the dataset while retaining a trace to the qualitative content it represented. This paved the way for an interrogation of the dataset where questions provoked from a qualitative reading could be inspected qualitatively and vice versa. The screenshot provided in figure 9 provides an early illustration of this method in action where the visualisation tool, Tableau, enables an interactive inspection of the values assigned to each tweet with a click of the button. The contribution of going back and forth between the quantitative and qualitative modes of inquiry is most saliently represented in Article IV. In fact, the conceptualisation of this article is a direct result of the questions identified with this method. Figure 9 is taken from a symposium presentation (Burnett, 2017) based on a provisional exploration of a dataset with tweets on ‘#DAMD’ (N=826 tweets) prior to a refined cleaning.[16] Figure 10 exhibits a visualisation of the final database previously detailed that

---

[16] The ‘#DAMD’ dataset contains tweets between 3 October 2014 and 3 March 2017.

highlights some of the observations that made its way into Article IV through the questions that enkindled.



Tweet (author's translation): **Think of everything you have told your doctor – well then. You have 6 hours left to prevent that it ends up in the national archives #damd rsyd.dk/wm461183**

Actor: **Media**
Form user: **nikolajsonne**
Post date: **March 27, 2015**
Favorite Count: **31** | Retweet Count: **83**

Tweet (author's translation): **#DAMD is deleted at #RegionSyd, but we must trust that researchers do not use data regionsyddanmark.dk/wm467297 #trustissues**

Actor: **Citizen**
Post date: **May 28, 2015**
Favorite Count: **0** | Retweet Count: **0**

Tweet (author's translation): **Does anybody know whether #DAMD the illegal health data is deleted? #dkpol #dksund**

Actor: **Politician**
Form user: **cekicozlem**
Post date: **August 20, 2015**
Favorite Count: **0** | Retweet Count: **2**

Figure 10. Screenshot of a visualisation based on the final Twitter dataset. The visualisation was produced through Tableau and used to interrogate the DAMD case between October 2014 and July 2017.

As is often the case with the quantitative gaze, the first question I asked my immediate attention was drawn to the extremities in the visualisation caused by sudden rise and fall in the number of tweets. The question then became, what caused the explosive peak on 27 March, 2015? As described in Article IV, this date marked the deadline for the citizens to opt-out of the DAMD database before action would ensue to preserve a copy of the illegal data collection in the Danish National Archives. This was derived from a qualitative inspection of the tweets in and adjacent to the peak (in a spreadsheet viewing of the dataset) that was supplemented through a document analysis of other sources (e.g. news media, websites, social media groups) in part linked to in tweets. The peak represented a moment's urgency with details of civic mobilisation that otherwise would have remained silent if the date was strictly examined through news media.

Looking beyond the major peak, my attention shifted to a more subtle observation of something unexpected. From inspecting the more modest peaks, I knew that the tweets on

28 May, 2015, centred on confirmed destruction of the DAMD data that was the sentence of the Danish Parliament. While this news is somewhat expected considering the preceding twitter activity of the DAMD case, I was puzzled why this activity remains active after its confirmed destruction up till the very point of data collection more than two years later. How could this be? Again, this question is taken up in Article IV in a discussion of the certainty of deletion prompted by the public discovery of residual traces of the database. Similarly to the previous one, this second was also qualitatively explored within and outside the twitter dataset.

The ethnographic utility of the quali-qualitative method in the thesis can be understood in the affordances it provides for eliciting questions that can serve as a starting point for analysis. Seen this way, my engagements with the dataset have informed my understanding of the case more broadly, bringing actors, themes and issues to my attention in ways that would be difficult to attain otherwise. The understanding I have gained from the dataset is then a particular one that is specific to twitter that has informed the questions I have posed in interviews, the selection of sites I have attended and my overall comprehension of the DAMD case.

4.4. Studying emergent platforms in Nordic welfare states

How can one gain a sense of what emerging platforms are about? As my second case focuses on infrastructural initiatives that champion trending ideas of ‘individual control’ over data through different platform models, in my research I sought to understand the positional work that platforms perform in relation to the market they wish to enter - or in this case disrupt.

Among the spaces where this form of labour is visible is at the annual MyData conferences, which besides representation from government, civil society and academia is also attended by entrepreneurial platform builders. MyData is a Finnish based initiative with a global outreach in pursuit of a new economic order. I attended the MyData annual conferences three times between 2017 and 2020 where Article I contributes a critical reflection on the kind of work required of ethnographers at such events.

In this sense, MyData can be seen as an activism movement committed to achieve a more “just, sustainable and prosperous digital society” where contemporary imbalances of who benefits from personal data are distributed in a more equitable manner (MyData, n.d.: para 2). In this congregation, GAFA (Google, Apple, Facebook and Amazon) has come to epitomise such injustice due to their monopolistic position to access and benefit from personal data in ways that remain opaque for data subjects until harm is done (e.g. public scandals). This concern is also registered by European regulators, as noted previously.

Discontent with the status quo, MyData and associated platform initiatives seek to establish a new personal data ecosystem based on trust where individuals are ‘empowered’ to exercise digital rights for control and choice over the handling of their own data - and ultimately benefit. To this end, the European right to data portability introduced in May 2018 has been assigned special status for disrupting data monopolies since it in principle entitles individuals to acquire or move their personal data from one service provider to another in a machine-readable format when technically feasible (Article V). It is this exchange of information that the platforms in question look to facilitate by providing a single interface (or personal health account) through which individuals can import, store and "share" personal data from public and consumer services for third-party use based on dynamic consent.

Name	**Status**	**Construction**	**Nation**
Data for Good	Quasi-operational	Non-for-profit foundation	Denmark
Cima Technologies	Discontinued	For-profit compay	Denmark
Sundhedsdatabanken	Discontinued	Unknown	Denmark
HealthD360	Conceptual	Private-public partnership	Denmark
Digi.Me	Quasi-operational	For-profit company	Iceland
Kanta Personal Health Record	Quasi-operational	Governmental initiative	Finland
Taltioni	Discontinued	Non-for-profit cooperative	Finland
Hälsa for Mig	Discontinued	Governmental initiative	Sweden

Table 5. Provisional candidates for in-depth case study

While innovation networks, conferences and other platform sites constitute a promissory space for digital entrepreneurship, the development ecology can also be volatile. During the MyData 2017 conference track called the "Digital health revolution" a member from the audience announced in frustration the unexpected collapse of what was supposedly a pioneering Finnish platform Taltioni ('My storage' or 'My cache') in absence of a viable business model. Taltioni was launched as a non-for-profit cooperative in 2012 with 27 members but was acquired by an IT firm making Taltioni a for-profit endeavour. More recently the governmental eHealth Authority in Sweden discontinued the development of the platform Hälsa for Mig ('Health for Me') for reasons that include legal tensions regarding offshore storage of personal health data (Ny Teknik, 2018).

For data intermediaries, attending conferences was the primary mode of investigation to understand the startups, but also the social, cultural and economic climate they engage (see Article I for a more detailed discussion). Noting the themes, actors, discussion flows, the feel allowed me to study the configuration of platforms as infrastructures for new models of health data use. They also alloww s me to elucidate how values are thought and sought. While the MyData whitepaper stresses the importance of business models if ecosystems are to "flourish" (Poikola, Kuikkaniemi and Honko 2015, p. 8), finding one that is viable in the domain of health remains a challenge (see Table 5). The platforms operate according to different organisational constructions (non-for profit, for-profit, government) and revenue models (account and transaction fees, data trade, provision of secure storage and other service) that have implications for how the economic costs should be financed and benefits distributed.

5. Articles, Findings and Contributions

This thesis is built from five articles. Three emerge from case study one, and two from case study two. The health data ecosystem concept has informed their construction, offering a view of the various phases in the life cycle of data as artificial distinctions that in practice are reciprocal in nature. In addition, each demonstrates the ideas of interdependency within an emergent ecosystem.

5.1 Case Study 1: Articles, II, III and IV

The three articles emerging from the first case study address different aspects of the life and death of DAMD. The first, *Sentinel Vision,* asks how the software underlying DAMD was envisaged to improve diabetes care. Starting with the origins of the infrastructure that was built to support quality improvement within primary care, this paper uses the concept of professional vision to illustrate where data intensive logics emerged, and how they have been promoted to strengthen preventive healthcare. It contributes a historical perspective into what is considered as relevant health data has expanded, and a technical perspective on what it takes to build an infrastructure in collaboration with various state and private partners. The second paper, *The Call of Data* focuses on the expansion of the database, building on its origins in *Sentinel Vision*. It asks how personally identifiable data on three-quarters of the Danish population ended up in the DAMD database, when its authorization was technically limited to four chronic conditions. Putting forward the idea of digital hoarding, it contributes to CDS and suggests that not only did actors fail to adequately recognize the problematics of the DAMD, they exhibited significant will, both through action and inaction, to ensure the longevity of what –in hindsight– should have been negotiated. By describing the data collection through the lens of the 'call' of data, the paper offers an analytic to explore both the material and aesthetic dimensions of data's appeal, extending Bennet's posthuman project to the consumption of data and its social sustainability. The final paper, *Refusing Retention* looks at the moment the database could no longer be sustained. Its question, 'what to do with illegal and yet invaluable data collection, and who decides' tackles the sociotechnical question of data deletion. It argues that what becomes available for deletion emerges from intertwined institutional, technical and temporal processes, from both those who refuse retention and those who seek to make deletion happen. As an illustration of the problem of deletion in a digital age, the paper invites colleagues in STS and cognate fields to focus on the challenges of deletion as a counter to the promise-filled positivity of data accumulation.

Overall, DAMD was a site of tension within my fieldwork, as described above. The subject of technoscientific failure is hardly a novel concern for STS with Bruno Latour's (1996) Aramis, or the Love of Technology among the canonical works in the field (see also Law and Callon 1992; Scott 1998; Bowker and Star 1999). The mystery of "who killed Aramis?" – a failed experimental project to build a large-scale personal rapid transit system in Paris between 1969 to 1987 – is narrated as a (fictional) detective novel to develop ideas of ANT and its significance to advance social inquiry beyond "big explanations" (Latour 1996, 133). The conclusion Latour arrives at is that Aramis died from an insufficiency of love as key stakeholders lose interest in the project. The case of the DAMD, however, seems to suggest an inversion, where its fatal outcome was caused by too much love – particularly for its infrastructural capacity to accumulate data from a sector that, for technical, legal, and cultural reasons, had traditionally been a hinterland to the national health data landscape. Put simply, the DAMD was loved for what it attempted to see – and became blind in the process.

5.2 Case Study 2: Articles I and V

Article I *"Let's Make it Happen"* addresses the practice of fieldwork in tech conferences. We present tech conferences as sites where stakeholders in tech convene to *make things happen*, be it a smart kitchen appliance or a re-imagined ecosystem for personal health data. As a central method in the present thesis, conference ethnography has taken me to 'dense and rich sites', including the annual MyData gatherings where the ecosystemic vision for personal health data was engaged with a 'temporal rhythm of expectation that oscillates between hype and disappointment in alternating cycles' (Article I: 17). Attending the social life of these conferences was ethnographically fruitful, yielding insights into the socio-political work and questions involved with the making of data futures, along with the life and death of startups envisioned to make such futures happen. What these conference gatherings provide is then an ethnographic entrance to *emerging* sociotechnical imaginaries and makings of ecosystemic change to reconfigure the conditions of possibility for how data can be used and thereby also understood. The social life of tech conferences is in this sense intimately related to the social life of data that can potentially encompass the previously summarised conceptions of lively data and more. Article V more deeply interrogates the building of platform infrastructures for health data by asking how infrastructure and ethics are co-implicated. Following the two platforms, Digi.Me and DfG, I ask how they promote themselves as 'the ethical choice' and what role infrastructure plays. Based on participant observation, interviews, document analysis as well as analysis of online materials produced by the companies, I offer a descriptive analysis of how claims to ethics are also underpinned by infrastructural practices. My aim in the paper is to show how data intermediaries, as envisaged in European documents, are making advances in the Nordic region. What this paper provides is an overview of how data collection and re-use is envisaged, as part of the life cycle of data.

5.3 Contributions Synthesis

To synthesise the contributions of the articles across the five pieces, I now return to the four types of data liveliness I described earlier.

Beginning with the fluid ontology of data, Article V provides a case study in which two data intermediaries - DfG and Digi.Me - attempt to repurpose a growing array of personal data for health (e.g. the donation of retail data for the development of cancer treatment), but also personal health data for purposes beyond health (e.g. the monetisation of health insights for commercial marketing). A similar repurposing of data is also demonstrated in the case studies concerning the DAMD case. Article II for instance illustrates how data from electronic health records in general practice acquired new purposes (e.g. quality development, research, patient safety, patient empowerment, honorarium reform) with the development of Sentinel (see also Article III). Article IV demonstrates how the DAMD database even gained value for additional purposes after it was confirmed to contain an illegal data collection, as attempts ensued to preserve a database copy in support of - but not limited to - public accountability. The fluid nature of data exemplified in the articles goes to show why critical scholars have argued that what data are is 'always an empirically situated question' (Maguire et al., 2020: 14). Indeed, Article IV furthermore points out how the regulatory realms of jurisdiction can themselves struggle to answer whether the contentious data collection should be treated as clinical data, personal data or archival data.

The question of where data should live, and how, is described as one of paramount importance to the two core cases. Article V demonstrates attempts from data intermediaries to move data from conventional silos to personal storage solutions hosted by the platform or a third-party vendor. Article II and III similarly describes endeavours to move data from various electronic health records to one national database. The reallocation of *digital* data in both cases draws attention to the 'multiplicity' of data (Prainsack, 2019a: 13) since it largely results in a clone or duplication where data simultaneously reside in more than one place at once. This ability to not only mutate but *multiply* injects an additional element of complication to the aforementioned question of "what are data" and others concerning the ontology of data. Article IV demonstrates how multiplication can provide data powers of resilience, as residual traces of the illegal data collection were found after its confirmed destruction. What ensued was a moment of uncertainty in which actors questioned whether the DAMD was dead or alive.

Instances of the datafication of life can also be observed within the article collection. This tracing has in fact already been presented in the *Literature Review* section in relation to Article II and Article V. This ongoing process of datafication is in part a result of repurposing of existing data sources - as described above - but also the development of new ones. The new data sources are most salient in Article II where they pertain to the introduction of questionnaires for patients to answer under the facilitation of GPs in their encounters. Attempts to increase the participation of patients in the datafication of health is a development gaining traction in Danish healthcare through data intermediaries, such as DfG (Article V), but also national initiatives on patient-reported outcomes (Langstrup, 2018).

Seen from the history of statecraft (Hacking, 1991; Scott, 1998), discussions of the datafication of life will often pivot to the *regulation* of life. Underlying this transition is the constructionist notion that the datafication of something brings that thing into being as something newly visible - and therefore also potentially governable. The most convincing examples of this 'looping effect' (Hacking, 1996) or *Making up people* (Hacking, 2006) can be found in Article II, which moreover foregrounds the politics of data. In Article II, the diagnostic coding of diabetes patients is established as a prerequisite if they were to be 'counted' in the diabetes population overview provided by Sentinel. Whereas undiagnosed patients would be excluded from the system, those assigned the appropriate diagnosis would enter a regime of preventive risk management. The governance mode in this regime was at least initially one of self-governance with the promotion of continual improvement in the diabetes care for general practices and self-care for patients (e.g. the promotion of healthy lifestyle changes in terms of smoking or exercise). General practices would later be confronted with a regime of enhanced external control, as the DAMD became enrolled in a dispute between the community of general practitioners and the employee regarding the limits of state legibility. Article IV presents how the models of DfG and Digi.Me also promote solutions to their users for the self-management of health. Users are however not only responsibilised for their health, as users are also assumed to take responsibility for the self-management of privacy.

Moving onwards, the fourth conceptual take on lively data concerns how the growing value of personal data has implications for the livelihood of industry actors. This conceptualisation is most salient in Article V that describes how data intermediaries - along with selective stakeholders - take active steps to participate in the growing political economy of health data. With a characterisation of data as ‘the lifeblood of economic development’, data intermediaries have in the EU Data Strategy been identified among the levers to secure Europe’s place in the global data economy (European Commission, 2020: 2). While the circulation of data is considered critical for the securement of economic development, it has also been proved to be a matter of life and death for the data intermediaries otherwise anticipated to inject newly liberated data in the European economy. In order to establish themselves where others have fallen (e.g. Taltioni and Cima Technologies are among the startups that failed to find a viable business model), DfG and Digi.Me draw on their infrastructural credentials to promote themselves as “the ethical choice” to gain circulatory momentum. To this end, DfG encourages citizens to ‘donate’ personal data for the advancement of cancer treatment that notably appeases industry demand for health data access in a framing of economic growth as a public value. A more explicit link to livelihood is exhibited in *Universal Basic Data Income*, which is the name of an app powered and promoted by Digi.Me for ‘ethical data monetization’. In this example, citizens can allegedly gain an annual income of €1000 in exchange for third-party access to anonymised insights from self-curated data profiles.

Article number	**I**	**II**	**III**	**IV**	**V**
Article title	**"Let's make it happen!"** An STS ethnographer's guide to tech conferences	**Sentinel vision** Data collection, disease prevention and professionals in Danish diabetes healthcare	**The call of Data** A post-mortem of the Danish General Medicine Database	**Refusing retention** Health data and its deletion in the Danish welfare state	**Infrastructures in the promotion of ethical tech** The rise of data intermediaries in the Nordic market for digital health
Status	*Revise and re-submit*	*Pre-submission*	*Accepted*	*Accepted*	*Pre-submission*
Journal	Science & Technology Studies	Science Technology & Human Values	Big Data & Society	Social Studies of Science	Big Data & Society
Empirical case	Data intermediaries	The DAMD case	The DAMD case	The DAMD case	Data intermediaries
Research question	How does one study the world of technology conferences?	How was the DAMD envisioned to improve diabetes care?	How could personally identifiable data on three-quarters of the Danish population end up in the Danish General Practice Database when its authorisation was limited to four chronic conditions?	What to do with the dubiously, likely illegal and yet potentially invaluable, 'unique' data, and, more importantly, who was to decide?	How are infrastructures enrolled when Digi.Me and DfG promote themselves as 'the ethical choice' to users, the market and the communities they engage?
Key concept(s)	The sociality of tech conferences	Professional vision Sentinel	Hoarding The call of data	Data deletion Refusal	Ethics Infrastructure
Main method of ethnography	Part. observation	Document analysis Interview	Document analysis Interview	Document analysis	Document analysis Interview Part. observation
Argument	Tech conferences are potentially rich and dense sites for ethnographic fieldwork and that *how* they are attended by ethnographers matters.	Preventative logics accompanying increased data-intensification in healthcare today have longstanding roots, and suggest that it is generative to trace the coimbrication of professional values, technological solutionism and the building of infrastructure for preventive healthcare.	Not only did actors fail to adequately recognise the problematics of the DAMD, they exhibited a significant will - through action and inaction - to ensure the longevity of what in hindsight should have been negotiated.	What becomes available for deletion emerges from intertwined institutional, technical and temporal processes, from both those who refuse retention and those who seek to make deletion happen.	Infrastructures are also used in the promotion of ethical tech and more ethnographic research is needed to illuminate the infrastructural practices that underpin the ethical claims put forward

Main contribution	Method: Reflexive, ethnographically supported guidance for ethnographers newly entering the fields of technology research Empirical: Demonstration of how tech conferences can offer unique opportunities to engage with stories that circulate in and construct fields of technology	Conceptual: Extends the scholarship on Sentinel by attending to its originating metaphor, following its paths of illumination through different professional contexts. Invitation for scholars to trace the coimbrication of professional values, technological solutionism and the building of infrastructure for preventive healthcare. Empirical: Case study of how data-intensive logics are promoted to strengthen preventive healthcare	Conceptual: Development of the 'call of data' as an analytic proposed for engaging with data while extending Bennett's sustainability project to data. Empirical: Descriptive account of how the unlawful collection of population-wide patient data occured, illustrating how failure to navigate legal and social thresholds acceptability can undermine the social sustainability of data initiatives. Demonstration of how attempts to exploit wasted opportunities in data can themselves result in waste with long lasting effect for public trust.	Conceptual: Development of 'refusing retention' as an analytic proposed deletion that challenges the positivity of data retention as a default mode of valuing data. Analysis of the close relationship between data deletion and data retention, illustrating in part how the latter can (paradoxically) become politically desirable to achieve the former. Empirical: Empirical analysis of how data becomes valued, not to be kept but because it should be erased. Illumination of the political difficulties of deletion as a sociotechnical practice and emerging site of negotiation between states and citizens.	Empirical: A demonstration of how infrastructures are enrolled in the promotion of ethics and how ethnographic research on infrastructural practices can enhance understanding on the the ethical claims Descriptions that shed light on how ethics is operationalized not simply in the tech industry in a generic sense, but in the specific context of highly regulated countries in Europe through data intermediaries. In the wake of an ethical crisis, DfG and Digi.Me represent a burgeoning solution to restore trust for a continued intensification of data reuse. The solutions they provide however introduce additional questions of ethical and societal importance that will require continued scrutiny from scholarly and regulatory actors.

Table 6. Overview of articles with journal information and academic content

6. Concluding remark

With each year the reach of ‘datafication’ extends further into the lives of individuals and societies. Between 2018 and 2025, the volume of data is according to the International Data Corporation (Reinsel, Gantz and Rydning, 2018: 3) predicted to grow five times its size, from 33 to 175 Zetabytes. While the accuracy of such numbers are necessarily subject to uncertainty, the general trajectory seems sound and is for instance taken up in the European Strategy for Data (2020) to formulate policy priorities. “Each new wave of data“, the strategy states, “represents major opportunities for the EU to become a world leader in this area” (2020, 2). In this context, health is one out of nine areas of strategic importance where the EU aims to create “a common European data space - a genuine single market for data […]” (ibid., 4). At this point in time, it is undeniable that datafication offers new opportunities for socioeconomic progress that in many cases will present data-intensification as an axiomatic policy objective. The promises attached to data-intensification are however not easily attained - and more seldom sustained - in practice, which is documented at length (e.g. Vezyridis and Timmons, 2017; Sterckx et al., 2016; Powles and Hodson, 2017; Smits, 2013; Garrety et al., 2014). This calls for more knowledge to better understand why the promise of data can be elusive in practice, but also efforts that propose answers to steer clear of this fate. This thesis responds to this call form a Danish perspective with an interrogation of two carefully selected case studies through the following research question:

> *How is data-intensification sought and brought about by infrastructural initiatives in the Danish ‘health data ecosystem’, and what tensions arise in the pursuit of valuations assigned to personal health data?*

In this thesis, I have presented empirical research on the rise and fall of infrastructures in their attempts to navigate a promissory space of uncertainty to unleash the promise of data. I have also shown how, in the work of infrastructuring an ecosystem for health data, specific values become contested and resolved. The first to venture into this space was DAMD that was at one point a beacon for chronic care in Danish healthcare and another a national health data controversy. The DAMD was however also a proposed model for data-intensification that from a technical perspective was exemplary until it proved to be unsustainable from a social and legal perspective. Through a series of three articles, I tell a story of the rise of the DAMD (Article II), the beginning of its end (Article III) and the negotiation leading up to the final judgement (Article IV). The second to make an advance were data intermediaries that proposed an alternative model for data-intensification based on ideas of individual control. While other data intermediaries in the Nordic region have struggled to find a sustainable model (Article I and V), Data for Good Foundation and Digi.Me look to succeed where others failed by targeting a market for *ethical* data sharing (Article V).

The examinations of infrastructures in the two core case studies have not been the end-point of the thesis. This interrogative work has also served as a probe to identify relevant movements and questions relating to the health data ecosystem in which the case studies are situated. The articles provide a collective analysis of the *life cycle of data within the emerging ecosystem*, from its collection to its (re)use and deletion. They contribute empirical, conceptual and methodological tools for understanding efforts to make use of personal health data in Denmark. The articles also demonstrate how emerging regimes of data accumulation are being challenged by questions of responsible conduct; How health and its datafication is expanding, along with logics of prevention and self-care and how the traditional roles and responsibilities of the Danish welfare state are changing. The changes for the Danish welfare state include emerging agendas of data entrepreneurialism that give shape to other changes in the form of collaborations, governance models and gatekeeper authority. Seen together, the articles contribute to the overarching argument that I put forward with this thesis: As the health data ecosystem evolves, so does the relation between citizen, state and market, which in turn engenders new questions of societal and political importance.

Future research

It is my hope that the work conducted in this thesis might inspire future research on data for scholars working in STS, CDS and adjacent fields. Theoretically, I suggest that the ecosystem frame can help scholars see the various phases in the life cycle of data as artificial distinctions that in practice are reciprocal in nature. The theme of life and death is one that scholars may consider that can concern data, databases, infrastructures, platforms, actors, metaphors, and ideologies, with an interest in how the past shapes the practices that follow. I propose that scholars consider the legacies of ecosystems with a view to highlight critical insights that may also be generative (Latour, 2004). More empirical research is needed to understand the models and values put forward by data intermediaries specifically. In which ways might they be considered to advance responsible data conduct and what new questions do they engender? How do they relate to existing and emerging movements, for instance in terms of the roles and responsibilities between state, market and citizen? In terms of method, I suggest participatory observations at conferences as a way to explore the fast-moving field of tech at sites that are both rich and dense.

Cultivating care for socio-environmental sustainability

As a contribution, one thing I pass on from ecological ecosystems to health data ecosystems regards the need to cultivate care. Today, it is widely accepted that the extraction of natural resources has an impact on the ecosystem on multiple scales. This has led to a recognition that extraction practices must respect certain limits, even though the costs involved can be difficult to predict due to the complexity involved. I propose that this lesson should be passed on to the health data ecosystem where the need for careful extraction practices is less accepted, let alone recognised. This is, as mentioned in the introduction, partly reflected in (ab)use of certain metaphors, such as the ‘golden egg’ articulated in relation to the DAMD. Contrary to what the system designers aim for, the DAMD turned out to be anything other than ‘future proof’ as its accumulation of data and purposes exceeded the social and legal limits of acceptability to the point of collapse (Article III). Part of cultivating care entails being careful about what devices (e.g. metaphors and narratives) we choose to make sense of data while fostering relevant critical sensibilities to for instance recognise what linear models of value creation fail to capture. As such, the present thesis contributes to scholarly conversations on the social sustainability of contemporary data practices, which is explicitly articulated in Article III in relation to Bennett's (2012) work on hoarding.

Future research could however also focus on the *environmental* sustainability of data practices. As noted Article III, Bennett's engagement with hoarding is also an eco-critique, which is viewed to convey 'the madness appropriate to a political economy devoted to over-consumption, planned obsolescence, relentless extraction of natural resources ("Drill Baby Drill"), and vast mountains of disavowed waste' (Bennett, 2012: 248). For Bennett, hoarding is then part of a posthuman project to make consumption practices more sustainable by foregrounding aesthetic and material aspects of the call of *things*. Whereas the analytic I propose with the 'call of *data'* extends this project to *social* sustainability of *data*, an environmental consideration is complimentary.

The *environmental* sustainability of data is in passing already registered in Schüll's (2018) work on digital hoarding. The point she makes is that even though the storage space occupied by digital data may seem to be 'virtually nil', retention can still be problematic due to its consumption of 'energy and other worldly resources' (Schüll, 2018: 44). During fieldwork at the *Data for Good Summit 2020* in Copenhagen, the director general of the trade association Digital Europe stated that 80 percent of the registered health data remains unstructured or untouched after its acquisition. At the event, the estimate was used to underscore the importance of *enhanced data availability* in Europe given the global competition in the political economy of health data. In the current data climate, data retention is more often than not the default mode of valuation that requires a significant burden of proof to make it otherwise (Article IV). Conversations on the *environmental* sustainability of data practices can in this regard add weight to the case for deletion, which is already used as an argument in the world of data management consultancy (Veritas, 2020, see also Veritas, 2015).

Obviously, there are many legitimate reasons (and obligations) *for* data retention in healthcare where timely access to relevant data can potentially be a matter of life and death - which can also be found in the case *against* retention (Article IV). The environmental sustainability of data can however still be relevant to potential cases of obsolete data as something to consider alongside the concerns for social sustainability. Consideration of the socio-environmental sustainability of data are particularly relevant in a time of ever-growing amounts of health data gaining increased value for non-health purposes. As actors look to repurpose the 'Scientific goldmine' for Denmark's new oil adventure, critical sensibilities to social sustainability will be important, as will an environmental perspective that introduces a more critical spin to the popularised characterisation of casting data "the new oil". If the objective is to create sustainable health data ecosystems then this must also be reflected in the infrastructural attempts to marshal and manage the life cycles of health data that they inhabit.

7. References

‘About’ (n.d.) Data for Good Foundation. Website. https://dataforgoodfoundation.com/en/about (accessed23 Sep, 2020)

Albris K, Otto EI, Astrupgaard SL et al. (2021) A view from anthropology: Should anthropologists fear the data machines?. *Big Data & Society,* 8(2): 1-7.

Anand N, Gupta A and Appel H (Eds) (2018) The Promise of Infrastructure. Duke University Press.

Anker P (2001) *Imperial Ecology: Environmental Order in the British Empire*. Harvard University Press.

Appel H, Anand N and Gupta A (2018) Introduction: Temporality, Politics, and the Promise of Infrastructure.

Aula V (2019) Institutions, infrastructures, and data friction–Reforming secondary use of health data in Finland. *Big Data & Society* 6(2): 1-13.

Ayres PG (2012) *Shaping Ecology: The Life of Arthur Tansley*. John Wiley & Sons.

Barry A (2001) *Political machines: Governing a technological society.* A&C Black.

Bart J and Popma J (2019) Medical research, big data and the need for privacy by design. *Big Data & Society* (6)1: 1-5.

Bauer S (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.

Beer D (2016) How should we do the history of Big Data? *Big Data & Society* 3(1): 1-10.

Bennett J (2012) Powers of the hoard: Further notes on material agency. *Animal, vegetable, mineral: Ethics and objects*, 237-269.

Bigger P and Rossi J (2012) *Review of all watched over by machines of loving grace*. Dir adam curtis. British Broadcasting Corporation 2011. Television. *Human Geogr* 5(1): 109–112.

Blasimme, A., Vayena, E., & Van Hoyweghen, I. (2019). Big data, precision medicine and private insurance: a delicate balancing act. *Big Data & Society* 6(1): 1-6.

Bonde M, Bossen C and Danholt P (2019) Data-work and friction: Investigating the practices of repurposing healthcare data. *Health informatics journal* 25(3): 558-566.

Borgman CL (2015) *Big Data, Little Data, No Data: Scholarship in the Networked World.* Cambridge/London: MIT Press.

Borra E and Bernhard R (2014) Programmed Method: Developing a Toolset for Capturing and Analyzing Tweets. In: Aslib *Journal of Information Management* 66/3, 262–78.

Bot BM, Wilbanks JT and Mangravite LM (2019) Assessing the consequences of decentralizing biomedical research. *Big Data & Society 6*(1): 1-6.

Bowker GC (1995) Second Nature Once Removed: Time, Space and Representations. *Time and Society,* 4(1): 47–66.

Bowker G (2005) *Memory Practices in the Sciences*. Cambridge: MIT Press.

Bowker GC & Star SL (1999) *Sorting things out: Classification and its consequences.* MIT press.

Bruns A and Burgess J (2011) The Use of Twitter Hashtags in the Formation of Ad Hoc Publics. In: Proceedings of the 6th European Consortium for Political Research (ECPR) General Conference 2011, University of Iceland, Reykjavik, pp. 1–9.

Bruun MH and Wahlberg A (2022) The Anthropology of Technology: The Formation of a Field. In: Brun MH, Wahlberg A, Douglas-Jones R et al. (Eds) *The Palgrave Handbook of the Anthropology of Technology.* Palgrave Macmillan, Singapore

Burnett (2017) The DAMD controversy: A digitally informed mapping of issues and their politics. Symposium: *Digital Healthcare: social logics, ethics and politics of data and technology*. University of Nottingham, UK.

Burnett (2018) Digitale Dilemmaer: Hvordan Forandrer Data Den Offentlige Sektor?. Public Dissemination. Available at: https://www.itu.dk/om-itu/events/events/2018/digitale-dilemmaer.

Law J and Callon M (1992) The Life and Death of an Aircraft: A Network Analysis of Technical Change In: Bijker W and Law J (eds) *Shaping Technology/Building Society — Studies in Sociotechnical Change.* Cambridge, Ma: MIT Press, pp. 21- 52.

Clarke AE, Mamo L, Fosket JR, et al. (2010) *Biomedicalization*. Duke University Press.

Cameron L and Earley S(2015) The ecosystem—movements, connections, tensions and translations. *Geoforum* (65): 473-481.

Chafkin M (2010) The Oracle of Silicon Valley. Inc. Available at: https://www.inc.com/magazine/20100501/the-oracle-of-silicon-valley.html (accessed 31 Mar 2022).

Cima Technologies (2017) Cash in on your data gold mine. Available at: http://www.cima.dk/2017/06/07/cash-in-on-your-data-gold-mine-in-danish/ (accessed 23 Aug, 2017).

Coleman DC, Swift DM, and Mitchell JE (2004) From the Frontier to the Biosphere. *Rangelands* 26(4): 8-15.

Copenhagen Institute for Future Studies (2019) Nordic Health 2030: Towards Preventive Health. Available at: https://www.norden.org/en/publication/nordic-health-2030.

Curtis A (2011) The Use and Abuse of Vegetational Concepts, Part II of *All Watched Over By Machines of Loving Grace*, Television, BBC documentary series. 30 May 2011.

Custers B and Ursic H (2016) Big data and data reuse. *International Data Privacy Law Advance*: 1–12.

de Humboldt A and Bonpland AIME (1805). *Essai sur la géographie des plantes: Accompagne D'Un Tableau Physique Des Regions Equinoxiales*. Paris, Schoell.

Delgado, N. A., & Cruz, L. B. (2014). Multi-event ethnography: Doing research in pluralistic settings. *Journal of Organizational Ethnography*.

Douglas-Jones R, Walford A, and Seaver N (2021) Introduction: Towards an anthropology of data. *Journal of the Royal Anthropological Institute* 27(S1): 9–25.

DR Medieforskning (2015) Medieudviklingen 2015. *Danish Broadcasting Corporation.* Available at: https://www.dr.dk/static/documents/2._medieudviklingen_2015_3_5a311a2a.pdf/ (accessed Aug 21, 2018).

Edwards PN (2003) Infrastructure and modernity: Force, time, and social organization in the history of sociotechnical systems. Modernity and technology, 1: 185-226.

European Commission (2020) A European strategy for data.

European Commission (n.d.) What personal data is considered sensitive? Available at: https://ec.europa.eu/info/law/law-topic/data-protection/reform/rules-business-and-organisations/legal-grounds-processing-data/sensitive-data/what-personal-data-considered-sensitive_en

Evans-Pritchard, E. E. (1928). *The social organization of the Azande of the Bahr-el-Ghazal province of the Anglo-Egyptian Sudan*. Doctoral dissertation, The London School of Economics and Political Science (LSE)).

Faulkner-Gurstein R and Wyatt D (2021) Platform NHS: Reconfiguring a Public Service in the Age of Digital Capitalism. *Science, Technology, & Human Values:* 1-21.

Fischer A and Tynell J (2014) OVERBLIK Det handler data-sagen om. *Danish Broadcasting Corporation*, 5 December.

Fox RC (2004). Observations and reflections of a perpetual fieldworker. *The ANNALS of the American Academy of Political and Social Science* 595(1): 309-326.

Frank L (2000) When an entire country is a cohort. *Science* 287(5462): 2398-2399.

Garrety K, McLoughlin I, Wilson R, et al. (2014) National electronic health records and the digital disruption of moral orders. Social Science & Medicine 101: 70–77.

Geertz C (1973) Thick description: Toward an interpretive theory of culture. In: Basic Books, pp. 310-323.

Geertz C. (2000) Deep play: Notes on the Balinese cockfight. In *Culture and politics* Palgrave Macmillan: New York, pp. 175-201.

Gillespie T (2018) *Custodians of the Internet.* Yale University Press.

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

Golley FB (1991) The ecosystem concept: A search for order. *Ecological Research* 6(2): 129-138.

Golley FB (1993) *A History of the Ecosystem Concept in Ecology.* Yale University Press.

Gordon C (ed) (1980). Power/knowledge. Pantheon: New York.

Graham S and Thrift M (2007) Out of Order: Understanding Repair and Maintenance. *Theory, Culture & Society*, 24(3): 1–25.

Guyer JI (2016) *Legacies, Logics, Logistics.* University of Chicago Press.

Hacking I (1982) Biopower and the avalanche of printed numbers. *Humanities in Society* 5(3–4): 279–295.

Hacking I (1991) How should we do the history of statistics? In: Burchell G, Gordon C and Miller P (eds) *The Foucault Effect: Studies in Governmentality*. Chicago: The University of Chicago Press, 181-195.

Hacking I (1996) The Looping Effects of Human Kinds. In Causal Cognition*:* Sperber et al. (eds) *A Multidisciplinary Approach*. Oxford: Oxford University Press: 351-83.

HackingI (2007) Kinds of People: Moving Targets. *Proceedings-British Academy* 151: 285-31.

Hagen JB (1992) *An entangled bank: the origins of ecosystem ecology*. Rutgers University Press. New Runswick: New Jersey.

Hafen E, Kossmann D and Brand A (2014) Health data cooperatives–citizen empowerment. *Methods of information in medicine* 53(02), 82-86.

Haraway D (1988) Situated Knowledges: The Science Question in Feminism and the Privilege of Partial Perspective. Feminist studies 14(3): 575.

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

Hastrup K (1995) *A Passage to Anthropology*. Routledge: London.

Heaphy (2020) Social Studies of Scientific Knowledge. International Encyclopedia of Human Geography 4: 263-270.

HedaX (n.d.) Gevinsten ved projektet. Available at: https://www.hedax.dk/gevinsten-ved-projektet (accessed Feb 02, 2022).

Heijlen R and Crompvoets J (2021) Open health data: Mapping the ecosystem. *Digital Healt*h (7): 1-16.

Hockenhull M and Cohn ML (2021) Hot air and corporate sociotechnical imaginaries: Performing and translating digital futures in the Danish tech scene. *New Media & Society* 23(2): 302-321.

Hoeyer K (2016) Denmark at a crossroad? Intensified data sourcing in a research radical country. In: Mittelstadt BD and Floridi L (eds) *The Ethics of Biomedical Big Data.* Cham: Springer, 73-93.

Hoeyer K (2018) Lost and found: Relocating the individual in the age of intensified data sourcing in European healthcare. *Personalised medicine, individual choice and the common good,* 133-154.

Hoeyer K (2019) Data as promise: Reconfiguring Danish public health through personalized medicine. *Social Studies of Science* 49(4): 531-555.

Hoeyer K and Winthereik BR (2019) Smukke, rå og beskidte: Data som antropologisk anliggende. Tidsskriftet Antropologi, (79): 135-141.

Hogan M and Vonderau A (eds) (2019) *The Nature of Data Centres.* Culture Machine 18.

Hogle L (2016) The ethics and politics of infrastructures: Creating the conditions of possibility for big data in medicine. In: *The Ethics of Biomedical Big Data,* Springer: Cham, pp. 397-427.

Holm-Larsen (2017) Danmarks sundhedsdata. Den nye Nordsø-olie? Børrsen. Available at: https://borsen.dk/opinion/blogs/view/17/4551/danmarks_sundhedsdata_den_nye_nordsa-olie.html.

Karasti and Blomberg, 2017 (2017) Knox H and Nafus D (2018) *Ethnography for a data-saturated world.* Manchester University Press.

Krogness KJ (2011) Numbered individuals, digital traditions, and individual rights: Civil status registration in Denmark 1645-2010. *Ritsumeikan Law Review* 28: 87-126.

Kvale S (1999) Constructing Knowledge Through Interviews. In Maiers WB et al. *Challenges to Theoretical Psychology*, Captus Press, York.

Langhoff TO, Amstrup MH, Morck P, and Bjorn P (2016) Infrastructures for Healthcare: From Synergy to Reverse Synergy. *Health Informatics Journal*, (24)1: 1-11.

Langstrup H (2019) Patient-reported data and the politics of meaningful data work. *Health informatics journal* 25(3): 567-576.

Larkin B (2013) The politics and poetics of infrastructure. *Annual review of anthropology,* 42: 327-343.

Lathrop D and Ruma L (2010) *Open government: Collaboration, transparency, and participation in practice.* O'Reilly Media, Inc.

Latour B (1987) *Science in action: How to follow scientists and engineers through society*. Harvard university press.

Latour B (1996) *Aramis, or, The love of technology.* Harvard University Press.

Latour B (2004) Why has critique run out of steam? From matters of fact to matters of concern. *Critical inquiry* 30(2): 225-248.

Latour B (2005) *Reassembling the Social: An Introduction to Actor-Network-Theory.* Oxford Press.

Latour B and Venturini T (2010) The Social Fabric: Digital Traces and Quali-quantitative Methods. In Proceedings of Future en Seine 2009. Cap Digital. http://www.medialab.sciences-po.fr/publications/the-social-fabric-digital-traces-and-quali-quantitative-methods/

Law J (1987) On the social explanation of technical change: The case of the Portuguese maritime expansion. *Technology and Culture* 28(2): 227-252

Law, John and Michel Callon (1992) The Life and Death of an Aircraft: A Network Analysis of Technical Change, In W. Bijker and J. Law (eds. (1992). *Shaping Technology/Building Society — Studies in Sociotechnical Change*. Cambridge, Ma: MIT Press, pp. 21- 52.

Lehtiniemi, T. (2017). Personal Data Spaces. *Surveillance and Society* 15(5), 626–639.

Lehtiniemi T and Kortesniemi Y (2017) Can the obstacles to privacy self-management be overcome? Exploring the consent intermediary approach. *Big Data & Society* 4(2): 1-11.

Lehtiniemi T and Ruckenstein M (2018). The social imaginaries of data activism. *Big Data & Society*, 6(1): 1-12.

Leonelli S, Rappert B and Davies G (2016) Data shadows: Knowledge, openness and absence. *Science, Technology, & Human Values* 42(3): 191–202.

Leonelli, S., Rappert, B., & Davies, G. (2017). Data shadows: Knowledge, openness, and absence. *Science, Technology, & Human Values,* 42(2): 191-202.

Leonelli, S., & Tempini, N. (2020). *Data journeys in the sciences*. Springer Nature.

Lindeman LR (1942) The trophic-dynamic aspect of ecology. *Ecology* 23(4): 399-417.

Lippert I and Douglas-Jones R (2019) Doing Data: Methodography in and of STS. *EASST Review,* 38(1): 35-39.

Lippert ML, Kousgaard MB and Bjerrum L (2014) General practitioners uses and perceptions of voluntary electronic feedback on treatment outcomes – a qualitative study. 5 ed. BMC Family Practice, *BioMed Central* (15)1: 85–9.

Lippert ML, Reventlow S, and Kousgaard MB (2017) The Uses and Implications of Standards in General Practice Consultations. Health. *An Interdisciplinary Journal for the Social Study of Health, Illness and Medicine* 21(1): 3–20.

Lucivero F and Prainsack B (2015) The lifestylisation of healthcare? ‘Consumer genomics’ and mobile health as technologies for healthy lifestyle. *ATG, Elsevier B.V.* 4(C): 44–49.

Lupton D (2017) Lively Data, Social Fitness and Biovalue: the Intersections of Health and Fitness Self-Tracking and Social Media. In: Burgess J, Markick A and Poell (Eds) *The SAGE Handbook of Social Media.* SAGE Publications Ltd, pp. 562-578.

Maguire J, Langstrup H, Danholt P et al. (2020) Engaging the data moment: an introduction. STS Encounters. (11)1: 7-26.

Maguire J and Winthereik BR (2019) Digitalizing the state: Data centres and the power of exchange. *Ethnos* 86(3): 530-551.

Malinowski, B. (2001). Coral Gardens and Their Magic: A Study of the Methods of Tilling the Soil and of Agricultural Rites in the Trobriand Islands. The language of magic and gardening. II (Vol. 8). Psychology Press.

Marelli L, Lievevrouw E and Van Hoyweghen I (2020). Fit for purpose? The GDPR and the governance of European digital health. *Policy studies* 41(5): 447-467.

McFall L, Meyers G and Van Hoyweghen I (2020) Editorial: The personalisation of insurance: Data, behaviour and innovation. *Big Data & Society:* 1-11.

Mackenzie A, Waterton C, Ellis R et al (2013 Classifying, constructing, and identifying life: standards as transformations of “The Biological”. Science, technology, & human values 38(5): 701-722.

MyData (n.d.) Declaration. Official website. Available at: https://mydata.org/declaration/ (Accessed 25 Nov, 2019).

Neyland D (2018) Something and nothing: On algorithmic deletion, accountability and value. *Science and Technology Studies* 31(4): 13–29.

Nordfalk F and Hoeyer K (2020) The rise and fall of an opt-out system. *Scandinavian Journal of Public Health* 48(4): 400–404.

Ny Teknik. (2018, October 2). GDPR spiken i kistan för statens digitala vårdtjänst | Ny Teknik. Available at: https://www.nyteknik.se/digitalisering/gdpr-spiken-i-kistan-for-statens-digitala-vardtjanst-6933413 (accessed Feb 15, 2019).

OECD. (2013). OECD Reviews of Health Care Quality: Denmark Executive Summary, Assessment and Recommendations. Available at: https://www.oecd.org/publications/oecd-reviews-of-health-care-quality-denmark-2013-9789264191136-en.htm. (accessed Mar 30, 2022)

OECD (2015) Data-Driven Innovation Big Data for Growth and Well-BeIng. OECD Publishing, Paris. http://dx.doi.org/10.1787/9789264229358-en

O'Rand, A. M., & Krecker, M. L. (1990). Concepts of the life cycle: Their history, meanings, and uses in the social sciences. *Annual review of sociology* 16(1): 241-262.

Oliveira S, Lima GFB and Lóscio BF (2019) Investigations into Data Ecosystems: a systematic mapping study. *Knowl Inf Syst* (61): 589–630.

Pasquale F (2015) *The black box society.* Harvard University Press.

Pickering, A. (2010). The cybernetic brain. University of Chicago Press.

Poikola A, Kuikkaniemi K and Honko H (2015) MyData – A Nordic Model for human-centered personal data management and processing. Available at: https://julkaisut.valtioneuvosto.fi/bitstream/handle/10024/78439/MyData-nordic-model.pdf.

Purtova N (2015) The illusion of personal data as no one's property. *Law, Innovation and Technology,* Taylor & Francis 0(0): 1–29.

Powles J and Hodson H (2017) Google DeepMind and healthcare in an age of algorithms. *Health and technology* 7(4): 351-367.

Prainsack B (2019a) Data donation: How to resist the iLeviathan. In: Jenny Krutzinna and Luciano Floridi (eds). *The Ethics of Medical Data Donation*. Dordrecht: Springer. pp. 9–22.

Prainsack B (2019b) Logged out: Ownership, exclusion and public value in the digital data and information commons. *Big Data & Society*: 1-15.

Prainsack B (2020) The value of healthcare data: to nudge, or not?. *Policy studies* 41(5), 547-562.

Puschmann, C., & Burgess, J. (2014): The Politics of Twitter Data. In Weller K, Bruns A, Burgess J, Mahrt M, and Puschmann C (Eds) *Twitter and Society*, Peter Lang, pp. 43–54.

Rapport DJ (2002) The health of ecology and the ecology of health. *Human and Ecological Risk Assessment (8)*1: 205-213.

Ratner H and Gad H (2019) Data warehousing organization: Infrastructural experimentation with educational governance. *Organization* (26)4: 537-552.

Reinsel, Gantz and Rydning (2018) Data Age 2025: The Digitization of the World From Edge to Core. Available at: https://www.seagate.com/files/www-content/our-story/trends/files/idc-seagate-dataage-whitepaper.pdf (accessed 31 Mar, 2022).

Riso B, Tupasela A, Vears DF et al. (2017) Ethical sharing of health data in online platforms – which values should be considered? *Life Sciences, Society and Policy* 13(12): 1-27.

Roberts C and Mackenzie A (2019) *Living Data: Making sense of health biosensing.* Policy Press.

Rogers R. (2009) *End of the Virtual.* Inaugural Lecture Series.

Rogers R (2013) Situating Digital Methods. In *Digital Methods*. MIT press: 1-18.

Rogers R (2014) Debanalising Twitter. In Weller K, Bruns A, Burgess J, Mahrt M, and Puschmann C (Eds) Twitter and Society. Peter Lang: ix–xxvi.

Ruppert E, Law J, and Savage M (2013): “Reassembling Social Science Methods: The Challenge of Digital Devices.” Theory, Culture & Society (30)4: 22–46.

Rose N (2009) The Politics of life itself: Biomedicine, power, and subjectivity in the twenty-first century. The Politics of Life Itself: Biomedicine, Power, and Subjectivity in the Twenty-First Century, Biomedicine, Politics, and Subjectivity in the 21st Century, Princeton University Press.

Reeves, M. (2016). The black list: on infrastructural indeterminacy and its reverberations. In Harvey P, Jensen CB and Morita A (Eds) *Infrastructures and Social Complexity.* Routledge: NY pp. 314-326.

Riso B, Tupasela A, Vears DF et al. (2017) Ethical sharing of health data in online platforms – which values should be considered? *Life Sciences, Society and Policy*, 13(12): 1-27.

Ruckenstein M (2014) Visualized and Interacted Life: Personal Analytics and Engagements with Data Doubles. *Societies* 4(1): 68–84.

Ruckenstein M and Schüll ND (2017) The Datafication of Health. *Annual Review of Anthropology* 46(1): 261–278.

Schüll ND (2018) Digital containment and its discontents. History and Anthropology 29(1): 42-48.

Scott, J. C. (1998). *Seeing like a state: How certain schemes to improve the human condition have failed*. Yale University Press.

'Search Tweets' (2018). Developer.Twitter.com. Available at: https://developer.twitter.com/en/docs/tweets/search/overview/standard.html (Accessed Aug 22, 2018).

Sharon T (2016) The Googlization of health research: from disruptive innovation to disruptive ethics. *Personalized Medicine*, 13(6): 563-574.

Sharon T (2018) When digital health meets digital capitalism, how many common goods are at stake? *Big Data & Society* 5(2): 1-11.

Sharon T and Zandbergen D (2017) From data fetishism to quantifying selves: Self-tracking practices and the other values of data. *New Media & Society* 19(11): 1695-1709.

Sharon, T., & Lucivero, F. (2019). Introduction to the special theme: The expansion of the health data ecosystem–Rethinking data ethics and governance. Big Data & Society, 6(2): 1-5.

SitraFund (2010) Taltioni - Finnish PHR platform. Youtube. Available at: https://www.youtube.com/watch?v=5h-W7PP9K1o (accessed 30 November, 2017).

Skinner D and Wienroth M (2019) Was this an ending? The destruction of samples and deletion of records from the UK Police National DNA Database. *British Society for the History of Science* 4: 99-121.

Slota SC, Hoffman AS, Ribes D, et al. (2020) Prospecting (in) the data sciences. *Big Data & Society* 7(1): 1-12.

Solove, D. J. (2012). Introduction: Privacy self-management and the consent dilemma. *Harv. L. Rev.* 126: 1880-1903.

Srnicek N (2016) *Platform Capitalism*. John Wiley & Sons.

Srnicek N (2017) The challenges of platform capitalism. *Juncture* 23(4): 254–257.

Star SL (1999) The ethnography of infrastructure. *American Behavioral Scientist*, 43(3): 377–91.

Star SL (2002) Infrastructure and ethnographic practice: Working on the fringes. *Scandinavian Journal of Information System* (14)2: 107-122.

Starkbaum J and Felt U (2019) Negotiating the reuse of health-data: Research, big data, and the European General Data Protection Regulation. *Big Data & Society* 6(2): 1-12.

Starosielski N (2015) *The Undersea Network: Sign, Storage, Transmission.* Durham: Duke University Press.

StatCounter Global Stats (2018) Social Media Stats Denmark. Available at: http://gs.statcounter.com/social-media-stats/all/denmark. (accessed Aug 21, 2018).

Sterckx S, Rakic V, Cockbain J, et al. (2016) 'You hoped we would sleep walk into accepting the collection of our data': Controversies surrounding the UK care.data scheme and their wider relevance for biomedical research. Medicine, Health Care and Philosophy 19: 177–190.

Stolton S (2019) Vestager's grand plan for regulating the digital economy. EURACTIV. Available at: https://www.euractiv.com/section/digital/news/vestagers-grand-plan-for-regulating-the-digital-economy/?fbclid=IwAR0al09Jg4DyM0OP0NlCYdtRyM_Yr--5FOJKMivYbhaL-fr4RQdrVgfCp68 (accessed Jun 19, 2022).

Strathern M (1996) Cutting the network. *Journal of the Royal Anthropological Institute.* 517-535.

Strebel I, Bovet A and Sormani P (Eds) (2019) *Repair work ethnographies: Revisiting breakdown, relocating materiality*. Palgrave macmillan.

Taylor L, Floridi L and Van der Sloot B (Eds.). (2016). *Group privacy: New challenges of data technologies* (Vol. 126). Springer.

Taylor L (2016) The ethics of big data as a public good: Which public? Whose good? *Philosophical Transactions of the Royal Society A: Mathematical Physics and Engineering Sciences* 374(2083): 1-13.

Taylor L and Purtova N (2019) What is responsible and sustainable data science?. *Big Data & Society*: 6(2): 1-6.

Technology, Methods and Infrastructure for Networked Medical Research (2016) „Lebenszeit-Epidemiologie“ in Dänemark möglich. Available at: https://www.tmf-ev.de/News/articleType/ArticleView/articleId/3037.aspx.

The Danish Government (2018) *A Coherent and Trustworthy Health Network for All: Digital Health Strategy 2018–2022.*

Torenholt, R., & Langstrup, H. (2021). Between a logic of disruption and a logic of continuation: Negotiating the legitimacy of algorithms used in automated clinical decision-making. *Health*: 1-19.

Tupasela A (2021) Populations as Brands: Marketing National Resources for Global Data Markets. Springer International Publishing.

Tupasela A, Snell K and Tarkkala H (2020) *The Nordic data imaginary.* Big Data & Society 7(1): 1–13.

Vallgarda S (2007) Public health policies: A Scandinavian model? *Scandinavian Journal of Public Health*, 35: 205–211.

Van Dijck J, Poell T, and De Waal M (2018). *The platform society: Public values in a connective world.* Oxford University Press.

Van Dyne GM (1969) *The ecosystem concept in natural resource management.* Academic Press: New York/London.

Vayena E and Blasimme A (2017) Biomedical big data: new models of control over access, use and governance. Journal of bioethical inquiry, 14(4): 501-513.

Vayena E and Blasimme A (2018) Health research with big data: time for systemic oversight. *Journal of Law, Medicine & Ethics*. 46(1): 119-129.

Vayena E, Dzenowagis J, Brownstein JS et al. (2018) Policy implications of big data in the health sector. *Bulletin of the World Health Organization*, 96(1): 66.

Vayena E and Gasser U (2016) Strictly biomedical? Sketching the ethics of the big data ecosystem in biomedicine. In *The ethics of biomedical big data.* Springer: Cham, pp. 17-39.

Venturini, T. (2010). Building on faults: How to represent controversies with digital methods. *Public Understanding of Science,* 21(7), 796–812.

Veritas (2015) The data report. See what others don’t - identify the value, risk and cost of your data. Available at: http://images.info.veritas.com/Web/Veritas/%7B364a7ca5-e05c-4fce-971b-88e18c62eafb%7D_45145_EMEA_Veritas_Strike_Report_Gulf.pdf.

Veritas (2020) Media Alert: Veritas Technologies Projects Dark Data to Waste Up to 6.4m Tons of Carbon Dioxide this Year. Available at:

https://www.veritas.com/news-releases/2020-04-21-veritas-technologies-projects-dark-data-to-waste-up-to-6-4-m-tons-of-carbon-dioxide-this-year.

Vogt H, Hofmann B and Getz L (2016) The new holism: P4 systems medicine and the medicalization of health and life itself. *Medicine, Health Care and Philosophy, Springer Netherlands* 19(2): 307–323.

Walford A (2021) Data Aesthetics. In: Carroll T, Walford A and Walton (Eds) *Lineages and Advancements in Material Culture Studies*. London/New York: Routlegde, pp. 205-217.

Wang, T. (2013, May 13). Big Data Needs Thick Data. Ethnography Matters. http://ethnographymatters.net/blog/2013/05/13/big-data-needs-thick-data/

Wadmann S and Hoeyer K (2018) Dangers of the digital fit: Rethinking seamlessness and social sustainability in data-intensive healthcare. *Big Data & Society* 5(1): 1–13.

Wahlberg, A. (2022). Assemblage Ethnography: Configurations Across Scales, Sites, and Practices. In: Bruun, Wahlberg, Douglas-Jones et al. *The Palgrave Handbook of the Anthropology of Technology*. Palgrave Macmillan, Singapore. https://doi.org/10.1007/978-981-16-7084-8_6

Weltevrede EJT (2016) *Repurposing Digital Methods: The Research Affordances of Platforms and Engines*. PhD Thesis, Universiteit van Amsterdam, NL.

Wessels B, Wadhwa FR and Sveinsdottir T (2017). *Open data and the knowledge society.* Amsterdam University Press.

Willis AJ (1997) The ecosystem: an evolving concept viewed historically. *Functional ecology* (11)2: 268-271.

Winthereik BR and Gad C (2015) *Data as Relation: Governance in the Age of Big Data.* Research Proposal. The Velux Foundations.

Winthereik BR, van der Ploeg I and Berg M (2007) The Electronic Patient Record as a Meaningful Audit Tool: Accountability and Autonomy in General Practitioner Work. Science, Technology, & Human Values. 2007;32(1):6-25. doi:10.1177/0162243906293884

Yarrow T (2017) Remains of the future: rethinking the space and time of ruination through the Volta resettlement project, Ghana. *Cultural Anthropology* 32(4): 566-591.

Zuboff S (2019) *The age of surveillance capitalism: The fight for a human future at the new frontier of power*. New York: PublicAffairs.

@atornes and @leanne_tru (2021) Enabling the future of academic research with the Twitter API. Twitter Developer Platform Blog. Available at: https://blog.twitter.com/developer/en_us/topics/tools/2021/enabling-the-future-of-academic-research-with-the-twitter-api (accessed 29 Mar, 2022).

PART 2

“Let’s make it happen!”: An STS Ethnographer’s Guide to Tech Conferences

Rachel Douglas-Jones, John Mark Burnett, Ester Fritsch and Michael Hockenhull

Abstract

Tech conferences are openings to the field for ethnographers of technological development. Analysing accounts from nine different conferences in the fields of IoT, biomedical big data and business innovation in big data, in this paper we develop a guide to the modes of critical and ethnographic engagement demanded by tech conferences. As events, they prompt us to become conversant in the languages of our area of study and offer unique opportunities to engage with stories that circulate in and construct our fields. As fieldsites, they challenge norms of presence, relation and observation. By connecting histories of conferencing as practice with the our observations about specific characteristics of tech conferences, we provide reflexive, ethnographically supported guidance for ethnographers newly entering their own fields of technology research.

Keywords: tech, conference, ethnography

How does one study the world of technology conferences? As locations where STS ethnographers open up their field-sites, tech conferences gather diverse arrays of people and things together. Some conferences are annual events, attended over the years of a study. Initial encounters in these fields might be characterized by the ethnographer’s nerves, that moment of entering a bustling room knowing nobody. Some conferences form part of a field’s conference ‘circuit’, where over the years, longer-term immersion brings familiarity with particular advocates, positions, ways of talking, languages and tacit norms. This article results from the relative absence of literature addressing the practicalities and implications of conference ethnography for STS scholars. Motivated by a fascination with the form of tech conference sociality that results at these events, and, in particular, what attending them can do for research projects within STS, our paper details both methodological considerations for fieldwork and reflections on where we might generatively turn our attention as conference ethnographers. Based on several years of ethnographic work and discussions about the role of conferences in STS ethnographies, we explore the techniques, affects and politics that occur within the boundedness of conference time.

Over the course of three years, the authors ethnographically attended more than 90 tech events – such as innovation meetings, hackathons, data sprints, meetups, workshops, seminars and conferences – in the context of research projects on emerging technologies. As such, the specific technology in focus varied across our respective projects and conferences. The conference vignettes are drawn from three different focus areas. During a project on the Internet of Things (IoT), the 2017 *MobileWorldCongress* in Barcelona was central. Inaugurated in 1987, it claims the status of the largest mobile event in the world. Today referred to as MWC – reducing the importance of the word "mobile" – it brings together global manufacturers, sales teams, innovators and 'visionaries' to exhibit, discuss and debate the future of connective technology. IoT technologies are present at the conference in advanced and prototype form, varying from industrial applications to wearables and home devices, collecting ever increasing amounts of data, and coming online in vast numbers. As part of a project on the emerging ecology of health data, the *MyData* conference became empirically central, with its "Nordic Model for human-centered personal data management and processing" (Kuikkaniemi, Poikola and Honko 2015). From its roots in the Finnish (open) data activism, the annual MyData conference has, since its first outing in 2016, returned to Finland's capital Helsinki where entrepreneurs, private individuals, activists, academics, listed corporations, public agencies and developers from various nations (re)convene.[i] *MyData* sees itself as a social movement with a common objective to strengthen individual control over personal data and thereby restore balance to what it views as the asymmetries of the contemporary personal data economy/ecosystem (Lehtiniemi and Ruckenstein 2019). And finally, in the context of a project on data-driven government in Denmark, conferences such as *Join the Data-Driven (R)evolution, Driving IT* and *Big Data Week London* all allowed for the socio-material investigation of tech discourses and imaginaries. The *Join the Data-Driven (R)evolution* conference was hosted by the lobby group Danish Industry, whereas *Driving IT* was hosted by IDA, the Danish union of engineers, and is its annual conference on IT. These two conferences thus gave differing but complimentary perspectives on tech discourse from a corporate and union perspective. The *Big Data Week London* conference, on the other hand, was organised as a for-profit conference, paid for by tech companies and recruiters, and provided a view of the international tech discourse in contrast to the local Danish discussions.

From these events, we observed shared challenges of tech conference ethnography, and sought ways to pay attention to and describe both the 'start up' energy and the discourses of hype (Hockenhull and Cohn 2020) present in these conference spaces. We became interested kinds of communities created in the wake of knowledge (Strathern 2004: 30, see also Strathern 2006), and in the way that the conference as a format convenes and shapes experiences of innovation. In the analyses that follow, we draw explicitly material from nine conferences, primarily in four European countries, [ii] which ranged in scale from a few hundred participants to upwards of 3,000. What the conferences had in common was a desire to present themselves, their attendees and speakers as at the cutting edge (Wasson 2005) of their corner of digital technology.

Created to serve both the public and private sector, the conferences informing and appearing in this article were deliberately used by organizers and participants for purposes greater than networking and sharing information. We asked ourselves: What kinds of ethnographic attention do we as scholars need to develop, and what will this bring forward in our work at the self-defined cutting edge of technological worlds? To begin to answer this question, we consider the exiting literature on conferences as a social form and draw out some particularities we see as pertaining to the tech conference.

The Social Life of Conferences

From the poorly attended side-session to the effervescent, personalized pitch of a presentation, conferences are familiar to the point of banality for academics. They are also, as the twenty-first century rolls on, potentially on their way out: the effects of COVID-19 and calls to end the carbon intensive mega-conference gathering are again drawing academic attention to how communities share ideas, and the infrastructures they take for granted in doing so (Pandian 2018, Beard 2019, Felt 2020). This reflexive attention to the conference as a unique form of sociality is a return to an earlier era of curiosity about conferences. In the middle of the last century, the international conference was considered a sufficiently unusual phenomenon to warrant a conference of its own. The first *Conference on International Conferences* was held in 1960 in England. In an ensuing publication, the convener and organizer Mary Capes reflected on the changes in the world that led to her research: 'three international conferences were held in 1853, but in 1900 for the first time over one hundred occurred, and by 1953 they numbered at least two thousand' (2001 [1960]:1). Capes was not alone in her interest. Social scientists, anthropologists and behavioural science scholars were beginning to turn their attention to their own gatherings. Early analyses of conference life, characteristic of the mid 1950s, attempted to use it as an *experimental* site, a deliberate form of group life that modelled laboratory conditions, exemplified 'feedback' loops, or aimed to create knowledge about group dynamics. Foremost of these were the Macy Conferences on Cybernetics (Foerster et al 1950) where organisers stated to their participants that 'each group when it comes together is an experiment' (Fremont-Smith cited in Heims 1991:25, see also Pias 2003). Research on conferences was being fed back into conference action, with behavioural scientists even pinning hopes for improvements in international communication on what might be learned there. As Capes noted at the time, 'to discuss a complex process while actively involved in it was a novel situation and no easy task' (Capes 2001[1960:]:4-5).

This kind of participant observation was, however, familiar to anthropologists. Margaret Mead – an attendee of the first Conference on International Conferences - continued thinking about conferences into the 1960s, with the publication of *The Small Conference: An Innovation in Communication'*(Mead and Byers 1968). Mead places the conference in a long history of social gatherings, from the printing press to modern transportation to the early days of the United Nations. 'Throughout these historical developments', she writes 'there is found the contrast between the formal and the informal, between equality of opportunity to participate and the maintenance of hierarchy' (Mead and Byers 1968:4). She notes the way material environments participate, whether through the 'raised lecture platform' or the 'special seat' (Mead and Byers 1968:4). Yet the scale at which she aims her analysis is a conference of 'a group small enough to sit around one large table, called together for a specific purpose, at a specific place, for a limited time, one, or at specified intervals in a series of designated length to consider new aspects of a specified topic' (Mead and Byers 1968:5).

Today's conferences, as their ethnographers know, far exceed that scale, and exist for quite different purposes. STS ethnographers of policy, expertise and knowledge practices in contemporary institutions have found themselves conducting fieldwork at much larger gatherings (Gross and Fleming 2011, Schwegler 2008, Fortun 2001, Shore and Wright 1997). In their 2017 collection *Ethnographies of Conferences and Trade Fairs*, Leivestad and Nyqvist bring together accounts of medical, art, investor and fashion conferences, arguing that attending to such events means taking seriously the professional world of those with whom they work. Yet what they describe as the experience of the weary professional– 'We have all been there. Tiredly listening to yet another keynote' (2017:2)– was already prefigured in 1960, by Mary Capes. Her mid-century delegate had already 'participated in one conference after another' she observed, and 'feeling disappointed and disillusioned under the impact of conference ennui, [had] retreated from his efforts' (Capes 2001 [1960]:2).

In STS, reflections on methods for meeting scientific and technical cultures where *they* gather go back decades. Recalling a suggestion early in her fieldwork that she 'look at the website', Anne Beaulieu initially felt brushed off – she had not travelled 'to the field' to be told to look at websites she could view from home. Later, however, she realized that the website contained the achievements of the Lab she was setting out to study, that it was 'an object through which to establish relations with the researchers (2010: 464) it was where *they* went. Beaulieu suggests that 'co-presence' shapes fieldwork, inviting STS researchers 'to consider new aspects of knowledge production that may not be strongly tied to a physically defined space such as a lab' (2010: 454). For annual conferences on the move, it is the communities and networks built through their regular gatherings that matter. But conferences strongly shape perceptions of technologies, and approaches to them. At a recent conference, STS scholar Jessamy Perriam described corporate conferences as sites where the future user is configured, where 'a problem-solution relationship with customers is created, and users are encouraged to use products in a way 'considered ideal by the corporation' (Perriam 2017).

Whether the user of a technology is being configured, or a societal need being problematized, tech conferences manifest particular ideas of futures. Those futures may be tacitly, or explicitly elite, tacitly or explicitly inclusive. Such futures may enter the room in the form of prototypes, pitches, business proposals, business models, or other forms of material objects brought along with presenters. Data physicalisations, alongside data visualisations, produce persuasive and engaging sense-making (Jansen et al. 2015, Buur, Mosleh and Nielsen 2018). Conferences, which are transitory and – if serial – often on the move, need an approach that views them as a practice. While Leivestad and Nyqvist focus primarily on *knowledge* exchange and production that happens in conference spaces (2017), through our discussions we have become interested in tech conferences as a particular genre of conference.

Ethnographic work at the tech conference can thus be served by cultivating a critical attention towards how the form of the conference and the character of technology are co-produced. Conferences are rich and dense sites for material that is part of our fields. All ethnographers make `critical decisions' (Strathern 1999: 240) about what fieldwork events they attend and the kinds of things they will do (Jensen 2007:844, see also Candea 2007). However, as STS scholar Casper Bruun Jensen points out, this decision making 'is *shared with other actors* trying to accomplish tasks in environments they are not fully controlling: that is, everyone (Latour, 1999b)' (Jensen 2007:844, emphasis added). Participants in technology conferences, like ethnographers of them, are evaluating what to spend their time on, which panels to see, what stalls to visit. Taking these shared concerns seriously, our analysis argues that how one attends tech conferences matters.

What follows is a joint analysis of tech conferences not only as an increasingly important part of STS fieldwork but a framework for what it is possible to glean from close observation – and analysis – of conference work. We have organized our analysis around vignettes that emphasize the techniques involved in conference scheduling and space, the deliberate and accidental cultivation of tech-space affect, and the different kinds of politics conferences host, and bring into being. Drawn from across our field-sites, our examples are offered as potential heuristics for others to use in their fields, shaping ethnographic attention to the ways that tech conferences shape emergent fields.

Methods Interlude I: A User's Guide

This is the paper we wanted to have read entering the field. Conferences are brief, yet within a research project, form part of a larger methods assemblage. None of us found a satisfactory guide to conducting conference ethnography, nor analyses of tech conferences that could help orient our attention. There may be time clashes: conferences are scheduled – research projects have their own temporalities. As a researcher, you have little control over where or when a conference central to your field may be held, or how much it might cost for you to attend at corporate rates. You find out about conferences by asking – interviewees, colleagues, practitioners. You find out about conferences on industry list-servs, and on Twitter. You find out about conferences on corporate newsletters, even perhaps from your supervisor. Learning to attend conferences takes time, and raises questions for your research. What is your own position within this field? Who are you as an attendee? Are you primarily observing, from a chair in the back row of a darkened room? Or are you on stage with tech practitioners, shaping the conversation at the conference? What happens when you are asked to run a session, perhaps based on your research? What are the opportunities, and costs, of access? How do you introduce yourself and your project to others, in settings where many attendees may have a foot both in academia and in industry? Do you stand out – as a woman in tech? What will you wear? Does the hackathon dress code require a "pizzas and coke" outfit, or a shirt for networking? While conferences feel special for us, they are often "more business as usual than carnivals" (Nyqvist 2017:11). They are part of day-to-day functions, and not as separated from what people do as they might feel for us. Empirically then, we look for carnivalesque ways of performing – where are inhibitions loosed? We reflect through the analysis on the relationship between the ethnographer and the conference, offering these questions as a way of structuring reflection.

Techniques of Conference

In 1968, the anthropologist Mead noted the use of the platform or 'special chair' in select gatherings, to denote elevated or important participants. Following her suggestion to consider the physical techniques deployed by conference organisers, we extend our attention to the careful programming, lighting, staging and music that conference planners use to pace and measure participants' experiences. Conference planners, organizers and participants engage in these – and many more – techniques to 'make conferences happen'. The vignettes in this section from the *MyData* and *MobileWorld* conferences are organised to show three distinct techniques. First, the Plenary places the reader in the opening moments of a conference, a space where speakers make claims about the gathering, perhaps note its history, and create a sense of collective purpose. From framing and setting the scene, our second moment considers a moment in the choreography of scheduling. At large conferences, panels, demos and networking often receive designated program space: our example comes from a 'pitch' event in which start-ups compete for a place in the future. The third and final story shows how conference-goers come online – making the parallel space of digital participation. Across our settings, we found these became repeated and recognizable coordinations of people, establishing a frame for the doing of technological conferences.

The Plenary

The conference is about to begin. I [XX] am at MyData, in Helsinki, Finland, and it is 2018. With my name tag around my neck and welcome package in hand I take a seat with a camera-friendly view of the stage and slides. This opening session has gathered all attendees in plenum – later we will move into parallel sessions and thematic tracks. As an opening plenary speaking to the full gathering, it seeks to accomplish much of what one would expect. The speakers provide a warm welcome, and begin to establish a sense of belonging, creating enthusiasm about the "exciting" programme ahead, sharing practical information, and providing a framing for the conference through this year's 2018 slogan "From vision to action – future now". To my surprise, however, despite its emphasis on making the future present, the slogan prompts brief moments of critical introspection and caution.

"Why do we value 'newness' so much?", asks the first presenter from the MyData team, whom I will call Emily. Addressing the room, the 'we' is the conference gathering of actors from academia, activism, software engineering, business and public bodies all working in the innovation space of personal data that includes digital health. Emily's follow up question is rhetorical, alluding to the risk of historical amnesia: “Isn’t the fact that we value ‘newness’ so much some of what has created the problems in our current personal data management systems?", she asks. In times of rapid and unpredictable change, Emily argues that “slowness” can be considered a quality worth valuing when tasked to “imagine a new ecosystem” for the circulation of personal data.

Emily’s caution about the pace of change was taken up by a second speaker from the *MyData* team, who I will refer to as Hannu. Being among the co-founders of *MyData*, Hannu reminded the room of the slogan used at the first annual *MyData* conference in 2016, two years before: “Make it happen, make it right”. *MyData*, as a movement, Hannu argued, could ‘happen’ from a technical and business stance without it being at all ‘right’ for society. Conversely, Hannu said, overt focus on figuring out how to get *MyData* ‘right’ could lead to a state of paralysis where nothing ‘happens’, leaving *MyData* without “influenc[e]” as other “change agents” take the lead.

This exchange from the opening plenary hints towards an unrest that resides within the *MyData* “family”. Despite their differences, the family members can for the most be boxed as data enthusiasts, but what constitutes ‘right’ is however not necessarily shared among them. *MyData* is interdisciplinary, and in a short period has repeatedly assembled broad societal representation at its events. The plenary however speaks to a point of contention that I observed at each conference between 2017 and 2019, namely how *MyData* should reconcile its origins in data activism with a growing presence of market values. When I attended the annual conference the previous year, 2017, one organizer had claimed that *MyData* was a “special movement” that “had no enemies” since the cause of *MyData* was “logical”. In that year’s closing plenary, the STS scholar Linnet Taylor, whose research focuses on data justice, gave a talk on the unintended consequences of ‘data solutionism’. The opening plenary of the following year could almost have been a direct response to Taylor’s insights. The 2018 slogan “From vision to action – future now” was being treated with a level of critical literacy that had been largely difficult to observe the year before. Thus “making it happen” was essential to achieving something considered equally, if not more important: “influence” over others they considered change agents. This kind of ‘influence’ could be narrated by a conference community to itself, over the course of several years.

As a technique of convening, plenaries are sites where the conference to come – or just completed – is narrated. As STS scholars, we ‘cannot simply repeat the analysis suggested by the actors [s]he is studying’ (Callon 1986:187), so attending to this narration and its subsequent effects in the life of the conference is vital. While the point is designed to develop a critical distance to the (often persuasive) arguments put forward by often impassioned technology advocates, in this example it is enriched by *MyData* being an environment where academics and activists overlap. That the *MyData* community creates a space for socio-critical scholarship to be heard differentiates it from other technology conferences, many of which focus more exclusively on digital innovation and disruption. With its roots in data activism for ecosystemic change, *MyData* blurs the boundaries between academia and industry, with regular shifts in register. These examples of how past conference slogans come to be seen from later years can be viewed as an attempt to manage such tensions by establishing legitimacy for critical thought and action under the same roof with a careful encouragement to slow and shape the pace of change.

The Pitch

Our second illustration of a conference technique takes us in to the annual *Mobile World Congress* (MWC). During this huge event, global companies present their latest devices (market ready, of course) alongside a host of unknown start-ups showcasing inventions and aspirations, many of which are yet to be realized in practice. In 2017, the whole city of Barcelona was turned upside down to host MWC, which unfolded at parallel satellite events across the city. Between the city's distributed venues, *Four Years From Now* (4YFN) ran a scheme for the small, relatively unknown start-ups. "Some find this name amusing", a professional from *Starupbootcamp Barcelona* tells me as our paths cross in the midst of multiple tech showcase booths: "many of the start-ups participating will not exist four years from now!" One way to tell who *might* be is to attend the late night *IoT Stars* event, a startup competition within 4YFN. After a competitive selection round, chosen applicants are invited to present their emergent IoT sensing technologies to a panel of judges.

The colour code throughout the 4YFN 2017 has been purple. It is the colour decorating the logo and the badges of temporary belonging that we are all wearing as we wander around, continuously immersed in purple light. The late-night event, however, glows red and gold with sponsorship from *Estrella*, a local Spanish beer with its characteristic red logo. On a stage in a room packed with lines of chairs placed for the audience, a jury of five people take their seats in a half-circle both facing the audience. One by one, representatives from IoT start-ups are invited on stage to pitch their inventions. Excitement flows from the stage. Each pitch is followed by questions. Jury members (including the Chief Technology Officer (CTO) of Barcelona, a founder of the IoT Council and various tech-innovators) ask the start-ups whatever questions they like. The room is upbeat, but questions from the jury carry an edge: they are challenging, critical, and concerned. After an IoT developer has just presented great confidence the founder of IoT Council asks ‘What happens to the data you collect?’. The developer hesitates. He nervously seeks to formulate a response the jury will accept. Where words were, just minutes before, flowing freely from the self-same person, they now seem stuck, appearing in incohesive fragmented sentences as the IoT developer chooses to argue that this is an important matter to attend to in the *future*. The next eleven pitches proceed in a similar fashion. As the night comes to a close, the jury announces the winners and the tall charismatic host of IoT Stars closes the session with the words: “Let’s make IoT happen!”

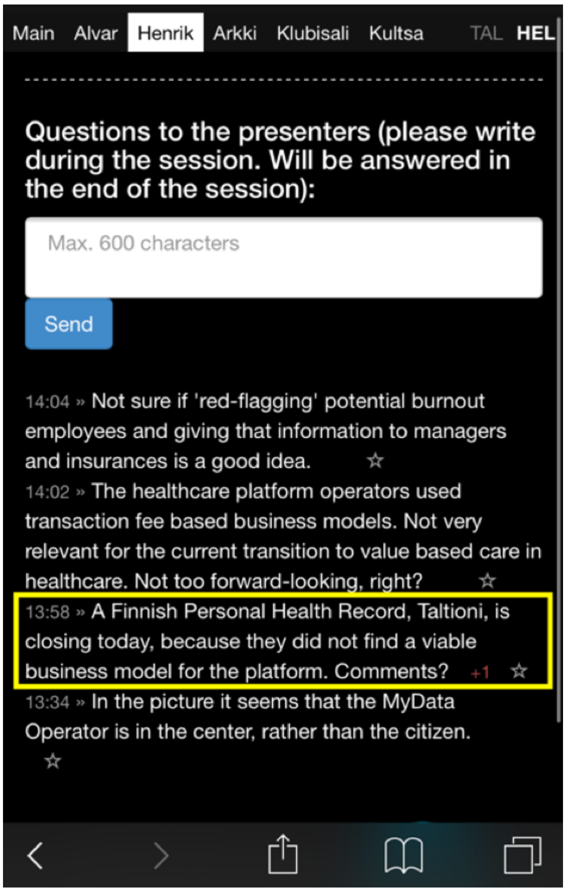
From timing to sponsor branding and colour shifts, name tags, logos and drinks, the IoT Stars event draws on many of the large-scale gathering markers identified by Nyqvist et al. (2017). Here, the formalised ‘pitch’ plays out through staging and theatricality, a big screen, and circle of questioning judges. It is strongly performative – presenters are clearly aware of the genre and have styled both themselves and their products to create a performance that will seem captivating and convincing. Even those who do not win will hope for their company’s name to circulate in tweets, recordings and discussions in the aftermath.

The Digital

Whether live streaming from Twitter, asking conference-goers to use program apps, or live polling devices, tech conferences often make an effort to perform their cutting-edge status. Before a conference begins, conference websites, social media platforms, hashtags and streaming platforms are starting points for exploring promotional material, what attendees say online and videos of pre-recorded sessions.

However, steps are increasingly being taken to integrate technologies into participation. In the conference’s combination of physical and discursive space, digital technologies reshape the topology of dialogue (Topologie der Tekinik 2015, Knox and Nafus 2018), bringing commentary and events from ‘outside’ the boundedness of Conference space in. We return to the MyData conference for this illustration, which has a tradition of crowdsourcing input from the audience via smartphones, tablet or laptop to provide real-time engagement in polls and Q&A sessions. On the second day of the 2017 MyData conference, Helsinki, I [XX] am crammed alongside 40 other conference attendants in a four-hour panel session entitled “The Digital Health Revolution”.[iii] I had chosen the panel to observe and better understand what ideas of revolutionary change were envisioned for the future of healthcare. As the fourth speaker was about to approach the stage, a member from the audience pointed the crowd’s attention to a comment on a large screen, placed right next to the presenter's table. The comment, digitally sent in from the floor, announced the collapse of a Finnish Personal Health Record platform, named Taltioni. It is closing today, wrote the anonymous commentator "because they did not find a viable business model for the platform". The participant asked the panel for comments.

Taltioni, as I later came to understand, had been considered a pioneer of the 'digital health revolution' in Finland. Initially funded by Sitra, the Finnish Board of Innovation (and MyData partner), Taltioni was established as a cooperative in 2012 that provided a technological platform to enable citizens to aggregate and share various sources of health and well-being relevant data based on individual consent. It had aspired to build an "ecosystem" wherein third-party developers could access personal data in exchange for self-care technologies. The timing of this announcement, and the visible mark of a closing platform during a panel about "The Digital Health Revolution", was presumably an unwelcome backdrop for the next speaker. He nonetheless took the stage to pitch his well-being innovation under the presentation title of "Taking MyData into Action", with the digital display of questions to the presenters still visible in the background.



Questions to the presenters (please write during the session. Will be answered in the end of the session):

14:04 » Not sure if 'red-flagging' potential burnout employees and giving that information to managers and insurances is a good idea. ☆

14:02 » The healthcare platform operators used transaction fee based business models. Not very relevant for the current transition to value based care in healthcare. Not too forward-looking, right? ☆

13:58 » A Finnish Personal Health Record, Taltioni, is closing today, because they did not find a viable business model for the platform. Comments? +1 ☆

13:34 » In the picture it seems that the MyData Operator is in the center, rather than the citizen. ☆

Figure 1: Screenshot of audience remarks regarding the closure of Taltioni taken during the Digital Health Revolution session from Screen.IO, accessed via the author's [XX's] internet device.

Here, a platform – celebrated within the community attending – had collapsed in front of the conference-goers. Communicated and made visible through the Conference's Finnish based discussion technology Screen.IO, the speakers on stage were left grappling to explain the "value proposition" of the MyData person centric data management model. The collapse infused the ensuing conversations, which aimed at describing the "Nordic" model and its potential for international export for a white paper. As the panel came to an end, an audience member posed the rhetorical question: "if we cannot get this to work in Nordic countries, then where can it work at all?"

Across these examples are things we are calling techniques of conference. They are material, social, designed in advance and while not specific to tech conferences, often given a particular shape by the aspirations of those holding a conference. Will the event strive to be participatory? Deliberative? Connected from one year to the next? What kind of 'macrostructuring' work (Latour and Callon 1988) in which a world is constructed by actors, will be generated through scales of significance at the opening Plenary? As noted above, in these professional settings, participants are invited to shape how the conference unfolds. Live pitch events perform the competition of the market, live comments reporting on live company collapses show no opposition between the 'digital' and 'real' (Boellstorf 2016). These techniques shape the conference experience for conferencegoers, as sociomaterial artifacts of design, intent and practice. But it is to the experience that we now turn.

Affective Conference

Conferences, as we have just seen, may be physically choreographed and scheduled to produce a specific set of experiences for attendees: the sense of common purpose or togetherness plenary, or competition in designated spaces[iv]. Conferences are lived. We now highlight the affective dimension of that lived experience. As Stark has recently argued, the interdisciplinary affective turn across the social sciences 'provides a set of conceptual springboards for work at the intersection of computer media studies and STS' (2019:119). These conceptual springboards include analyses of sociotechnical entanglement and intimacy (Latimer and Gómez 2019), links to the field of HCI through physiology and emotion as design relevant within computing (Picard 2000, Leahu et al 2008), an increasing attention to the deliberate targeting of affect (Lury 2004, 2009. Schüll 2012) and – by extension – STS studies that focus on the 'orchestration of technique and sentiment' (McFall 2014:7, cited in Deville 2015:13; see also Cochoy et al 2017). It is Deville's work that is closest to our intention with bringing affect and conference together. Presenters draw in audience members and through emotions such as hope or direct enthusiasm. These experiences are embodied, largely 'nonconscious and unnamed, but nonetheless registered, experiences of bodily energy and intensity' (Gould 2010:26, cited in Stark 2019:119). The technology conference is home to a dizzying array of characters, from the passionate true-believer activist to the cool, detached operator networking their way to the next position. Deville ties 'possibilities [that] are not inherent or given, but made and constantly re-made' (2015:12) to STS's longstanding deployment of 'affordance', usually used 'to describe the way in which objects and environments offer themselves up for the conduct of some form of human action' (Deville 2015:12). Both help us attend to the relationship between the orchestrated and emergent.

Performing Passion, Producing Desire

It is after lunch at the Driving IT conference, 2017, and I [XX] have decided to attend a panel in the Start-up/AI track. When I enter the room and take a seat at the back, the presenter, Warren, is already up on stage, first PowerPoint slide on display. Warren is dark-haired but balding, in his late thirties, and wearing a shirt, blazers and jeans. The room has not filled up, as the other sessions have. Undeterred, Warren launches into explaining about his project: a smart kitchen. The smart kitchen, as far as I can tell, amounts to a set of pans, pots and cooking tops with *sensors* inside them, which, connected to an app, could keep cooking temperatures precise. Warren is passionate. He is an engineer at heart, doing the project out of a desire to "create something", develop himself, and of course, make money. His slides recount pitching in San Francisco, visits to his factory in China and testimonials from chefs, test-users and his children. It is a deeply personal and heartfelt presentation, and I find it impossible not to root for him - on some level. However, in the quietness of the almost empty room, the disconnect between Warren's passion and the sparse clapping of a largely absent audience leaves me deflated. Compared to the bustling rooms of other presentations that day, our room felt empty and flat.

Warren is but one example of the many kinds of characters who present at tech conferences. Like the start-up pitches of our earlier section, he performs the person of passion, seeking to make connections to further the agenda of a project in which he is emotionally, imaginatively and financially invested. An important objective is to produce desire for the future in which his smart kitchen is in production, and in peoples' homes. By attending the tech conference, Warren has a venue to simultaneously spar technically with peers, get input on the feasibility of an idea or seek inspiration for a business model that will give the project an edge. A central aspect of technology conferences to consider is the way in which they cultivate this kind of subjectivity, irrespective of whether the prototypes shown or ideas produced actually lead to tangible products. As Lilly Irani has argued, events such as hackathons sometimes accomplish more in the formation of entrepreneurial subjects than they do in terms of tangible solutions or ‘hacks’ (Irani, 2015). Similarly, technology conferences can be said to also form subjects as part-entrepreneurs, part tech-enthusiasts.

Glass, Steel and Concrete

Constrained by requirements of being able to hold large numbers of people, but also drawn by the association with gilt and glitz and fed by the notion that the tech scene holds the key to the ‘next big thing’, it is no surprise that many technology conferences take place in the halls of industry or great constructs of glass and steel marking modern conference venues.[v] These non-places (Augé, 1995) are similar to what Felix Stein, writing on German management consulting, calls a ‘closed-off international corridor’ consisting of ‘incredibly homogenous’ architecture which ‘stressed glass, steel and concrete’ (Stein, 2017:16). Conferences of this genre punctuate the clinical, modernist aesthetic with a variety of professionally designed logos, banners, and merchandise.

Figure 2 & 3: Figure 2 on the left-hand shows the entrance to the *Big Data Week London Conference*, while Figure 3 on the right-hand shows the inside of the Danish Industry main building, where the *Join the Data-Driven (R)evolution* conference was held.

The *Big Data Week London* conference took place in a building within view and walking distance of The Gherkin, Norman Foster's widely recognized 'rocket' that, in the year following its construction was said to inspire 'such a kind of cosmic awe that makes Christianity [represented by St Paul's Cathedral] look a bit like yesterday's faith' (Jencks 2005:13-14, cited in Charney 2007:201). The area was ostentatiously inner-city London, next to the gentrified Spitalfield Markets. A shinning glass escalator leads attendees to sliding doors of glass, carefully decorated with full-size stickers emblazoned with the branding of the conference and its sponsor. Marble floors reflected the stern but polite face of suited guard standing outside. The glass slid aside to reveal something that like a mix between a hotel foyer and an Apple store, complete with lounge furniture, ambient lighting, floor-to-ceiling screens and attentively awaiting staff. I gave my credentials at the reception and they 3D-printed my name-tag on the spot. Conference sponsor booths lined the area where food was offered - canapés, fresh juices and mini-croissants - like a trap, luring in the hungry to be turned into customers or recruits-to-be. The space was clean, well-organised and the food was delicious.

Some tech conferences are ostentatious, catering to a level of service associated with business, surplus and wealth. Sponsorships lubricate networking and recruiters become temporary, gregarious hosts. Other tech conferences, such as DefCon which has been running since 1992 and is the world's largest 'underground hacking conference', aim for subversive aesthetics – DefCon uses a smiley face over skull and crossbones. Similarly, some conferences cater more to DIY sensibilities, focusing on sustainable surroundings or placing itself purposefully in an inhabited but hip area, such as the Copenhagen TechFestival, which symbolically positions itself as part of society, by taking place in an accessible part of the Danish capital. Each location and aesthetic produces a different atmosphere, and suggests different kinds of affective dimensions of 'tech' to which the ethnographer can pay attention.

Expectation

It is the second day of MyData 2017 and I [XX] am at the packed session on "The Digital Health Revolution"[vi] which is coming to an end. A member from the audience has been holding back what emerges now as an emotional and incisive interruption. Narrating her experience with related digital health endeavours in the Swedish public sector, she erupted in frustration how no one had managed to showcase a “viable” business model based on a concrete use case of the MyData principles. She challenged someone, anyone, to prove her wrong. An awkward silence followed. In this silence I recalled the message from the main track of the opening csonference day in Tallinn, Estonia, which had stated that the MyData movement was “happening!” (see Figure 4). I did not think I was alone in feeling the dissonance between the immense possibility of that opening day, and the impossibility held in the collective silence. After a brief exchange of thoughts and collective introspection, the panel resumed the scheduled program, returning to making the revolution of digital health happen.

Figure 4: Author’s [XX] photo from the opening conference day in Tallinn showing the visual rhetoric deployed to cultivate a collective sense of momentum, which in the health sector might be considered to exceed expectations.

Whereas the example with Warren – the smart kitchen entrepreneur – shows how presenting to an audience usually poses an opportunity to promote, test and develop early prototypes and innovations, in the case of the frustrated Swede we see a reversal of roles. The practitioner this time is placed in the audience rather than at the table, and challenges the narratives being put forward.

Scholars versed in innovation theory will be familiar with the temporal rhythm of expectation that oscillates between hype and disappointment in alternating cycles (Borup et al. 2006: 290). Whereas 'hype' usually carries a strong and at times blind[vii] future orientation with it, 'disappointment' is engendered in the present realizations and experiences more broadly (see also Hockenhull and Cohn 2020). In volatile innovation spaces, the distance between hype and disappointment can be very slim indeed, especially when taken into consideration that early technology tends to be packaged in techno-utopian terms (ibid.) where digital health is by no means an exception (e.g. Lupton 2017, Petersen 2019).

Experienced ethnographically, the expectation made present by the opening slides is suspended during the outburst from the Swedish practitioner. Such moments are not necessarily unusual – raising expectations is explicitly the objective of the event. Thinking with scholars from the sociology of expectation, we can see how emergent technology is undergirded by conditions of uncertainty that practitioners – and regulators – must navigate in their quest to establish themselves in markets that are not always present in advance (Borup et al 2006, see also Rotolo, Hicks and Martin 2015). And expectation is shared between the ethnographer and the participant, albeit with different reasons. In both cases, it can be viewed as a sensory device or set of affective sensibilities embodied and enrolled to support knowledge practices. Whereas the goal of the ethnographer is to *describe* the phenomenon in question, expectation is for the entrepreneur a *reference point* for assessing the veracity, robustness and reliability of claims inform future action (Borup et al. 2006, 295). Use cases, proof of concepts and business cases exemplify some of the recognized artefacts that in entrepreneurial settings count as a form of 'evidence' to corroborate promissory claims of technology and thereby instill belief in the community. These artefacts and their vehicles are then a means to not necessarily know, but become less uncertain about instances of what standard to adopt, which investments to legitimize, and who to collaborate with or follow out of inspiration.

Politics of Conference

Anthropologist Lucy Suchman's characterization of politics 'as matters of encounter, ordering and distribution' (Suchman 2016, cited in Gorur et al 2019:1) makes tech conferences staging grounds for technological politics. Through discussion and demonstration, the relationships between funders, policies, national agendas, and global competition are laid out. Of course, the specific technologies themselves have a politics warranting our attention. Whether we listen for the politics in design (Winner, 1999, Benjamin 2019), the struggles to construct and stabilize networks (Callon and Latour, 1984), or the interplay of imaginaries, sociotechnical systems and state politics (Jasanoff and Kim, 2015), politics familiar and new are enacted. Through our analysis, we have come to consider ethnographically salient politics of conferences in two key ways. Our first examples show the ways in which there are politics *in* conferences, which may be overtly done through speeches and invitations, or more subtly unfolding in the guise of conference sociality. Our second demonstrate how conferences can *act* with political intent, taking advantage of the gathering of a broad range of expertises and perspectives to generate statements, manifestos and declarations. These practices often blur in the field, and are sometimes difficult to disentangle. The vignettes of this section demonstrate how these different types of politics can be done in and through conference settings.

Politics *in* Conferences

A central and easily identifiable politics at conferences is the appearance of conventional political actors. A minister giving a talk, a union representative leading a workshop or an elected official participating in a panel debate about the future of a technical field. In the keynote we describe below, a Danish government official was invited to speak on the use of data in private companies. He championed the specific institutions set up to help such companies to participate in external markets and as Minister of Higher Education, from a sitting government which was understood to be pro-business, his talk also highlighted the research activities of several Danish universities.

The minister takes to the stage in the huge auditorium hall; a couple of cameras flash. An experienced public presenter, he gives a confident "good morning," and the audience responds audibly, echoing the greeting. He dives into his speech, talking about thought experiments; google maps in fairy tales, peasant satisfaction indexes in 16th century France. A phone rings and he weaves it seamlessly into the flow of his speech ("the aristocrats are already calling"). He is charismatic, and people laugh. He discusses the data revolution, for which the conference is named, explaining customer-acquisition, unicorns, and how data is the next big resource, uncoupled from geography or climate. He underlines how

Danish companies must join this movement, that is simultaneously a revolution and an evolution, and how the national Innovation Centres can help them think and act internationally. He finishes his talk – ten short minutes in time, but wide in breadth – and people applaud. The minister leaves almost immediately, his assistant in tow.

This is overt politics: elected or leading actors making statements, taking the opportunity to be photographed, using the conference as a stage through which to demonstrate their own technological interest and expertise. Later at the same event, the Danish Foreign Minister also gave a short talk, again highlighting the importance of the State's Innovation Centres, and the recent appointment of Kasper Klynge as the world's first Tech Ambassador (see also Schiølin 2019, Marzouki 2019). These centres were established in different cities around the world (Shanghai, Silicon Valley, Boston, Münich, São Paulo, New Dehli/Bangalore, Seoul and Tel Aviv), as part of Danish export interests, potential venues for the technologies under discussion at the conference. But the conference format provides a venue for discussions between their representatives and Danish companies to talk both business and regulations, trade secrets and opportunities. While actors and settings might not be identified as belonging to conventional national politics, they may still be very much *political.*

These appearances by two ministers are explicitly political: their very presence shows and is interpreted by attendees to focus attention and favour the topic. They exert political muscle to draw attention to the nation's bureaucratic infrastructures (Innovation Centres) and they champion political accomplishments (the Tech Ambassador). The relationship of such explicit politics to tech conferences, whether it be official politicians, organizations or political figures, can be fraught. It is not always the case that governments and officials are welcome. Many wish to keep their technologies separate from the shifting political winds of the day, to avoid politicization that may jeopardise their longevity. Paying attention to when and how political moves are made at tech conferences can highlight how they are easily turned into sites through which existing relations between politics and technology are either displayed, qualified or developed, and how political achievements are staged in such settings

Not all politics at conferences is as overt as political speechmaking - ethnographers know about the importance of informality, the talk that happens in spaces between formally structured events. They also know that sometimes, the important work of brokering of alliances, and the exercise of soft power happens here (Nyqvist et al. 2017:10, see also Aspers and Darr 2011). Conference politics can be more subtle, happening in the "in between" spaces, as conversation topics move between different formats. Let us return to the *IoT Stars* event at the Mobile World Congress in 2017, the pitch session in between formal events, and pick up with the event's closing phrase: "Let's make IoT happen!". What happened in the aftermath?

The *IoT Stars* pitch and jury session has ended, and we the audience are set free. We leave our chairs and, with encouragement from the organizers, explore the venue and enjoy the sponsored beers, music and networking. The next day I return to the Mobile World Congress start-up venue 4YFN. I recognize one of the IoT developers who had pitched last night. He's in front of a booth, holding a coffee in one hand, and apparently trying to get as many free KitKats as possible into the other. I approach him and introduce myself, telling him I'd seen his pitch and we sit down nearby. Over sugar and caffeine, the IoT developer tells me that he got really nervous in front of the jury. It had seemed to him that one of the jury members - IoT Council representative, I know from being present at the event - asked absolutely everyone about data issues. Even before presenting he knew it would be difficult for him to answer a question about data. He tells me that through the IoT devices he makes, he gains access to very sensitive data. Perhaps he notices something cross my face, so he reassures me he will not sell it or anything. There is a pause, and a moment later he concludes that maybe it doesn't matter if he does or not, "there will probably be so much data in the future".

Here, we can see the aftermath. The IoT Council member's sustained questioning about data ethics during the *IoT Stars* event shapes conversations after the pitches are finished, proliferating into the following day. The uncertainty that begins on the public stage seeps into the conference breaks and corridors, haunting the *IoT Stars* participants even after the event has ended. We see how effectively the jury member from the IoT Council has planted unease – likely in more than just this one developer - through his questions. He has taken the opportunity of being in the role of a jury member to push an agenda about data ethics[viii], probing developers on how much they have (or have not) considered the implications of their policies. Ethnographers build relationships, and even those built over a few days at a conference can be valuable in coming to understand indirect political efforts on the part of participants, deliberate planting of conversational seeds, or provocative lines of enquiry. Coming to see these more subtle forms of conference politics draws on the ethnographic skill of building knowledge about the participants involved at the conference, their respective relations, and agendas.

Declarations and Policies

All tech conferences promote agendas embedded with ideas of how the world – current and future – should be shaped. Such ideas can be conservative, largely supporting a continuation of the status quo, but they can also actively advocate for change. The *MyData* initiative is arguably an example of the latter. It promotes an alternative to the contemporary power imbalance that favours data monopolies. In the *MyData* future, individuals regain control over their personal data. While some might consider this a "logical move", (as explored above) it is normative too.[ix] Thus the conference invites its attendees to set the agenda – literally. At the *MyData* conference of 2018, organizers dedicated six hours of the programme to "open space" sessions where attendees "pitched" an idea for their own session in plenum, which interested peers then attended in the chosen room. The 2018 *MyData* conference also invited its participants to set the agenda in a second sense: after the formal programme ended on the second day, participants were invited to participate in a "Founding meeting". Divided into groups, those who stayed live-commented on drafts of the organisation's bylaws or its business plan in an online document. Discussions continued in plenum, along with the legal standing and NGO status of the conference group. And when the conference format provides a moment for attendees to sign a newly formulated declaration from the main stage, it brings its membership together to articulate a common purpose, and create a means by which that document could influence tech agendas beyond the contained time and space of the event.[x] Participatory measures are not a given, particularly not in settings where control over the narrative, direction and structure of the event are paramount. It is more often organised into networking, question and answer sessions, and, exceptionally, digitally mediated Q&A sessions.

Tech conferences are sites where futures are described. To return to the quote with which this section began, Suchman continues her definitional work on politics by returning to the STS position that the 'means' by which politics are done matter, and that other means are possible (Suchman 2016, cited in Gorur et al 2019:1). As STS scholars drawing on her work note, 'exploring how things come to be ordered in a particular way, bringing together particular actors in a particular set of relations… things might still be otherwise' (Gorur et al 2019:2). Conferences shape these actors and relations: an ethnographer might be ideally positioned to ask a difficult question, or might observe different visions playing out over the course of Q&A. Thus the politics of conference – whether explicit, between the lines or written in to the design of the event – help us better understand the dynamics of a given technology as some paths are taken up, and "otherwises" left behind.

Methods Interlude II

> *These vignettes are drawn from our fieldnotes, pulled off our phones and written up from notebooks. Many observations made their way into our articles, here they are put to work in showing the challenges of conference ethnography. Reflecting on how they were made, we draw out practical suggestions: Be open and collect a lot of material. Do your analysis afterwards. Take fieldnotes, and where possible, record. Take photos. Plan breaks and get rest – you do not have to attend everything on offer. Some of our advice is tailored to the timebound nature of conference life: Contact people ahead of time. Be ready for impromptu interviews. Sketch the floorplan and layout. Collect programs and free swag. Keep your nametags, hang them by your desk. But know that conferences are dense. You will always feel you are missing something – even more intensely than regular fieldwork. You may be following a prepared presentation, whilst attending to the subtext of an audience response, noting their questions and simultaneously photographing visual materials used by the speaker. Organise your fieldnotes and consolidate the stuff you have put in different places. Write one-pagers summarizing the conference: when you leave, sit down and give yourself an overall impression. What were the key moments? Are there any initial themes of your conference? Write out vignettes and moments that feel particularly salient, and print materials for your office. Consider free writing and memo-ing, especially if you are working to draw links across multiple events.*

Conclusion

Ethnographers preparing for fieldwork on emerging technologies cannot help but attend conferences. Our article has sought to prepare them. We have drawn out stories from tech conferences, situating them within a longer history of gatherings for knowledge and positioning them analytically as insights into the specific genres of sociality that arise at tech conferences. But we have also sought to differentiate and explore the facets we see as specific to technology, exploring the central role they have in configuring arenas where technological change is performed, done and experienced. “Let’s make it happen” is a refrain heard across our field-sites, simultaneously a summoning of a collective and an invocation of an active future. Through our examples and methods interludes, we have suggested that tech conference ethnography can benefit from certain field attunements and preparation. By way of concluding our analysis, we offer the following questions, building on the ethnography and analysis above, to orient the ethnographic gaze for STS scholars working with technology conferences in their own fieldwork.

Techniques

- How has the conference been planned? What is the scheduling?
- Where are people gathered and separated? How is the conference physically distributed?
- What venue has been chosen for this event? Why? What would be different if it were elsewhere?
- How does the merchandise provided by the conference shape belonging and participation? E.g. branded lanyards, tote bags, stress balls, notebooks, pens and pencils, water bottles, umbrellas, USBs etc.
- What 'special' sessions have been created? E.g. demonstrations, pitch events, networking session, opening and closing panel, keynotes, breakout sessions, sponsored events, parties and afterparties?
- What technologies are being used to mediate and moderate – are people live tweeting, are there dashboards, how are questions being taken, live feedback, conference apps, LinkedIn profiles, email lists, listserv signups, QR codes, etc?
- What technological spectacles have been brought to the conference e.g. VR booths, AR demonstrations, 3D printers, robots

Affect

- What is the feel of the space, and what kinds of affect does this space inspire? What is it indicative of in terms of the aspirations and identity of the conference? Is there any sense of history to the venue? It glass and steel, corporate conference hotel, convention centre, hippy commune, fablab, meetup space – attend to smells, colour, music, jingles
- What kinds of foods and drink are being provided – at booths and poster sessions and in the halls?
- What is the dress code of the conference?
- What is the tone and tenor of conversations and presentations? Is there an entrepreneurial spirit, activist, techno-anxiety, cutting-edge-tech-bro, future fetishism? How (and by whom) are these activated? How do these affects change between the spaces of the conference (e.g. breakouts, keynotes, pitches, charging stations, exhibits, networking lounges, silent spaces).
- Is FOMO a tangible part of how people choose to spend their conference time?
- What is at stake for attendees? What are they trying to achieve by being present? Are they selling? Buying? Promoting? Hustling? Learning? Criticizing? Networking? Going through the motions?
- How is downtime created or managed within the structure of the event? (if at all)

Politics

- Are official representatives of public and private organisations invited? If so, what is their role?
- How is the conference used to make political declarations and statements? E.g. claims about policies, announcements about investments, contestations over funding made by representatives.
- How do presenters draw on ongoing political events, geopolitical events, known politicians? Are these national, international? How do these travel into conversation after the talks?

- What is the relationship between corridor talk and the creation of consequences outside of the conference? Are there conversational seeds planted? How do they travel and grow? Is the keynote resonating in other talks or break times?
- How are challenging issues neutralised or closed down? What role do facilitators have? What topics are described as 'outside the scope' of discussions?
- How is official status conveyed and exhibited through the organisation of the program and space? Are there restrictions on access?

During COVID-19, and in the light of the climate crisis, the future of conferencing as a practice is uncertain. In many fields, experiments with online conferencing have been active for a few years, from the un-conference to further digitally mediated gatherings (Bastian 2020, Felt 2020). In 2018, anthropologist Anand Pandian, reflecting on the first online conference of the American Association for Cultural Anthropology, asked the prescient question, 'could a platform be devised that would give participants a sense that they were sharing a common experience in time, even at a distance from each other?' (Pandian 2018: np). To announce the end of conference culture would be to risk a premature, performative obituary. But the sudden changes of 2020 have shown that technology and conference life are more intertwined than ever, with new habits, affects and politics emerging in these spaces[xi]. As scholars of technology and its emergences, we should be ready to enter them ethnographically, and see how people "make it happen" in the years to come.

References

Author 2020

Aspers, P and Darr, A (2011) Trade Shows and the Creation of Market and Industry. *The Sociological Review* 59(4):758–778.

Augé, M (1995) Non-Places: Introduction to an Anthropology of Supermodernity. London, Verso. Trans John Howe.

Bastian, M (2020) Synchrony at the Online Conference. http://www.michellebastian.net/home/synchrony-at-the-online-conference (accessed 2 July 2020).

Beard M (2019) The death of the conference. In: *Times Literary Supplement.* Available at: https://www.the-tls.co.uk/the-death-of-the-conference/ (accessed 26.02.2020).

Beaulieu, A (2010) From co-location to co-presence: shifts in the use of ethnography for the study of knowledge. *Social Studies of Science* 40(3): 453-470.

Benjamin, R. (ed) (2019) *Captivating Technologies: Race, Carceral Technoscience and Liberatory Imagination in Everyday Life*. Durham, NC: Duke University Press.

Boellstorff T (2016) For Whom the Ontology Turns: Theorizing the Digital Real. *Current Anthropology* 57 (4): 387–407.

Borup M, Brown N, Konrad K and Van Lente H (2006) The sociology of expectations in science and technology. *Technology Analysis & Strategic Management* 18(3-4): 285–298.

Buur, J, Mosleh SS, and Nielsen CF 2018. Physicalizations of Big Data in Ethnographic Context. Ethnographic Praxis in Indusetry Conference -Honolulu.

Callon M and Latour B (1981) Unscrewing the Big Leviathans How Do Actors Macrostructure Reality. In: Knorr K and Cicourel A (eds) *Advances in Social Theory and Methodology. Toward an Integration of Micro and Macro Sociologies.* London: Routledge, pp. 277–303.

Callon M (1986) Some elements of a sociology of translation: domestication of the scallops and the fishermen of St. Brieuc Bay. In: Law J (ed) *Power, action and belief: a new sociology of knowledge?* London: Routledge, pp. 196–223.

Candea, M (2007) Arbitrary locations: in defence of the bounded field-site *Journal of the Royal Anthropological Institute* 13(1); 167-184.

Capes M (2001 [1960]) *Communication or Conflict: Conferences: their nature, dynamics and planning.* London: Routledge.

Charney I (2007) The politics of design: architecture, tall building and the skyline of central London. *Area* 39 (2): 195-205.

Cochoy F, Deville J and McFall L (2017) *Markets and the Arts of Attachment*. Oxford: Routledge.

Danyi E (2015) The Parliament as High Political Programme. In: Müller AL and Reichmann W (eds) *Architecture, Materiality and Society: Connecting Sociology of Architecture with Science and Technology Studies*. London: Palgrave Macmillan, pp. 99-119.

Deville J (2015) *Lived Economies of Default: Consumer Credit, Debt Collection and the Capture of Affect.* Oxford: Routledge.

Felt, U (2020) EASST letter to members. https://easst.net/wp-content/uploads/2020/05/EASST-letter-May-2020.v2.pdf accessed October 16, 2020.

Foerster, H von, Mead, M., and Teuber HL (1950) *Cybernetics: Transactions of the Seventh Conference*. New York : Josiah Macy, Jr. Foundation

Fortun, K (2001) *Advocacy After Bhopal: Environmentalism, Disaster, New World Orders.* Chicago: University of Chicago Press.

Gross, N and Fleming C (2011) Academic Conferences and the Making of Philosophical Knowledge. In Charles Camic, Neil Gross and Michèle Lamont. *Social Knowledge in the Making*. Chicago: University of Chicago Press.

Heims S (1991) *The Cybernetics Group.* Cambridge: The MIT Press.

Irani, L (2015) Hackathons and the Making of Entrepreneurial Citizenship. *Science Technology and Human Values* 40(5); 799-824.

Jansen Y, Dragicevic P, Isenberg P, et al. (2015) Opportunities and Challenges for Data Physicalization. In: *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI),* New York, NY, USA, 18-23 April 2015: 3227–3236. New York: ACM

Jasanoff, S and Kim, SH (2009. 'Containing the Atom: Sociotechnical Imaginaries and Nuclear Power in the United States and South Korea.' *Minerva* 47(2); 119-146.

Jencks C (2005) *The Iconic Building*. New York: Rizzoli.

Jensen, C B (2007) Infrastructural Fractals: Revisiting the Micro-Macro Distinction in Social Theory *Environment and Planning D: Society and Space* 25(5); 832-850.

Knox, H and Nafus, D (2018) *Ethnography for a data-saturated world.* Manchester: Manchester University Press.

Kuikkaniemi, K, Poikola, A. and Honko, H. 2015. Mydata: A Nordic Model for Human Centered Personal Data Management and Processing. Working Paper.

Latimer J and Gómez DL (2019) Intimate Entanglements: Affects, more-than-human intimacies and the politics of relations in science and technology. *The Sociological Review Monographs* 67(2): 247–263.

Lehtiniemi T and Ruckenstein M. (2019) The social imaginaries of data activism. *Big Data & Society* 6(1). DOI: 10.1177/2053951718821146

Lupton D (2017) *Digital Health*. London: Routledge.

Mead, M and Byers, P (1968) *The Small Conference: An Innovation in Communication.* De Gruyter, Berlin.

Strathern, M. (1999) *Property, Substance and Effect: Anthropological Essays on Persons and Things.* Athlone Press, London.

Strathern, M (2004) *Commons and Borderlands: Working Papers on Interdisciplinarity, Accountability and the Flow of Knowledge*. Sean Kingston, Oxon.

Strathern M (2006) Knowledge on its travels: dispersal and divergence in the makeup of communities, *Interdisciplinary Science Reviews* 31(2): 149-162

Marzouki M (2019) Techplomacy: Towards an Increasingly Privately Ordered Digital World Politics? In: *The 60th ISA (International Studies Association) Annual Convention, International Studies Association,* Toronto, Canada, Mar 2019. Available at: ⟨halshs-02076612⟩

Müller AL and Reichmann W (2015) *Architecture, Materiality and Society: Connecting Sociology of Architecture with Science and Technology Studies*. London: Palgrave Macmillan.

My Data Global (n.d.) Founders. Available at: https://mydata.org/founders/ (accessed 24 February 2020).

Pandian, A (2018) Reflections on #displace18 https://culanth.org/about/about-the-society/announcements/reflections-on-displace18 accessed September 19, 2020.

Perriam, J (2017) Corporate conference workshops as settings to configure the user. Paper presented at 4S Annual Meeting *STS (In)Sensibilities* Boston, MA. USA.

Petersen A (2019). *Digital Health and Technological Promise.* London: Routledge.

Pias, C (1963) *Cybernetics: The Macy Conferences 1946–1963: The Complete Transactions.* Diaphanes, Zurich.

Rotoli D, Hicks D and Martin BR (2015) What is Energinet Technology? Research Policy (44)10:1827-1843

Schiølin K (2019) Revolutionary dreams: Future essentialism and the sociotechnical imaginary of the fourth industrial revolution in Denmark. *Social Studies of Science*. Epub ahead of print 29 August 2019. DOI: 10.1177/0306312719867768.

Schwegler, T (2008) Trading Up: Reflections on Power, Collaboration and Ethnography in the Anthropology of Policy *Anthropology in Action* 15, 2: 10-25.

Shore, C and S. Wright (eds) (1997) *Anthropology of Policy: Critical Perspectives on Governance and Power.* London: Routledge.

Stein, F.(2017) *Work, Sleep, Repeat: The Abstract Labour of German Management Consultants*. London: Bloomsbury Academic.

Topologie der Technik (2015) *Topological Manifesto*. Darmstadt: Topologie der Technik.

Gorur R, Hamilton M, Lundahl C and Sundström Sjödin E (2019) Politics by other means? STS and research in education. *Discourse: Studies in the Cultural Politics of Education* 40(1): 1-15.

Stark L (2019) Affect and Emotion in digitalSTS. In: Vertesi J, Ribes D, DiSalvo C et al. (eds) *DigitalSTS: A Field Guide for Science & Technology Studies*. Oxford: Princeton University Press, pp.117-135

Van Lente H (2012) Navigating foresight in a sea of expectations: lessons from the sociology of expectations. *Technology Analysis & Strategic Management*, 24(8): 769–782.

Wasson C (2005) Celebrating The Cutting Edge. In*: Ethnographic Praxis in Industry Conference Proceedings*, Redmond, WA, USA, 14-16 November 2005: 140-145. Redmond, USA: EPIC

Winner, L)1999) Do Artifacts Have Politics? *Daedalus,* 109(1): 121-136.

---

[i] In the October 2018, the *MyData Global* was established as a legal entity (MyData.org n.d.) where 2020 will mark a break in tradition as the annual conference will be held in four continents (Latin America, Asia, Europe and Africa) over the course of the year.

[ii] Our ethnographic work has a European focus. However, important comparative work needs to be done with the formalized components of international conferences as well as regional tech conference cultues in India, China and the United States.

[iii] The name of this track was taken from the title of an interdisciplinary project funded by Tekes, the Finnish Funding Agency for Technology and Innovation, that was active from 2014-2018.

[iv] Many conferences are indeed explicitly set up to offer attendees specific agendas or ideas, as we shall return to in our next section.

[v] For work at the intersection of architecture and STS see Müller and Reichmann 2015, and Danyi 2015.

[vi] See above for the origin of the track name.

[vii] While hype can evidence and indeed contribute to a severe case of "historical amnesia" (Borup et al. 2006, 290), this is not necessarily always the case since hype comes with its own rituals of production to obtain desirable effects in specific environments (see Hockenhull and Cohn 2019).

[viii] The *IoT Stars* event took place in 2017, a year after the General Data Protection Regulation (GDPR) had been agreed, but a year before it came into force, making questions of data collection, use and erasure pressing for all who would soon need to comply with it.

[ix] The presentation from figure 4 is entitled Mastering your Data can Benefit to Everyone, which (misleadingly) frames the MyData movement as an apolitical one where 'everyone wins'.

[x] Participation should however not be conflated with influence, as the former does not always result in the latter (see Valtysson 2014). Valtysson B (2014) Democracy in disguise: the use of social media in reviewing the Icelandic Constitution. Media, Culture & Society, 36(1): 52-68.

[xi] The Silent Conference, a "hot new trend" was announced in February 2020, showing that conference genres are themselves sites for technological innovation. For more see https://agentmajeur.com/mission/silent-conferences-axa-research-fund/

Sentinel Vision: Data Collection, Disease Prevention and Professionals in Danish Diabetes Healthcare

John Mark Burnett and Rachel Douglas-Jones

Abstract

Sentinel, the software at the centre of this article, is a decision support technology designed for general practice. Developed on the basis of early potentials of electronic health records (EHRs) for improving diabetes management in the late 1980s, the technology went on to become the basis for data intensification in Danish general practice. In this article, we trace what we argue is the reshaping of professional medical vision by recounting the history of its introduction, and its envisaged impacts on managing responsibility, risk and treatment futures for diabetes. As new actors gained access to diagnostic criteria, new fields of visibility opened up. We argue that preventative logics accompanying increased data-intensification in healthcare today have longstanding roots, and suggest that it is generative to trace the coimbrication of professional values, technological solutionism and the building of infrastructure for preventive healthcare.

Keywords: Data Intensification, Health Data, Sentinel Technology, Prevention, Professional Vision

> A patient came to visit me who had an ulcer on his leg, a foot ulcer. I looked in the journal and saw that he had diabetes and also that it had been three years since his last visit. And that is where you conclude he hasn't been to check-ups. Okay, that might be his fault, but it is certainly also my fault too. I should in one way or another have followed up and made sure that he showed up so we could have had him better regulated so he could have avoided that foot ulcer. That is why it is said that you need to have data, you need to have knowledge about what patients you have in your practice. And you need to have a hold on the vulnerable patients in particular whom otherwise do not show up. Because *that* is where you can make a difference. (Interview, Henrik Schroll, 2017)

In chronic disease care, keeping track is a long-term challenge (Manderson and Warren 2016). The idea of 'noncompliance' to treatment foregrounds health provider concerns (Hunt and Ahar 2001: 347), while discussions of responsibility often individualise, missing structural constraints on patient action and engagement (Trnka and Trundle 2017). As former GP Henrik Schroll speaks in this interview extract, we hear the voice of a doctor working with a diabetes patient over time, reaching for knowledge, wanting to 'make a difference' to the progression of their disease. His reflections on *following up,* on *check-ups* and on the possibility that this foot ulcer presented to him *could have been avoided* move directly into *having a hold* on patients you do not see, the ones who do not show up, the particularly vulnerable. Schroll places data in this role, anchoring the capacity to know, manage and ultimately transform the provision of care.

The development of Sentinel Data Capture (hereafter "Sentinel"), the technology at the centre of this article, was led by a small group of pioneering GPs to support quality improvements in the delivery of diabetes care. The interviewee, Henrik Schroll, held a central role in the development of Sentinel, envisioned to support fellow GPs keep track of patients, follow up and ultimately prevent the progress of disease. In this article, we build on his account of Sentinel's development, sharing a walkthrough of the system as designed, to trace its emergence and logics through a lens of technological and professional vision. We argue, building on an in-depth interview with Schroll, augmented through document analysis of official documents, professional journal articles, and news media, that this software history is important. Its importance rests partly through the way it introduced a novel form of chronic disease management into general practice, but also because, in the years following its development, Sentinel went on to form the basis for a range of data collection technologies within Danish public healthcare. Building the basis both for preventative disease intervention and data intensification in the sector, the capacities of the software were significant, and over time, more disease classifications were collected, arguably without legal basis (Burnett and Douglas-Jones, forthcoming). Today, Sentinel as it was originally designed is no longer in use, its use ceasing in

September 2014, following a public debate in Denmark about the use, re-use and broader role of primary care data in research.

In 2005, as Sentinel was under development, about one-third of the Danish population was estimated in a policy report to suffer from chronic disease, posing a serious threat to public health (Kronisk Sygdom, 2005: 32, see also Moth 2012). The pressure in the national healthcare system was reported to be most urgently located among hospitals where complications from diabetes and other chronic conditions occupied the vast majority of finite clinical resources. With the amount of people living with chronic conditions expected to grow, investment in data-intensive logics were expected to curtail the costly hospitalisation of chronic patients by strengthening prevention in primary healthcare, thus mitigating the incidence of comorbidities (OK-2006). Since general practice is generally the first point of contact to the national healthcare system for patients in Denmark, GPs are the 'gatekeepers' to other healthcare providers, which patients can access through referrals when health problems for instance escalate to the need of specialist and hospital care. While general practice in Denmark is a private enterprise, GPs operate as a contractual extension of the national healthcare system as self-employed entities remunerated per capitation (number of patients) and per fee-for-service (e.g. consultation or vaccination) from public funds.

Living with chronic disease is becoming increasingly common (Manderson and Wahlberg 2020, Hvidberg et al 2020) and at the same time digital practices are increasingly being taken up both by patients (Kingod et al 2017, Kingod 2018) and physicians (Kaufman 2010, Sieverdes 2013) as a means of managing chronicity. The promises of digital management of chronic disease are substantial; in their summary of the development of different kinds of technologies for diabetes, from smart monitoring to the use of AI in detecting diabetic retinopathy Fagherazzi and Ravaud argue that we are

> [m]oving from a world in which patients are characterized by only a few recent measurements of fasting glucose levels and glycated haemoglobin to a world where patients, healthcare professionals and research scientists can consider various key parameters at thousands of time points simultaneously (Fagherazzi and Ravaud, 2019: 322).

They contend that in this new world of digital monitoring, diabetes can be differently ‘prevented, managed and characterized’ (Fagherazzi and Ravaud 2019:322). As we trace some of the starting points for these hopes for technological management of diabetes through the development of Sentinel Data Capture, we investigate the making of a technology upon which Fagherazzi~~e~~ and Ravaud’s present vision for the future rests. We examine the introduction of preventive logics via systems of quality management, and argue that it is necessary to trace the coimbrication of professional values, technological solutionism and the building of infrastructure for preventive healthcare. In what follows, we explore what ideas were embedded in Sentinel’s design, interrogate the particular ways of seeing it created for professionals, patients and administrators, and trace the subsequent social and political ramifications as the capacities of the technology expanded over the years. The article is organised around different ways of seeing that Sentinel brought into being for general practitioners, patients and administrators. From a walk-through of the test software identifying how the data collection shaped where to look, who to target and how to intervene, we trace the way that the ‘lighthouse’ beam extended visibility and sight into unexpected domains, offering an analysis of the politics of illumination. To provide a conceptual and practical background to the software tool, we begin our analysis with a review of sentinel technologies in healthcare.

Understanding Sentinel Technologies in Healthcare

Sentinel is a semantically rich concept. A sentinel stands watch, on guard, senses on alert, monitoring (Oxford English Dictionary, 2022). Human or machine, they may detect the otherwise imperceptible, through their distributed watch or their sensitised anticipation of threat (Keck and Lakoff 2013: 2). Within medical establishments around the world, the term has been attached to projects of disease surveillance at a population level. Today, sentinel techniques increasingly make use of electronic health data to generate datasets upon which to conduct computational modelling and predict disease progression from those ‘on watch’, reporting back. Their construction may be prompted by a specific and intense event: for example, when chikungunya broke out in the Indian Ocean in 2007, Madagascar developed a new sentinel surveillance system that relied on daily reporting of presenting cases through encrypted text messages, the ‘first nationwide real-time-like surveillance system ever established in Madagascar’ (Randrianasolo 2010: 1). Sentinel systems more generally operate in the background, watching. The European CDC maintains a clinical surveillance of influenza using a network of ‘sentinel general practitioners’, which it figures at ‘1-5% of physicians working in the country or region’ (ECDC 2022). Similarly, sentinel is the name taken by the United States medical product safety monitoring mechanism, introduced to build a multi-sited database for devices (US FDA 2019). And it also gives its name to the watch that the Canadian Primary Care Sentinel Surveillance Network (CPCSSN) has over its population, founded to store 'longitudinal electronic health records' on a national level (Ehsani-Moghaddam et al 2019, see also Williamson et al. 2014).

Being able to detect and anticipate disease outbreaks or drug side effects through population surveillance has been well described throughout biomedical literature, with systems looking for 'sentinel events', markers that sound the alarm, 'whose occurrence serves as a warning signal that the quality of preventive and/or therapeutic medical care may need to be improved' (Rutstein et al 1983). Within the sphere of primary care, sentinel systems are complex socio-technical infrastructures, with their own literature. They sit at the intersection of preventative medicine, the increasing use (and capacities) of information technology in the sector, and the temporal chronicities of specific diseases.

Prevention is a key concept operating within diabetes care, with widespread studies exploring forms of effective intervention (Lindstrom et al 2006) and researchers increasingly exploring technology based interventions not only for the management of diabetes (Quinn et al 2011), but to slow its onset (Grock et al. 2017). Since Schroll's encounter with the unfortunate patient with the foot ulcer, much has changed regarding how information is registered and used in Danish primary care. The adoption of information technology, particularly electronic health records (EHRs) marked an epochal shift from paper to digital with repercussions for informational practices throughout the sector. However, a review of the co-evolution of norms in diabetes management and their entwinement with digital systems gives us some insight into the infrastructure Schroll was aiming to build.

Schroll set out to tackle the handling of diabetes through general practice. Diabetes is a chronic condition which, left untreated, can lead to lifelong complications. Those diagnosed with the condition are required to 'manage' their health, food intake, activity levels, in order to ensure steady blood sugar levels in the absence or ineffective working of insulin (Zimmet et al 2014). Among the early adopters of electronic health records, Schroll saw a potential in data to engineer a new way by which GPs could monitor and, crucially, keep a "hold" of diabetes patients less vulnerable to the shortcomings of human memory – and especially for the marginalised no-shows since, as Schroll put it, that is "where [GPs] can make a difference". Marginalisation often figures heavily in discussion of the management of diabetes, which tends to draw contrasts across cultures of individualization and responsibility. Working with elderly Russian Jewish émigrés, for example, Borov and Hine found that the United States, regime of biomedical 'self-management' fails – and is consequently classified as 'noncompliance'– due to distinct readings of the values of individual agency and self control (2008:1). In her work with Turkish migrants in Berlin, Germany, Guell describes 'self-care at the margins' to describe the role that migrant patients took as experts and technology users to make 'chronic illness…manageable and tangible where formal support by the German healthcare system was inadequate' (2012:518).

Improving the quality of disease management started with making data management more *manageable*. Sentinel originated in 2001, when Schroll finalized his doctoral study on methods for registration and analysis of episodes of care in general practice using ICPC codes (Schroll 2001). There, his interest was in supporting data collection for research and quality improvement *without* placing additional information registration burdens on GPs. Two years later, he found himself at the helm of the IT operations – and later the organization itself – in what is today known as the Danish Quality Unit of General Practice (DAK-E hereafter). After the development of a prototype between 2004 and 2006, between 2007 and 2014, this unit implemented the pilot project on a national scale under the name Sentinel Data Capture. During the interview from which the opening quote is extracted, Schroll pointed to a painting of a lighthouse on his office wall, received as a retirement gift. Sentinel, he explained, has an etymological link to "lighthouse" or "watchtower". "I have always leaned more towards the lighthouse, he continued, as it "illuminates and shows which way you should go, so you avoid running into problems".

A successful lighthouse beam, illuminating the darkness of the unknown, was to be made up from the assemblage of various elements: GPs, by changing the way they registered patient visits, could collect data more systematically, data that computers could then reflect back to them. To do this more systematically, the International Classification of Primary Care (ICPC) needed to be translated into Danish. Schroll shared a prominent role in translating the first version of ICPC into a Danish setting during the 1990s (Bentzen and Schroll 1992) and travelled across the nation with his team to train GPs through the provision of courses and more (Bentzen and Schroll 1990). As the decade rolled on, Schroll's primary interest in the ICPC was the prospect of access to uniform data from general practice to acquire statistical knowledge on disease pathway and, crucially, explore its utility for quality development[1]. In our interview, Schroll described what was attractive about the classification scheme from the perspective of GPs at this early phase:

---

[1] This exploration was piloted in the FLUKS project where 42 GPs voluntarily endured the manual labour of data registration in exchange for statistical feedback reports produced by specialists. See **ref**

> And the reason for using ICPC, my best reason, and that was the one people bought, it was actually… When the paper journal was used, you generally wrote nothing. You wrote: '18/03/92, Tonsillitis, Penicillin '. That was a consultation. When IT was beginning to be used, it [the documentation requirement] started to grow. Eventually, we got discharge letters from the hospitals, which we had only received on paper before, and we received all sorts of other things electronically. So we ended up with a journal that grew voluminously. Vehement.
> […] If an ICPC diagnosis was used for each consultation, then it was possible to say 'L03' - that is, lumbar spine problems - I will filter everything between all the consultations with all contacts associated with 'L03'. And then you will see that the 30- 45 screen pages have become one or 1.5 screen pages.
> […] What this means is that you can suddenly get the overview back in consultation in the voluminous journal that was growing. That was my best argument for the doctors: You should use diagnostic codes because you will get the overview back and you will be able to manage the consultation, which you previously couldn't - or you were about to lose - because there had been such a huge success with the communication. You get so much information from hospitals' outpatient clinics, hospitalizations, etc. And it [the journal] constantly grows. That was my best argument.

The introduction of IT in general practice presented new ways for digital information to flow at unprecedented rates and routes, which in turn created new challenges in the process. The proliferation of documentation requirements and access to dispersed patient information (from new actors) contributed to what Schroll describes a "tremendous" growth in the journal volume, resulting in more work and a heightened risk of information overload for the GPs individually. As Schroll told his colleagues, having access to structured digital information and query mechanisms or algorithms to isolate information relevant to a specific diagnostic code however provides a way to render journals, and thus also consultations, more manageable.

The implementation of ICPC was however only part of the work needed to establish the national setup of data-driven quality development of chronic care that was established in years to come. At the time, ten or twelve different electronic patient record (EPR) vendors were on the market with distinct preferences of how information should be stored. This meant that the development of an algorithm could not be transferred across doctors subscribing to vendors with incompatible EPR systems, thus complicating any ambition of national scalability. Rather than reinventing the wheel for each EPR system, Schroll and his associates accomplished the challenge of interoperability through the provision of SDC that offered a common tool for the entire primary healthcare system.

Using “Sentinel” as a prefix to Data Capture, Schroll explains, underscores the core design principle of the system from a technical standpoint: it should never be necessary to enter the same data twice and always accessible from one place only. This paraphrasing highlights principles of centralised data storage and data recycling that have underpinned policies and programmes on digital modernisation in the public sector in Denmark for the past years. Key rationales include efficiency gains from the elimination of parallel registrations that aside from the extra data work paves way for inconsistent data entries and formats that can be problematic for epistemological and integrational reasons. Sentinel was designed to deliver similar gains in general practice where these principles formed the building blocks for a system that would manage to capture structured data from general practice and render them amenable for reuse in a uniform manner for the purpose of quality development and research.

Our analysis of Sentinel as a system that shaped practice draws inspiration from Charles Goodwin’s publication *Professional Vision* (1994). For Goodwin, ‘professional vision’ refers to “socially organized ways of seeing and understanding events that are answerable to the distinctive interests of a particular social group” (Goodwin, 1994: 606). In attending to practices of coding, highlighting and graphical representation in archaeology and legal argumentation, Goodwin argues that these shape certain ways of seeing. It is these kinds of ways of seeing, set up through Sentinel, that we follow in this article. From codes to the classifications they rely on, visualisations and early dashboards, Sentinel carried the capacity to re-shape professional vision, and not just for GPs. As scholars who have taken up Goodwin’s work have observed, professional vision is an enskilled practice, and it is not individual. It reflects expertise of social value to specific professional communities, which “not only convey[s] ideas, meaning and beliefs, but configure[s] them” (Grasseni, 2007: 5). To illustrate our argument that Sentinel had a role in configuring beliefs about quality of diabetes care, we also draw on Barbara Maria Stafford’s essays on the virtues of images. Written in the 1990s during the ‘epistemological uncertainties and educational upheavals of an electronic future’ (1998:4), Stafford’s work offers us ways of examining the power of visualising, a contemporary phenomenon deserving of significant critique (Thylstrup and Veel 2017, D’Ignacio and Klein 2020). As such, to look for spaces where professional vision is reshaped and redirected requires attention to the ways that rationalising technologies like Sentinel are discursively promoted as ‘good looking’ (Stafford, 1998).

Much of Sentinel’s interventions into diabetes management support treatment decision-making, as much as ‘keeping track’. STS scholar Marc Berg has argued, that the prospect of “transforming the 'art' of medical decision making into a 'science'" (Komaroff 1982)” by means of technology has long had its advocates and critics (Berg 1997: 4). To configure beliefs also means ‘transforming practices in the process of acting on them’ (Higgins and Larner 2010: 7). In order for protocols to work, the elements that constitute medical practices must be disciplined ‘to behave in a uniform, stable and predictable way’ (Berg, 1998: 234). As we explore below, introducing Sentinel to work practices in general practice is arguably a starting point in the study of what is lost, gained or changed by other means departing from actual work practices. This work-practice archaeology is, we suggest, necessary as the uptake of algorithmic automation expands to new professional spaces (Pasquale, 2015; Slota et al., 2020; Beer, 2019). Attending to the reshaping of professional vision helps us strive for a ‘double vision’ (Haraway, 1991; Verran, 2001) that sees what is present cognizant of what is absent, in particular, the fullness of biomedical social worlds (Cruz 2020). Such politics also apply to the industry of data analytics as its advancement into new professional spaces call into question established ways of seeing and doing (Beer, 2019, Ruppert et al., 2017). Reviewing the introduction and sociotechnical workings of Sentinel thus provides an opportune chance to explore what such advances look like in the clinics of Danish general practice and understand the politics involved when data-intensive modes of seeing gain authority in new spaces.

Managing Risk

In the opening interview extract, we caught a glimpse of the possible consequences of disease progression when clinical responsibilities of diabetes care fall out of sight. The problem that Schroll identified was not merely *anecdotal*, limited to the unfortunate encounter with the potentially avoidable foot ulcer. As the years went by, Schroll became convinced that the problem was *systemic* based on the organisational practice of information management in general practice.

Whilst teaching medical students in the late 1980s, Schroll was confronted with a line of questions that he struggled to answer. How many diabetes patients did he have? How well were they treated? Could he document his treatment? He realised that despite his conviction that his practice adhered to professional guidelines, he could not document the quality of treatment, let alone answer how many diabetes patients he had in his practice. In the interview, he explained: "we operated with a paper system[…]. You could remember Mrs. Hansen, Mr. Sørensen etc., until you had one-and-a-half handfuls[…], because it had to be remembered in the head, there was no system to look it up from”. With a patient population of around 1600 in his practice, he knew from the national prevalence of diabetes that 56 to 60 should be diabetics. But for Schroll, the paper-based organisation of patient records did not accommodate his finite memory to gain a systematic way of monitoring and managing diabetes among his patient population to for instance follow up on those who had not attended check-ups.

As general practice began to embrace EHR, Schroll and a small group of pioneering GPs saw a potential in digital technology. Drawing on the lighthouse metaphor, I am told by my Schroll that Sentinel was to provide data analytics or feedback “that illuminates for the individual user ‘what is it I have done well’ and ’what is it that I can do better’” based on data from their own practice. In anticipation of what I address below, critical scholarship would notice how “what is done well“ and can be done “better” necessarily valorises a particular visual discourse of 'good looking‘ (Stafford, 1998). Before we proceed with this analysis, it is first useful to introduce the technical and organisational setup in support of Sentinel.

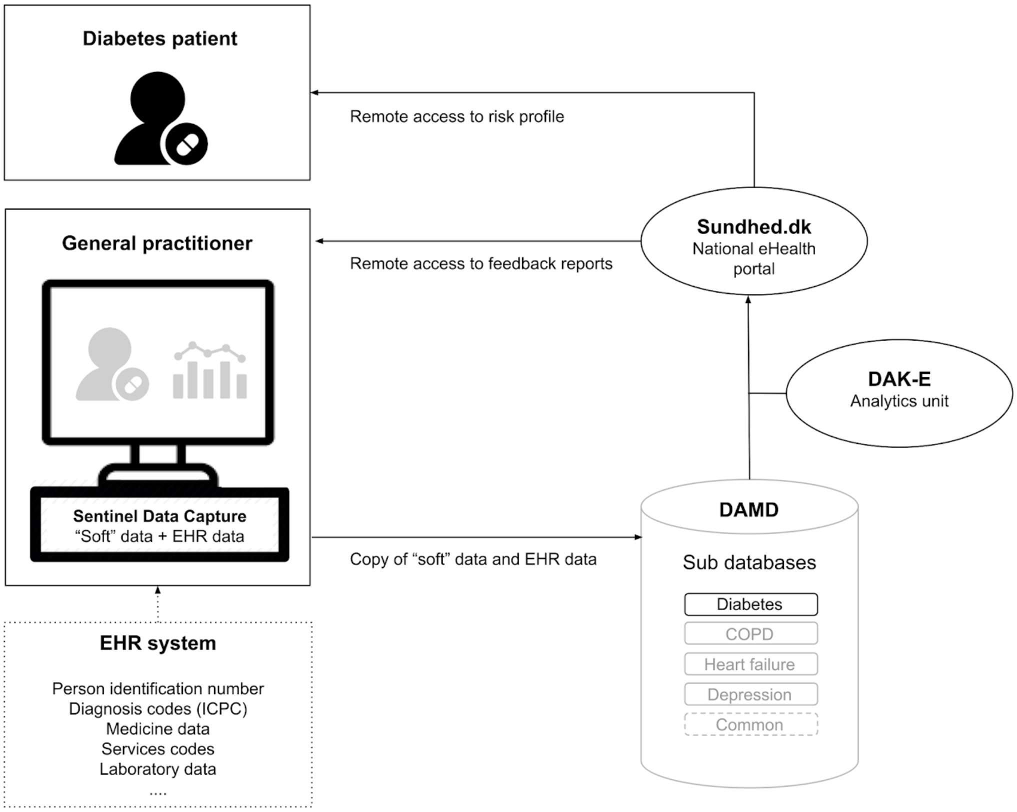


Figure 1 - Author's diagram of the technical and organisational landscape regarding the provision of feedback reports

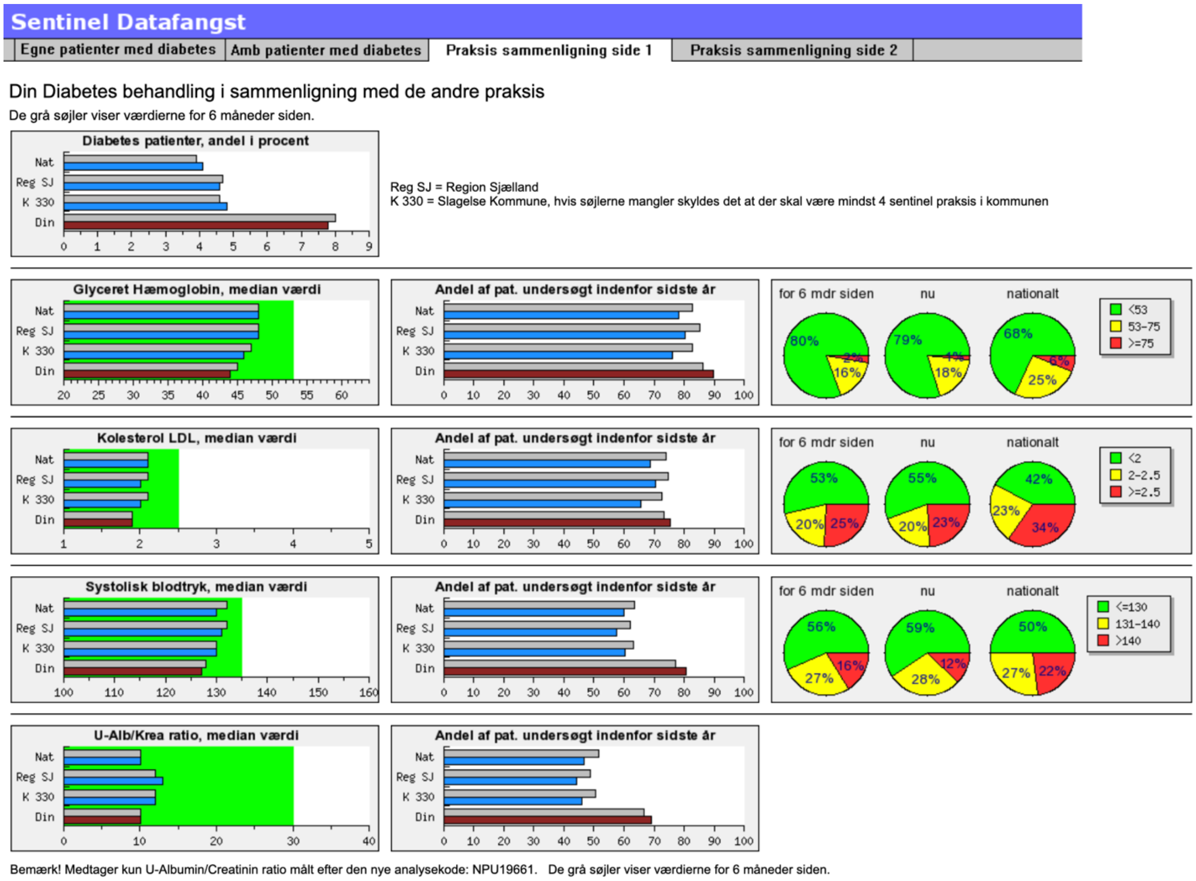
Technically, Sentinel required a local installation of software on the computer of registered users across the clinics in general practice. Once installed, diagnostic codes, prescribed drugs, laboratory results and other types of data were copied to a disease-specific sub-database for patients assigned the designated ICPC-code for diabetes type-1 ('T89') and type-2 ('T90'). Unlike the early pilot, the capture of most data was automated from twelve EHR systems operating in the Danish market, recognising that a manual registration after working hours was an untenable requirement for users in the long run. Users were however required to manually register supplementary data that was not available for reuse in the information landscape of the healthcare system. This "soft data" (Kristiansen, 2013), as Schroll terms it, was to be registered during annual controls via a questionnaire form that would appear once a year for each diagnosed diabetic.[2] Centralised in a database known as the *Dansk AlmenMedicinsk Database*

---

[2] "Soft" data included comorbidity risk factors (e.g. existing complications and lifestyle habits) and procedural recommendations (e.g. whether the patient has undergone a foot examination or received lifestyle counselling within a recommended time frame).

(DAMD hereafter), the data would be cleaned and algorithmically processed through disease-specific “indicator sets” developed and overseen by DAK-E in coordination with the National Indicator Programme. This coordinative collaboration was in part established to support the production of anonymised public audits across the various sectors in Danish healthcare regarding specific diseases. Once processed by DAK-E, users - GPs and nurses - would be able to access data analytics in the form of weekly “quality reports” via *Sundhed.dk,* the national eHealth portal.[3] To ensure readily available decision support, the data would first be stored locally on the user’s computer in a Microsoft SQL database – generated upon the installation Sentinel – from which data would be copied to DAMD. With an overview of the technical landscape the following proceeds to review Sentinel, highlighting how it was envisioned to help users improve the quality of diabetes care.

Where to look



Din Diabetes behandling i sammenligning med de andre praksis

De grå søjler viser værdierne for 6 måneder siden.

Reg SJ = Region Sjælland

K 330 = Slagelse Kommune, hvis søjlerne mangler skyldes det at der skal være mindst 4 sentinel praksis i kommunen

Bemærk! Medtager kun U-Albumin/Creatinin ratio målt efter den nye analysekode: NPU19661. De grå søjler viser værdierne for 6 måneder siden.

---

[3] Indicator sets were eventually developed for Chronic obstructive pulmonary disease (COPD), cardiovascular disease and depression too with corresponding sub-databases (see figure 1).

Figure 2 –Screenshot of ‘population indicators’ in the demo version of Sentinel

Given that Sentinel had been discontinued in general practice at the time of the interview, in March 2018, Schroll was kind enough to walk me through a preserved demo from his personal computer to illustrate the envisioned benefits. The first feature that Schroll highlights is the instant count of the patients diagnosed with diabetes (n=330) and overall patient population (n=4585), along with a metric of the ‘diabetes frequency’ that accounts for the comparative difference in percent (7,2%) (figure 2).

Thirty seconds into the demonstration, Schroll has from that one metric already established that “This is a practice that has really done a lot to find their [diabetes] patients” . In order to convey the underlying logic, I am taken to a page labelled *Practice comparison* that is populated with bar charts with the occasional inclusion of pie charts (see figure 2). The first bar chart is ‘diabetes frequency’ that allows the user to compare the current percentage against colleagues on an aggregate level (municipal, regional and national) and their respective development six months prior. “If you are 25% under the expected,” Schroll narrates, “then you are probably not good enough at finding your diabetes patients”. Exhibited in this quote is a logic of statistical probability that is envisioned to accompany the use of Sentinel incidences of *un*diagnosed diabetics that are made visible through their statistical absence. Users are thus expected to pay attention to what in statistical quality management has been phrased the ‘information in variation’ (Deming, 2000 [1982]: 309) as a means to render visibility to diabetics who are ‘off the radar’ in Sentinel.

Coding undiagnosed diabetics is then a condition to be ‘counted’ by Sentinel and thereby made visible to the user through various indicators. As the review advances, it will become increasingly clear how seeing is not an end in itself. On the contrary, the following demonstrates that “[s]eeing something is the first step to controlling it” (Espeland and Stevens, 2008: 415). This is for instance indicated by the political allocation of funds to introduce a financial incentive for the coding of undiagnosed diabetes in the formative years of Sentinel (PLO and Regional Government, 2006). For those counted in the system, the logic of statistical probability would extend to other ‘population indicators’.[4] The basic idea was to provide the users a general overview of the quality of care and thereby assist with the identification of

---

[4] The 'diabetes frequency' indicator is the only one designed to make *un*diagnosed patient visible through statistical reasoning.The other population indicators are designed to support users apply statistical reasoning to assess the overall health and treatment of *diagnosed* patients.The 'diabetes frequency' is then a condition for the usage of other population indicators that only illuminate diagnosed patients with diabetes.

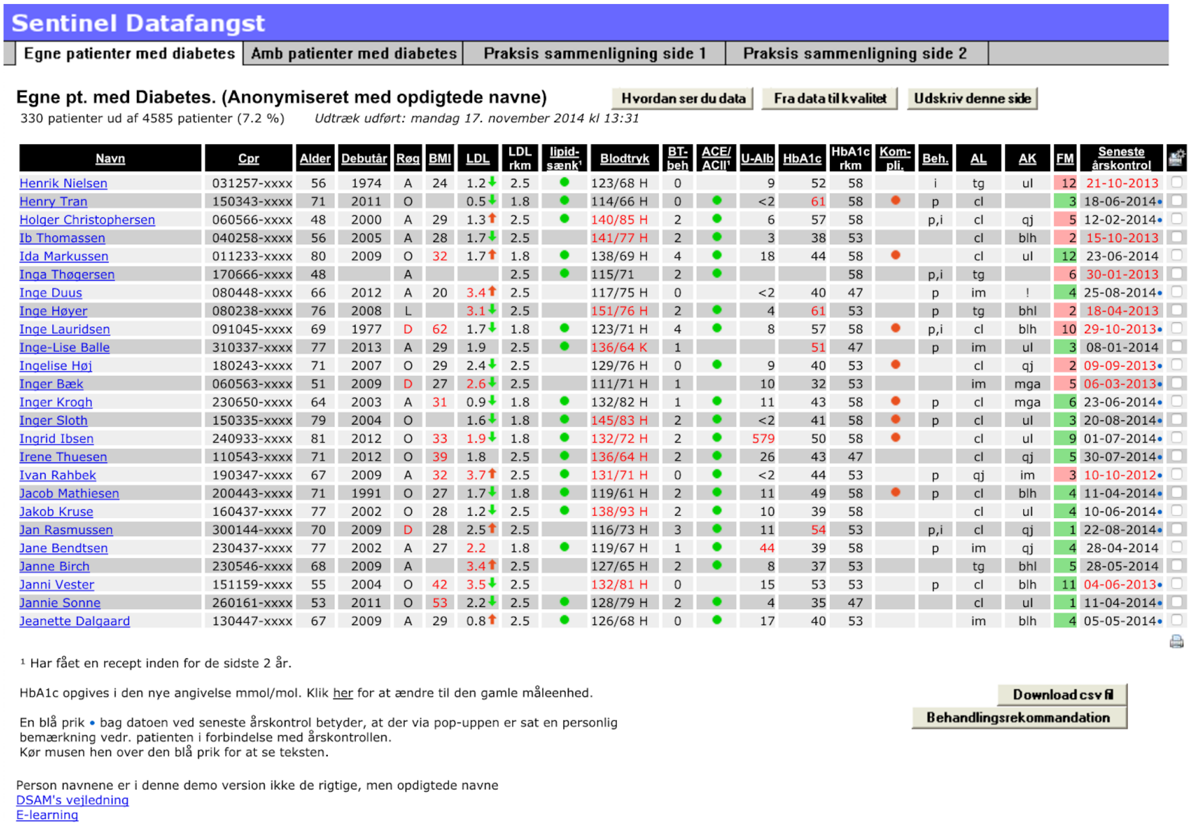
potential areas of improvement. Population indicators include the median value of biomedical measures (bottom four bar charts on the left) that can be compared to colleagues, as before, and benchmarks of national recommendations of acceptable levels (green area in aforementioned bar charts) defined under the stewardship of The Danish College of General Practitioners (DSAM).[5] The biomedical measures are blood sugar cholesterol, blood pressure and albumin/creatinine, which are all known within scientific medicine to increase the risk of comorbidity.[6] Together, these biomedical measures comprise a profile of the diabetes population that users can use to identify comparative anomalies as potential areas of improvement. Other charts would by contrast lend themselves to the examination of comparative anomalies in the delivery of recommended procedures. So, when a recommended medication is *not* prescribed to patients with excessive value outcomes (e.g. anti-hypertensive medication [*ACE/ACII*] for high blood pressure) or controls are *not* carried out within the recommended time frame (e.g. annual check-up or foot examination). Population indicators are then a way for users to understand the profile of their diabetes population and the status of treatment. They provide a basis for self-evaluation to inform *where* users should look in order to improve their "performance levels", as Schroll puts it, which ties into quality management philosophies of continual improvement.

Who to target

---

[5] Whereas PLO focuses on the contractual conditions of general practice, DSAM looks to strengthen the profession on the basis of science.

[6] The threshold for recommended blood sugar levels is for instance defined at HbA_{1c}>= 53 mmol/mol where excessive levels are known to increase the risk of microvascular complications such as blindness, kidney failure, impotence and foot amputation (Wei 2014).



Egne pt. med Diabetes. (Anonymiseret med opdigtede navne)

330 patienter ud af 4585 patienter (7.2 %) *Udtræk udført: mandag 17. november 2014 kl 13:31*

Navn	Cpr	Alder	Debutår	Røg	BMI	LDL	LDL rkm	lipid-sænk[1]	Blodtryk	BT-beh	ACE/ ACII[1]	U-Alb	HbA1c	HbA1c rkm	Kom-pli.	Beh.	AL	AK	FM	Seneste årskontrol	
Henrik Nielsen	031257-xxxx	56	1974	A	24	1.2↓	2.5	●	123/68 H	0		9	52	58		i	tg	ul	12	21-10-2013	☐
Henry Tran	150343-xxxx	71	2011	O		0.5↓	1.8	●	114/66 H	0	●	<2	61	58	●	p	cl		3	18-06-2014•	☐
Holger Christophersen	060566-xxxx	48	2000	A	29	1.3↑	2.5	●	140/85 H	2	●	6	57	58		p,i	cl	qj	5	12-02-2014•	☐
Ib Thomassen	040258-xxxx	56	2005	A	28	1.7↓	2.5		141/77 H	2	●	3	38	53			cl	blh	2	15-10-2013	☐
Ida Markussen	011233-xxxx	80	2009	O	32	1.7↑	1.8	●	138/69 H	4	●	18	44	58	●		cl	ul	12	23-06-2014	☐
Inga Thøgersen	170666-xxxx	48		A			2.5	●	115/71	2	●			58		p,i	tg		6	30-01-2013	☐
Inge Duus	080448-xxxx	66	2012	A	20	3.4↑	2.5		117/75 H	0		<2	40	47		p	im	!	4	25-08-2014•	☐
Inge Høyer	080238-xxxx	76	2008	L		3.1↓	2.5		151/76 H	2	●	4	61	53		p	tg	bhl	2	18-04-2013	☐
Inge Lauridsen	091045-xxxx	69	1977	D	62	1.7↓	1.8	●	123/71 H	4	●	8	57	58	●	p,i	cl	blh	10	29-10-2013•	☐
Inge-Lise Balle	310337-xxxx	77	2013	A	29	1.9	2.5	●	136/64 K	1			51	47		p	im	ul	3	08-01-2014	☐
Ingelise Høj	180243-xxxx	71	2007	O	29	2.4↓	2.5		129/76 H	0	●	9	40	53	●		cl	qj	2	09-09-2013•	☐
Inger Bæk	060563-xxxx	51	2009	D	27	2.6↓	2.5		111/71 H	1		10	32	53			im	mga	5	06-03-2013•	☐
Inger Krogh	230650-xxxx	64	2003	A	31	0.9↓	1.8	●	132/82 H	1	●	11	43	58	●	p	cl	mga	6	23-06-2014•	☐
Inger Sloth	150335-xxxx	79	2004	O		1.6↓	1.8	●	145/83 H	2	●	<2	41	58	●	p	cl	ul	3	20-08-2014•	☐
Ingrid Ibsen	240933-xxxx	81	2012	O	33	1.9↓	1.8	●	132/72 H	2	●	579	50	58	●		cl	ul	9	01-07-2014•	☐
Irene Thuesen	110543-xxxx	71	2012	O	39	1.8	2.5	●	136/64 H	2	●	26	43	47			cl	qj	5	30-07-2014•	☐
Ivan Rahbek	190347-xxxx	67	2009	A	32	3.7↑	2.5	●	131/71 H	0	●	<2	44	53		p	qj	im	3	10-10-2012•	☐
Jacob Mathiesen	200443-xxxx	71	1991	O	27	1.7↓	1.8	●	119/61 H	2	●	11	49	58	●	p	cl	blh	4	11-04-2014•	☐
Jakob Kruse	160437-xxxx	77	2002	O	28	1.2↓	2.5	●	138/93 H	2	●	10	39	58			cl	ul	4	10-06-2014•	☐
Jan Rasmussen	300144-xxxx	70	2009	D	28	2.5↑	2.5		116/73 H	3	●	11	54	53		p,i	cl	qj	1	22-08-2014•	☐
Jane Bendtsen	230437-xxxx	77	2002	A	27	2.2	1.8	●	119/67 H	1	●	44	39	58		p	im	qj	4	28-04-2014	☐
Janne Birch	230546-xxxx	68	2009	A		3.4↑	2.5		127/65 H	2	●	8	37	53			tg	bhl	5	28-05-2014	☐
Janni Vester	151159-xxxx	55	2004	O	42	3.5↓	2.5		132/81 H	0		15	53	53		p	cl	blh	11	04-06-2013•	☐
Jannie Sonne	260161-xxxx	53	2011	O	53	2.2↓	2.5	●	128/79 H	2	●	4	35	47			cl	ul	1	11-04-2014•	☐
Jeanette Dalgaard	130447-xxxx	67	2009	A	29	0.8↑	2.5	●	126/68 H	0	●	17	40	53			im	blh	4	05-05-2014•	☐

[1] Har fået en recept inden for de sidste 2 år.

HbA1c opgives i den nye angivelse mmol/mol. Klik her for at ændre til den gamle måleenhed.

En blå prik • bag datoen ved seneste årskontrol betyder, at der via pop-uppen er sat en personlig
bemærkning vedr. patienten i forbindelse med årskontrollen.
Kør musen hen over den blå prik for at se teksten.

Person navnene er i denne demo version ikke de rigtige, men opdigtede navne
DSAM's vejledning
E-learning

Figure 3 - Screenshot of 'individual indicators' in the demo version of Sentinel with fictional patients

By clicking on the page *Own patients with diabetes* Schroll shows me how insights from 'population indicators' can be examined on a patient level to help users prioritise who to target (figure 3). Patients within the diabetes population are listed vertically (rows) by name and profiled horizontally (columns) through 21 parameters including biomedical outcomes, prescriptions and latest annual check-ups to provide 'patient indicators' of the quality of care. Unlike paper records in filing cabinets, Schroll showcases the affordances of digital technology by resorting to the listing of patients according to various parameters by clicking the column header. While technologically rudimentary in the 2020s, this feature was nevertheless crucial to the envisioned operation of singling out patients contributing to identified anomalies on a population level (e.g. patients with excessive blood sugar levels or those overdue on their annual check-up). In order to help users see who to follow up on, patient indicators exceeding national recommendations were automatically highlighted in red and personalised to the profile in question when relevant.[7] Together, the selection of patient indicators and highlighting

---

[7] The threshold for highlighting 'HbA1c' (blood sugar) red for instance varies according to the presence of ICPC coded complications – kidney disease (K89, K90, K91, K92) and cardiovascular disease (K74, K75,

techniques form an algorithmic assemblage programmed to cast light on patients with greatest need of clinical attention. The next quote shows how this instance of patient profiling was intended to “prompt” users into a proactive mode of corrective disease regulation through intervention:

> *Part of what we worked with in DAK-E [was] to figure out if we could get a hold on diabetes patients and prompt the doctors by providing reports saying ‘you have 120 diabetes patients in your practice where 80 of them are well-regulated. But there are 30 who you haven’t seen the past year and it is these and these patients.*

Embedded in these features of Sentinel is a specific and normative idea of what counts as ‘good’ diabetes care. While the literature on what constitutes ‘good care’ is extensive, encompassing contestations over patient choice (Mol 2008) practitioner documentation (Bødker et al 2019) and the role data itself in shaping accountability (Cruz 2022), Sentinel originates in indicators for quality. Quality indicators, on a population and patient level, are designed to highlight factors that according to national guidelines are known to constitute (e.g. excessive biomedical values) and mitigate (e.g. recommended examinations and treatment) the risk of comorbidity. In this respect, Sentinel is a quality assurance technology to strengthen compliance with national guidelines in general practice based on quality indicators that ultimately promote a *preventive* logic of care to regulate the risk of comorbidity.

To this end, the user is imagined to be capable and willing to evaluate and improve their own performance levels by statistically scrutinising the information in variation on a population and patient level. Comparisons with colleagues are in this process intended to yield insights on the practical feasibility of national benchmarks in a given municipality to for instance factor in socioeconomic differences. The normative aim of Sentinel, as a social and political project, is thus to illuminate patients at risk of comorbidity so resources can be allocated accordingly. When interviewed in *Dagens Medicin,* Schroll described the implications of his design:

> This will have a massive significance. There will be fewer consequential complications, and it will benefit everyone - also society. Including those with high blood pressure who do not have chronic diseases but are at risk of developing one. You can […] get hold of them before it becomes critical. The doctor can look at a risk score and see which patients to act on, and ultimately save lives. Just press a button and you get the risk factors for the patients. These are people who are not sick yet but have indications that something may be going in the wrong direction later. (Schroll, cited in Schultz, 2013)

K76, K77), as well as number of years diagnosed with diabetes (https://demo.dak-it.dk/diabetes/behandlingsrekommandation.pdf).

How to intervene

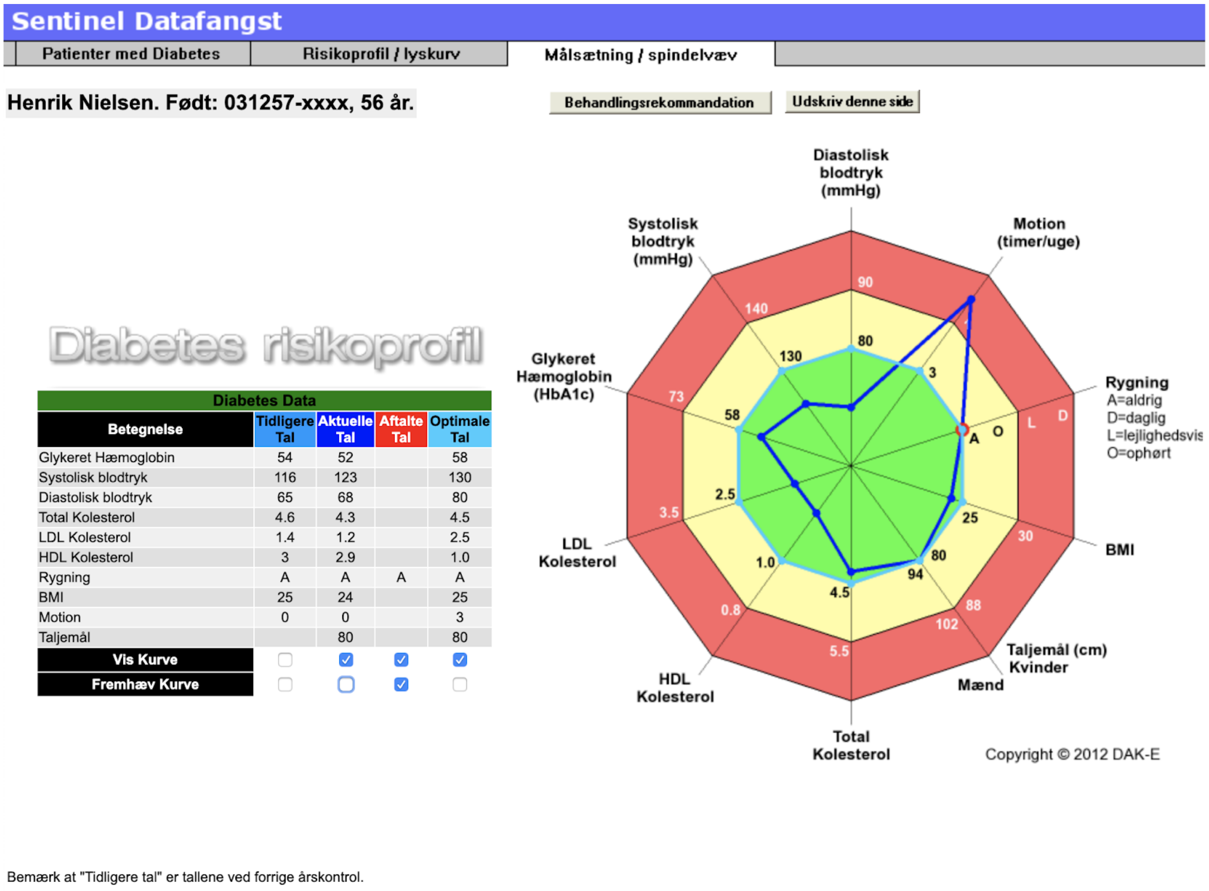
The last page that Schroll walks me through is called *Target setting / Spiderweb* that displays a “diabetes risk profile” intended to facilitate patient interventions in the promotion of healthy lifestyle choices.[8] Clicking on a hyperlinked name in the spreadsheet Schroll provides an example based on Henrik Nielsen, a patient of 56 years of age seemingly not too keen on the idea of exercise:

> There’s the patient [list]. If we choose Henrik Nielsen… How was your blood pressure level half a year ago and now? You have numbers available with graphically depicted values. There's the "spiderweb". He doesn't move this guy, his exercise level is zero (chuckle), but otherwise he is well-positioned. This is a patient you can talk with […]

Henrik’s risk profile consists of ten measurements (pertaining to blood sugar, blood pressure, cholesterol, smoking, waist size, body-mass-index, exercise) represented in a table but also a radar chart referred to as a “spider web” (see figure 3). The most recent measures are in the radar chart illustrated by the interconnected plots in dark blue and the “optimal” value is delineated by light blue plots, as reflected in the column headings too.

---

[8] The Diabetes Risk Profile was made available to patients via the National eHealth portal (Sundhed.dk) in August 2010 (DAK-E 2011, 11).



Henrik Nielsen. Født: 031257-xxxx, 56 år.

Diabetes Data

Betegnelse	Tidligere Tal	Aktuelle Tal	Aftalte Tal	Optimale Tal
Glykeret Hæmoglobin	54	52		58
Systolisk blodtryk	116	123		130
Diastolisk blodtryk	65	68		80
Total Kolesterol	4.6	4.3		4.5
LDL Kolesterol	1.4	1.2		2.5
HDL Kolesterol	3	2.9		1.0
Rygning	A	A	A	A
BMI	25	24		25
Motion	0	0		3
Taljemål		80		80
Vis Kurve	☐	☑	☑	☑
Fremhæv Kurve	☐	☐	☑	☐

Bemærk at "Tidligere tal" er tallene ved forrige årskontrol.

Figure 4 – Screenshot of "Diabetes risk profile" displayed as a table and radar chart in the demo version of Sentinel

In the radar chart, the current measurement of 'exercise' [*motion*] - defined as number of hours per week - almost intuitively demands attention of the user since it is the only outlier plotted in the high-risk zone signalled by the colour red. Devoid of stratification and visual cues, the table representation by contrast fails to communicate the severity of risk that presumes a degree of medical expertise that cannot be expected of laypersons.[9] Unlike the two previous pages, Schroll narrates how the user would review the risk profile in with patients where it is envisioned to serve as an instrument for setting "shared targets" between controls: "And then [the physician] would say 'I can see that the blood pressure is my responsibility to regulate, but the thing about exercise… Could something be done there?' [...]Any suggestions?". In this way,

---

[9] In fact, I am in the same interview informed that a bar chart was introduced as an alternative display option since the radar charts proved challenging for some of the marginalised patients previously noted as the primary beneficiaries.

Schroll continued, the patients could reply: “No, the thing about running half an hour each day, I won’t do that. But I can take a walk”. The objective was then to establish *feasible* targets through dialogue (walking) to make a break with the “blame culture” in general practice where patients are held accountable to unreasonable *ideals* (running) by their physician. By setting targets through patient involvement Schroll asserted that patients would be more inclined to attend annual controls compared to the unmotivating scenario of being “scolded” for their imperfections – again.

The above scenario then illustrates another instance of how Sentinel was envisioned to help GPs keep a hold on their patients. In setting shared targets, the use case scenario also shows how the radar chart would help the physician draw up boundaries of responsibilities. In the quote, Henrik is confronted with the graphs and numbers to take more responsibility for his level of exercise but only after the physician *performs* the division of labour by establishing that she can “see” the medical regulation of blood pressure within her domain of responsibility.[10] Despite its fictional footing, the use case demonstrates how the exercise of medical authority also draws on graphical representations to promote healthy lifestyle changes where it in contrast to previous examples is the physician, and not only algorithm, who does the work of highlighting. In practical terms, Henrik’s targeted number of hours of exercise per week would, following the protocol, be entered in the corresponding field in the questionnaire prompted at the beginning of the annual check-up. Once recorded, the target would feature as a red plot on the radar chart accessible to both doctor and patient via Sundhed.dk, thereby providing a reference of accountability for the next annual control.

Exhibited in this scenario is how the manual capture of “soft” data is used to rationalise medical intervention *beyond* the body (Lucivero and Prainsack, 2015); habits and corresponding targets of patients are datafied, tracked over time and stratified through models or risk to inculcate a preventive logic of self-care (e.g. Rose, 1990; Schüll, 2016). This promotion of healthy lifestyle changes resonates with discourse in Denmark for the first decade of the twenty-first century, as demonstrated in a comparative review of the public health policies in Scandinavia this (Valgaarda, 2007). With its framing of individual behaviour as both cause and solution of disease burdens like diabetes, the discourse in Denmark was considered liberal in comparison to Norway and Sweden where differentiating factors included social relations, living conditions

---

[10] This is also reflected in the spatial disposition of the radar chart with the grouping of exercise, smoking, BMI and waste size furthest to the right.

and values. This differentiation of the “Scandinavian model” of public health helps to elucidate the politics involved (and reinforced by) with Sentinel with its promotion of health styles. Among the concerns that critical scholarship of social inquiry highlights is the moral encroachment of medical intervention into everyday life; On what grounds are such schemes justified? What impact they have on pathologisation (an exacerbated prevalence of ‘worried well’ patient) and overtreatment as notions of health, illness and disease are redefined on the basis of risk to prevent what may never be expressed (Webster, 2002)? To what extent should individuals be responsibilised for the treatment of conditions that have in part been predisposed in genetic, social and economic terms? Such questions are deeply political and yet easily evaded and de-politicsed when reduced to the technoscientific realm of numerical and visual units of representation (e.g. Porter, 1996; Lupton, 2018).

The rationale for the national introduction of Sentinel was formulated in 2006 by the regional government in the funding process. A key argument in the rationale was based on the assumption that the number of *diagnosed* diabetics in Denmark, estimated to be between 100.000 and 150.000, was “considerably” short of the number of actual diabetics (Statens Serum Institut, 2014a: 273). This was considered to constitute a social problem since the risk of comorbidity – regarding the circulatory, nervous, urinary and eye systems – for diabetics is medically known to become exacerbated when unregulated. Diabetics were additionally noted to have a “significant” excess mortality compared to non-diabetics (Statens Serum Institut, 2014a). The perceived disparity between the number of diagnosed and actual diabetics was however also articulated as an economic problem. Accounting for 3 percent of patient encounters in general practice – the third most frequent reason of all encounters – diabetes was considered a “resource heavy” disease area (ibid.). If the aforementioned disparity was true, then the undiagnosed diabetics would in terms of probability have a greater risk of comorbidity and by extension also the costly treatment of specialised or hospital care. The socioeconomic objective for Sentinel was then to improve the diagnostic coding and treatment of diabetes in general practice to reduce the risk of comorbidity through disease regulation - and thereby the decline of quality of life and increase of costly treatment.

The changed practice of diagnostic coding and diabetes care

The decision to implement Sentinel on a national scale was made in the collective negotiations between the PLO and the regional government. In the first collective agreement, Sentinel was between 2007 and 2009 introduced as a decision support tool that GPs could opt to use on a voluntary basis. While the initial focus was on diabetes, the scope of feedback reports was gradually expanded to other areas of chronic disease. With the second collective agreement of 2010 came the decision to make the use of Sentinel mandatory for seven disease areas by the end of 2012.

Whereas the adoption of Sentinel struggles to gain momentum in the period of the first collective agreement, the introduced mandate of the second agreement effectively made Sentinel a routine tool in the practice of chronic care (See Burnett, forthcoming).[11] Along with this mandate came a duty of registration that for the first time in Denmark made the use of the diagnostic coding obligatory for general practice, alongside the entry of questionnaire data that was also a requirement of the feedback reports. While disputes may occur regarding how much should be coded, by 2012 few questioned the ICPC classification system as a national standard. With this status of acceptance, the ICPC has today become somewhat black-boxed, as it so often happens with classification systems with the passing of time (Bowker and Star, 1999). Revisiting the history of the ICPC can however be a generative exercise to remember that classification systems provide schemas to render the world visible in certain ways.

The classification system for diseases used in general practice cannot simply be transferred to the hospital sector because they operate under different conditions. Contrary to the general practice, the hospital sector has for decades operated with a duty of documentation that in recent years has practised in accordance with the *International Classification of Diseases* (ICD) now into its 11th edition. One difference between the ICD and ICPC is that the former presupposes a clarified pathological picture, which is therefore considered unfit for general practice since it is not unusual that health problems never achieve full diagnostic clarity (Rosendal & Falkø 2009). What can be elucidated in general practice is by contrast a picture of the early stages of and pathways to illness, that is, how health problems evolve from the very first encounter.

This way of thinking gained prominence in the 1970s that is reflected in the ICPC. Contrary to most other classification systems, the ICPC sought to depart from the esoteric language of medicine by providing a terminology for GPs to systematically code the ‘reason for encounter’ rather than disease diagnosis alone (Lamberts & Wood, 2002). ICPC then

---

[11] That is, until it became the centre of a public controversy that resulted in infrastructural collapse to the dismay of health professionals (see Burnett, forthcoming).

represented a particular nosological school of thought that encouraged GPs to “objectively” record - and thereby attend to - the symptoms, complaints and fears of patients through dialogue *before* a “subjective” attempt to diagnose the cause (Lamberts & Wood, 2002). In this sense, the promotion of ICPC can be viewed as an attempt to re-educate the clinical attention as to bring the patient closer into view.

The ICPC can – in terms of what it deems necessary to code and prioritisation of dialogue – be viewed as complementary to the movement towards ‘patient-centredness' conceptualised earlier in the same century; a move in advocacy of a holistic gaze to look beyond the disease and recognise the patient as a ‘whole person’ (Royal College of General Practitioners, 1972). The claim that ICPC presented a different school of thought is in part indicated by the resistance it encountered upon its early reception within the community of general practice. In one study, the case of this resistance is reported to stem from the proposed embrace of a dialogue friendly language and objective documentation was perceived to undermine their professional judgement and expertise (Lambert and Wood, 2002). Although ICPC is today established as the national standard for diagnostic coding in Denmark, it is as a practice a relatively recent tradition within general practice.

Before the advent of EHR systems, as described above, Schroll recalls how "18/3/92, tonsillitis, penicillin" would typify the only record of a patient encounter in Denmark. Format and spelling might vary. With the early adoption of EHR systems in 1993, 1 percent of GPs coded their patients. By 2003, the adoption of ICPC had grown to 16 percent a decade later, followed by an increase to 45 percent in 2005 when Sentinel was initially piloted (Vedsted and Schroll 2008, p. 177). The study from which this development is described observes how the ICPC was adopted with variations in terms of its regional coverage (ranging from 33% til 93% in different regions). A ‘large’ variation was additionally noted in the actual practice of coding, as some would only code the organ in question without for instance specifying the symptom (the organ is indicated by the first letter of the diagnostic code where the symptom, process and disease are given a two digit number, e.g. “T” for ‘Endocrine/Metabolic and Nutritional’ and “07” for “weight gain”).[12] Too large a variation in terms of geographical coverage and the practical use of ICPC can respectively challenge the statistical representation and validity of the recorded data for certain purposes, such as clinical research. Where the former issue was addressed with the aforementioned contractual intervention, the latter was approached with a concerted effort to

---

[12] A second version of the Danish ICPC translation was implementation 2008, ICPC-2-DK: https://web.archive.org/web/20160330063758/http:/www.dak-e.dk/flx/en/general-practice/icpc/

make the coding practices more uniform. With Schroll involved, a national governance of training was established to educate users on the correct use of ICPC codes through courses, workshops and educational materials.

Changing Practice of Care

Sentinel was considered to provide useful mechanisms for identifying patients exceeding the recommended timeframe for routine control, for instance through the provision of automated follow-up notifications but also the option to manually sort the population overview by 'latest annual control'. Individual quality indicators in the population overview were - largely based on satisfactory and dissatisfactory visual markers - additionally found helpful by enhancing the awareness of poorly regulated patients. The performative effect of data was highlighted by one of Lippert et al's GP study subjects as articulating being "convinced" by access to the "actual data" (2014:6). The use of aggregated feedback for benchmarking and comparisons were by contrast observed to be limited. The exception was however the average prevalence of diabetes that (along with fiscal incentives) was seen to improve the awareness and coding of undiagnosed type-2 patients (Lippert et al. 2014: 5). In this sense, Sentinel can be viewed as an administratively useful tool to systematise work practices in a way that promotes the (re)allocation of clinical attention and resources to vulnerable patients, thus enkindling the representational politics of who receives care as previously envisioned by Schroll.

While Sentinel was generally received in commendable terms, its users also expressed concern regarding its datafication of everyday work practices. The stratified view of patients was unanimously considered inadequate as a basis for evaluating the "actual quality of care" and "the need for treatment in individual cases" (Lippert et al., 2014: 7). What would be evaluated as satisfactory treatment from a distance were for instance observed to stand in bleak contrast to enskilled judgements made in situated patient encounters. One concern expressed by Lippert's study subjects was that seeing the patients through numbers could derail the attention from what the author/s paraphrases as "'real' patients and 'real' problems" expressed to reconfigure professional priorities (Lippert et al 2017: 11). The indicators provided by Sentinel promote a particular view valorising biomedical view of the patients that necessarily is reductionist by nature and therefore inadequate for seeing the patient as a 'whole' (Lippert et al. 2017:6). This was echoed by another subject wary of an observed tendency to "forget about the patients" (Lippert et al. 2017: 11) because numbers and the questionnaire (used to produce data and strengthen guideline compliance) would drown out attention for “more important” questions such as “how the patient is doing” (Lippert 2017: 12). Here we can observe a tension between temporally distinct logics of care where Sentinel mostly operates according to a principle of risk mitigation to prevent *future* possible complications engaging negotiation with a principle of holistic care to tend to immediate needs known in the *present*.

Extending Sight

What began as a project to improve the visibility and manageability of diabetes in general practice ended in fierced dispute as the DAMD database became a coveted source of extended visibility for other stakeholders (Langhoff et al., 2016 and Wadmann; Hoeyer, 2018; Burnett, forthcoming; Burnett and Douglas-Jones, forthcoming). In Schroll’s mind, there is little doubt about when the quality development project took a turn for the worse. In our interview, Schroll tells me that “Something happened in 2010”, which according to his view was “the poison in the entire project”. In the interview, he traces this “poison” to two outcomes of the collective agreement the same year that reconfigured infrastructural arrangements, extending the visibility of Sentinel for patients and health professionals, and also public administrators.

Extended visibility for patients and health professionals

The first outcome was the agreement to ensure ‘that relevant data […] can be seen by the remaining healthcare system, including the duty doctor’ (Danish Regions and PLO, 2009: 151). With the exception of unstructured journal notes, everything in the DAMD was by 2011 made accessible to health professionals via Sundhed.dk to ensure the accessibility of relevant information when needed to strengthen patient safety (The Ministry of Health, 2010). It was additionally decided to extend the previously mentioned possibility of self-access to patient records to all patients by 2013. The official objective of this latter decision was the promotion of patient empowerment.

Although Schroll describes the ambition as “sensible” and potentially life saving, his point of contestation is the way it was implemented. Unstructured data, such as the contents of case notes, had already been disregarded from the initial capture by Sentinel to protect the privacy of the patient and preserve a digital working space for the general practitioner. According to Schroll, social diagnoses (‘Z’-diagnoses) should however have been omitted too on grounds that they contained “highly personal information” given under the assumption of confidentiality. Problems regarding economy/poverty (‘Z01’), partner’s behaviour (‘Z13’) and compliance/disease comprehension (‘Z11’) are examples of ICPC diagnoses documented in Danish general practice to assess health issues in a “social medical light”. Social diagnoses can also be important in the line of specialist and hospital care, but Schroll sees it as the task of the GP to include the factors *specific* to the case at hand in the referral *when* relevant. The GP is in this conception not just a gatekeeper for patients to access the health system, but also for the healthcare system to access sensitive patient data in order to preserve patient confidentiality and thereby trust as prerequisites for good care.

According to Schroll, the implementation of self-access to patient records (containing prospective information on for instance treatment and medication) precipitated a gradual disintegration of the room for confidentiality. Transposing himself into the shoes of a patient, Schroll admits how he would “get damn furious” if he found confidential information of a sensitive nature about himself, including his spouse, registered on Sundhed.dk:

> If a GP had talked with a patient about some very intimate potency problem or infidelity in a marriage and he had coded it as ‘Z’, that is social factor, and it could be retrieved when the patient enters to view [his or her] own data on Sundhed.dk. Then I would [as a patient] get damn furious. Then I would say “I thought I told you this in full confidentiality so it should never go beyond the two of us”.

Contrary to the official objective of its introduction, the quote provides an use scenario of how patients might from a privacy perspective experience self-access as *disempowering*. The visibility of unwanted registrations could in principle additionally be encountered in dialogue with health professionals while unwanted journal viewers were made visible via an access log on the portal. Articulated in this example is a tension between privacy and patients safety. On the one hand, omitting visibility to social diagnoses could well mitigate the risk of unwanted information disclosure. On the other hand, the same omission could potentially compromise patient safety when social diagnoses are of clinical relevance. In practice, GPs cannot always be readily available to act as a gatekeeper in the curation of relevant information, for instance in acute hospitalisation. The omission of social diagnoses is however not a bullet proof measure to eradicate the risk of an experienced privacy invasion, as the experience is subjective and contingent to cultural norms. This challenges an essentialist conception of ‘sensitive’ data, as it becomes difficult to predict the content of this category certainty.

Extending visibility for public administrators

As indicated, the DAMD was of considerable interest to the public administration for its capacity to render visibility to the activities of general practice. Advancing the frontiers of state legibility (Scott, 1998) intro general practice has been articulated as a means to base administrative planning and public accountability on the data-driven ideal of knowledge as opposed to intuition (Burnett, forthcoming, see also Langhoff et al. 2016; Wadmann and Hoeyer, 2018). The metaphorical linkage between sight and knowledge was of course also present in the thinking of philosophers in the age of enlightenment. Thoughts "that more knowledge could cause problems, that light might prove another tyranny", Haridimos Tsoukas notes, were however not something that the philosophers that age were prepared to entertain (Tsoukas, 1997: 839). Whether or not the stakeholders of the DAMD were prepared, the question of what should be visible and known to whom caused considerable problems and power struggles enkindled particularly by a state-led push for strengthened public accountability. On the note of 'accountability', social anthropologist Marylin Strathern (2000: 310) observes that the demand for transparency suggests an absence of trust. Probable cause of such an absence was formulated in the advisory report *Chronic disease* from 2005, highlighting an 'extraordinarily high' level of reimbursement claims for services in general practice (Kronisk sygdom, 2005: 51).[13] From its inception to its demise, the DAMD was in parallel to its clinical and research utility enrolled in a series of attempts to reform the honorarium system in general practice.

The first reform came with the collective agreement of 2006 where GPs were encouraged to abandon the fee-per-service model of remuneration in favour of a fee-per-capitation (PLO and DR, 2006: 2). The new model was piloted for diabetes with a disease-specific 'pathway fee' that GPs could collect for each ICPC coded diabetic insofar an annual control was conducted via Sentinel (e.g. registration of questionnaire, self-care conversations) and a minimum of one intermediate control. Unlike the pay-per-service model, the fixed structure of the pathway fee was thus theorised to create an economic incentive for a reduction in the number of patient encounters in general practice, which was the long-term objective of preventive disease-regulation (Kronisk sygdom, 2005: 59).[14] The pathway fee however failed to catch on and was consequently decoupled from Sentinel in the subsequent collective agreement. The

---

[13] Remuneration for services accounted for 70% of the income in general practice.

[14] Following a review, the OECD published an assessment of GP provision in Denmark, noting that 'the fact that most GP income derives from fee-for-service may not be best suited to the provision of holistic, integrated care' (OECD 2013:22).

reason for this outcome was according to Schroll because the pathway fee was ‘economically unattractive’, since most GPs would in his view struggle to regulate patients sufficiently to limit patient encounters to 2-3 a year (Redaktionen, 2009).

The failure of one reform however made way for another: a pay-per-performance model of reimbursement. Early intimations of this second model were formulated in the collective agreement of 2010, which proposes a continued dialogue about making indicator data from the DAMD accessible to public authorities on individual general practices (PLO and Danish Regions, 2006: 103). Hitherto, public authorities were limited to an aggregated view of such data – based on a minimum of four general practices – for purposes that include public reporting (DAK-E, 2007: 20). Commenting on the ‘crisis agreement’, the head of PLO reassured his members that ‘data [from the DAMD] shall not be used for control, but professional development and learning’ (Boysen, 2011). Although the collective agreement of 2010 did not result in the implementation of the pay-per-performance model, it did mandate a national adoption of Sentinel with an obligatory transfer of ICPC-related data to the DAMD. As noted in an excellent analysis from Wadmann and Hoeyer (2018), the introduction of self-access to patient records held other interest than patient empowerment. One ulterior motivation behind the project of self-access is alluded to in the working title it received: “the snitch” (Wadmann and Hoeyer, 2018: 7). As one informant explained, it was internally “envisioned that citizens would login and check whether doctors were cheating and inform [the authorities], if the doctors had billed services which patients hadn’t received (Wadmann and Hoeyer, 2018: 7). At play is then an interesting inversion of subject positions where it is the patient who monitors the GP to potentially flag irregular billing activities.

By 2013, the subsequent round of negotiations had escalated into an outright conflict; PLO threatened to terminate their contract with the state, and the state threatened to intervene by law. On 28 May 2013, the Minister of Health tabled a bill mandating GPs ‘to provide information regarding the company to regional councils for planning, quality assurance and control of paid subsidies and remunerations’ (The Danish Parliament, 2013a). Yet again, the DAMD was on the verge of being repurposed to assign the subject position of the ‘observed’ to individuals GPs for purposes of external control – but this time from above.

In a professional journal, prominent figures involved in the development of the DAMD and DAK-E formulated a response on why *The individual GP must not be identified* (Redaktionen, 2013). Sanctioning GPs for having ‘poor data’, the authors argued, would make it unattractive to have wayward patients affiliated to a practice and thus deter new practice in socially burdened areas. Based on foreign experiences, a shift towards a pay-per-performance model of remuneration was furthermore argued to introduce a perverse incentive structure that would tempt GPs to ‘embellish the numbers’ and thus compromise the reusability of the data material for quality development and research. As an endnote, the authors cautioned that ‘going after the 5%, who perform the worst, would with strong probability get all GPs to protect themselves against reprisals’. The director of PLO subsequently reiterated the aforementioned threat that his members would be ‘inclined to protect [themselves] tactically by registering wrong and skewed data to obtain as good data as possible’ (Nielsen, 2013). A vocal figure among a growing antagonistic formation of GPs cautioned that a pay-per-performance model would inadvertently reward a medical quick-fix when long-term lifestyle counselling would be the cheaper, convenient and sustainable alternative (Jeppesen, 2013). Regardless, the bill was passed 27 June 2013 and entered effect two months later without the materialisation of PLO’s threat of state independence (The Danish Parliament, 2013b). The victory for state administrators would however be short-lived as the DAMD became the centre of a public controversy 15 months later that culminated with a parliamentary decision to destroy the database (see Burnett and Douglas-Jones, forthcoming).

The Politics of Illumination

At its origin, Sentinel was envisioned to “illuminate” the path to ‘good’ diabetes care for its users in general practice, bringing to light both individual patients in need of follow up, and allowing GPs to view their patients relative to national averages. Yet as researchers have argued for decades, there is ‘nothing innocent about making the invisible visible’ (Strathern 2000: 309). In her work, Strathern studied the operation of ‘audit, quality assurance and accountability’ and became concerned with situations where ‘visibility as a conduit for knowledge is elided with visibility as an instrument for control’ (2000:309 The direct connection between illumination and greater knowledge is clear in the Sentinel case, particularly in Schroll’s early description of it as more of a lighthouse, his hope being that the beam would “illuminate and show which way you should go, so you avoid running into problems”. However, while the beam of the lighthouse generated more knowledge, in the process it also created new challenges. As the beam of light extended, the politics to making visible intensified. Following from the introduction of Sentinel as a data capture tool for quality improvement, wherein the promise of data became less about seeing with data and more about what that data was seen to be and say about those who had generated it.

In documenting how Sentinel guided users in *where* to look, it was a tool for performance, a form of self-diagnosis for GPs, to be used towards the ends of improved patient care. Through the use of population indicators through which comparisons could be drawn, GPs could look at their own "performance levels". In her analysis of the role of EHRs in the United States, Cruz argues that 'data-centered accounts may … be used to outline a social reality that has yet to exist', (2022: 7) with a desire to 'leverage data to recognize more than the present social reality is possible (2022:8). Using new diagnostic coding capacities as a base, Schroll sought precisely to bring a new reality – one of greater overview and more personal follow-up – into being. This comparison, from cases many years apart, allows us to draw out a more general point about the collection and centralization of data within healthcare, namely that data-intensification carries a forward momentum which reaches into the everyday promises of what kind of social realities its use may bring forth.

However, at the same time, the management of a 'resource heavy' disease area such as diabetes from the perspective of a public health system also led to the transformation of that same data into a form of measurement of GPs by administrators. The elision described by Strathern – from 'a conduit for knowledge' to 'an instrument for control' (2000:309) is clear. From a pathway fee to a pay-per-performance model within general practice, there was clear concern that data would be used for control, requiring the head of the Organisation of General Practitioners (PLO) to explicitly state it would be used rather for 'professional development and learning' (Boysen 2011). More than this, as Wadmann and Hoeyer describe with 'the snitch' (2018:7), with increased patient access to data, monitoring for GPs moved the purpose of making data visible from self-assessment to reporting on GPs, reversing longstanding relations of power.

Finally, as we discussed *how* GPs were invited to look at their patients through data, different ways that professional vision was re-shaped by Sentinel. Well described by GP focused studies of Sentinel (Lippert et al. 2014, 2017), while software and visualisations offer considerable promise, it is also necessary to acknowledge and work with their limitations (Sofoulis, 2011, see also Winthereik and Jensen 2017). The good promoted by Sentinel is one of rationalised prevention that adopts a strict focus of disease-specific risk factors on the basis of scientific medicine. While this way of seeing was in many respects successfully promoted by Sentinel, the beacon of good diabetes could also lead users astray from the holistic gaze that draws on medical humanism to recognise the patient as a 'whole' (e.g. by including social, economic, cultural and psychological dimensions of the encountered health problem).This focus in turn promotes the institutionalisation of a preventive logic of care where proactive disease regulation and healthy lifestyle changes is to be sought by self-governing subjects as means of risk mitigation against comorbidity that raise political questions of moral, social and ethical importance. [7] In short, the analysis shows empirically why it would be wrong to assume the uptake of the data gaze in new professional spaces as a value-neutral enterprise, because it is always shaped by social logics that are political in the sense that they promote certain ways of seeing, being and intervening in the world. No patient can be wholly captured in a risk profile of ten parameters or a spreadsheet with 21 columns. As one of Lippert's informants put it, "[d]ata capture comes with the risk that it may make you concentrate on measurable aspects and forget about everything else" (Lippert, 2014: 7). Among the findings that vacated the space of "everything else" was aspects of work "cannot be measured in the same way as long-term blood glucose levels", such as lifestyle, but also the notion that "[q]uality is just as much about talking properly with the patients" (Lippert, 2014: 7). What these remarks convey is a tension between competing definitions of 'good' diabetes care (and also what constitutes 'relevant' data) that reflect distinct schools of thoughts in medicine. In Cruz's framework, data 'simultaneously *capture, obscure* and *envision* the social'. The concerns expressed by GPs in Lippert's study clearly identify spaces where data *obscures,* yet it may be the same data that 'in one context demonstrate good care, and in another obscure it' (Cruz 2020, cited in Cruz 2022: 9). This simultaneity returns us to professional vision, in that 'data representations create the very conditions for people to contextualize them: human actors recognize the realities represented as well as those are not' (Cruz 2020: 9).

Conclusion

Quality measures lend themselves easily to preventive measures in healthcare. Within diabetes and other chronic diseases, keeping track can be significantly assisted by technologies to which some of this work can be delegated, with the hope that disease progression may be slowed or halted. Sentinel has captured the attention of academics, for the changes its capacities for illumination brought to general practice, far beyond its initial envisaged outcomes. This article has sought to extend the scholarship on Sentinel by attending to its originating metaphor, following its paths of illumination through different professional contexts. Across studies of the incorporation of ever more sophisticated electronic health patient record (EHR) systems, the promise of what data can make visible remains enticing. As Cruz has recently demonstrated in her work, ‘data as visibility’ is very common amongst those who seek to build dashboards, with one programmer reporting that ‘[b]eing data-driven is all about *seeing what’s happening…..but it was impossible to look into our system’* (Programmer, cited in Cruz 2022: 5).

Data-intensification has the power to reshape professional vision in the context of diabetes care in Danish general practice. This question was examined by scrutinising how Sentinel was envisioned to help users (GPs) see where to look, who to target and how to intervene, which served the base for unpacking how embedded logics promoted a particular version of ‘good’ diabetes care. Looking into systems is a task that ethnographers and scholars of technology, studying increasing data-intensification in healthcare, can do alongside programmers, practitioners and professionals, attending to the multiple roles data is given, and takes. We see the intersection of sentinel technologies, preventive medicine and digitalization as a generative site for studies of shifting power relations, the contestation of professional values, and ‘socially organized ways of seeing’ (Goodwin 1994: 606). As scholars considering new ‘terrains’ for sentinel devices, Blanchon and colleagues remind us that ‘[o]ne of the most difficult questions is to understand whether the success of a sentinel depends on the number of actors enrolled or the adequacy of the signal’ (Blanchon et al. 2020: np). In closing, we suggest that future successful sentinel technologies depend on managing enrollment. To ensure that care is taken with pathways of illumination, we call for collaborative and comparative ethnographic work across healthcare systems, following the lighthouse beams of contemporary software on its travels towards chronic disease prevention.

References

Beer D (2019) *The Data Gaze. Capitalism, power and perception*. London: SAGE Publications.

Bentzen, N., & Schroll, H. (1990). ICPC. International Klassification for den Primære Sundhedstjeneste. Dansk selskab for almen medicin, Månedsskrift for Praktisk Lægegerning og Fyns Amt.

Bentzen, N and H.Schroll (1992). “Dansk bearbejdelse” Dansk Selskab for Almen Medicin, *Månedsskrift for praktisk lægegerning,* Fyns Amt.

Berg M (1997) Rationalizing Medical Work: Decision-Support techniques and medical practices. Cambridge/London: MIT Press.

Blanchon, D., F. Keck, FM Le-Tourneau, S. Tonnelat, A. Zuniga-Teran. 2020. Sentinel Territories: A New concept for Looking at Environmental Change. https://metropolitics.org/Sentinel-Territories-A-New-Concept-for-Looking-at-Environmental-Change.html accessed 30 March 2022.

Bowker GC & Star SL (1999) *Sorting things out: Classification and its consequences.* MIT press.

Bødker MN, Langstrup H and Christensen U (2019) What constitutes 'good care' and "good carers"? The normative implications of introducing reablement in Danish home care. Health & Social Care in the Community 45(6): 12166–8.

Cruz, T. M. (2020). Data politics on the move: intimate work from the inside of a data-driven health system. *Information Communication and Society* DOI: 10.1080/1369118X.2021.1954972

Cruz, T. M. (2022). The social life of biomedical data: Capturing, obscuring and envisionining care in the digital safety-net. *Social Science & Medicine* 294

DAK-E (2009) Dansk AlmenMedicinsk Database: DAMD Årsberetning 2008.

DAK-E (2010) Dansk AlmenMedicinsk Database: DAMD Årsberetning 2009.

Deming WE (2000 [2982]) *Out of Crisis.* MIT Press. Cambridge, MA

D'Ignacio, C and L. Klein. (2020). *Data Feminism.* MIT Press: Cambridge, MA.

Espeland WN & Stevens ML (2008) A sociology of quantification. *European Journal of Sociology/Archives Européennes de Sociologie*, 49(3): 401-436.

Goodwin C (1994) Professional vision. *American Anthropologists*, 96(3): _____

Grasseni C (2007) Introduction In: Grasseni C (Ed) *Skilled Visions: Between Apprenticeship and Standards.* New York/Oxford: Berghahn Books, pp. 1-19.

Grock, S., Ku, Jh., Kim, J. *et al.* A Review of Technology-Assisted Interventions for Diabetes Prevention. *Current Diabetes Reports* **17,** 107 (2017). https://doi.org/10.1007/s11892-017-0948-2

Haraway D 1991 *Simians, Cyborgs and Women: The Reinvention of Nature.* Routledge/New York

Higgins, V and W. Larner. 2010. Standards and Standardization as a Social Scientific Problem in V. Higgins and W. Larner *Calculating the Social: Standards and the Reconfiguring of Governing* Palgrave Macmillan: Basingstoke, UK. pp 1-19.

Hvidberg, MF, Johnsen SP, M. Davidsen and L. Ehlers. 2020. A Nationwide study of prevalence rates and characteristics of 199 chronic conditions in Denmark. *PharmacoEconomics Open* 4, 361-380.

Kaufman, N. 2010. Internet and information technology use in treatment of diabetes. *International Journal of Clinical Practice* 64(s166): 41-46.

Kingod, N, B. Cleal, A. Wahlberg and G. Husted. 2017. Online Peer-toPeer Communities in the Daily Lives of People with Chronic Illness: A Qualitative Systematic Review. *Qualitative Health Research* 27(1): 89–99. Doi: 10.1177/1049732316680203

Kingod, Natasja. 2018. *The Tinkering M-patient: An (Auto-)Praxiographic Study of Attuning to a Life with Type 1 Diabetes through Online and Offline Support'* PhD Dissertation, Copenhagen University.

Lamberts H and Wood M (2002) The birth of the International Classification of Primary Care (ICPC). Serendipity at the border of Lac Léman. *Family Practice* 19: 433–435.

Langhoff TO, Amstrup MH, Morck P, and Bjorn P (2016) Infrastructures for Healthcare: From Synergy to Reverse Synergy. *Health Informatics Journal*, (24)1: 1-11. DOI: 10.1177/1460458216654288

Lindstrom J, Ilanne-Parikka P, Peltonen M, Aunola S, Eriksson JG, Hemio K, et al. Sustained reduction in the incidence of type 2 diabetes by lifestyle intervention: follow-up of the Finnish Diabetes Prevention Study. *Lancet*. 2006; 368(9548):1673–9.

Lippert ML, Kousgaard MB and Bjerrum L (2014) General practitioners uses and perceptions of voluntary electronic feedback on treatment outcomes – a qualitative study. 5 ed*. BMC Family Practice,* BioMed Central 15(1): 85–9.

Lippert ML, Reventlow S, and Kousgaard MB (2017) The Uses and Implications of Standards in General Practice Consultations. *Health: An Interdisciplinary Journal for the Social Study of Health, Illness and Medicine* 21(1): 3–20. DOI: 10.1177/1363459315590245

Lupton D (2018) Digital Health. Critical and Cross-Disciplinary Perspectives. London and New Your: Routledge.

Manderson, L., and Wahlberg, A. (2020). Chronic living in a communicable world. *Medical anthropology* 39(5): 428-439.

Moth, G, Vestergaard M and Vedsted, P. (2012). Chronic care management in Danish general practice - a cross-sectional study of workload and multimorbidity. *BMC Family Practice* 13(52) http://www.biomedcentral.com/1471-2296/13/52

OECD. (2013). OECD Reviews of Health Care Quality: Denmark Executive Summary, Assessment and Recommendations. https://www.oecd.org/publications/oecd-reviews-of-health-care-quality-denmark-2013-9789264191136-en.htm accessed March 30th, 2022

Pasquale, F. (2015). *The black box society*. Harvard University Press.

Porter TM (1996). *Trust in numbers: The pursuit of objectivity in science and public life*. Princeton University Press.

Quinn CC, Shardell MD, Terrin ML, Barr EA, Ballew SH, Gruber-Baldini AL. Cluster-randomized trial of a mobile phone personalized behavioral intervention for blood glucose control. Diabetes Care. 2011;34(9):1934–42

Randrianasolo, L., Y. Raoelina, M. Ratsitorahina et al. 2010. Sentinel surveillance system for early outbreak detection in Madagascar. *BMC Public Health* 10 (31): 1-8. https://doi.org/10.1186/1471-2458-10-31

Rose, N. (1990). *Governing the soul: The shaping of the private self*. Taylor & Frances/Routledge.

Rosendal, M & Falkø, E (2009) Diagnoseklassifikation i Danmark med fokus på almen praksis. *Ugeskr Læger 2009*;171(12):997-1000

Ruppert E, Isin E, & Bigo D (2017) Data politics. *Big Data & Society*, *4*(2), 2053951717717749.

Schroll, H. (2001). *Methods for registration and analysis of episodes of care in general practice classified according to diagnosis. Elderly patients' contacts to their GP registered prospectively using ICPC codes*. PhD Dissertation,.

Schüll, N. D. (2016). Data for life: Wearable technology and the design of self-care. *BioSocieties*, *11*(3), 317-333.

Scott, J. C. (1998). *Seeing like a state: How certain schemes to improve the human condition have failed*. Yale University Press.

Schultz, J. (2013). Henrik Schroll er tilfreds med fangsten. *Dagens Medicin,* 3 May 2013. https://dagensmedicin.dk/det-vil-gavne-alle-ogsa-samfundet/ accessed 30 Mar

Sieverdes, J C., Trieber F, Jenkins C, and Hermayer . 2013. Improving Diabetes Management with Mobile Healht Technology. *The Americna Journal of the Medical Sciences* 345(4): 289-295.

Slota SC, Hoffman AS, Ribes D, et al. (2020) Prospecting (in) the data sciences. Big Data & Society 7(1): 1-12.

Sofoulis, Z. (2011): Skirting complexity: The retarding quest for the average water user, Continuum: Journal of Media & Cultural Studies, 25:6, 795-810.

Stafford, BM (1998) *Good looking: Essays on the virtue of images*. Mit Press.

Statens Serum Institut (2014a) Bilag til udredning vedr. ansøgninger om godkendelse af DAMD som klinisk kvalitetsdatabase, fra 2007 og frem. Available at: https://www.ssi.dk/-/media/arkiv/indhold/dk-dansk/aktuelt/nyheder/bilag-til-udredning-vedr-ansgninger-om-godkendelse-af-damd-som-klinisk-kvalitetsdatabase-fra-2007-og-frem.pdf?la=da

Thylstrup, N and K. Veel. 2017. Data Feminism from a Feminist Perspective. *Women, Gender and Research* 1: 67-71.

Trnka, Susanna and Catherine Trundle. 2017. *Competing Responsibilities: The Ethics and Politics of Contemporary Life.* Durham, NC: Duke University Press.

USFDA (United States Federal Drug Administration). 2019. Sentinel System: Five-Year Strategy 2019-2023. https://www.fda.gov/safety/fdas-sentinel-initiative. Accessed 28 June 2021

Verran, H. (2001). Science and an African logic. University of Chicago Press.

Wadmann S, and Hoeyer K (2018) Dangers of the Digital Fit: Rethinking Seamlessness and Social Sustainability in Data-Intensive Healthcare. *Big Data & Society* 5(1): 1–13. DOI: 10.1177/2053951717752964

Webster A (2002)Innovative Health Technologies and the Social: Redefining Health, Medicine and the Body. *Current Sociology* (50)3: 443-457

Wei, N. (2014) Empirically Establishing Blood Glucose Targets to Achieve HbA_{1c} Goals. *Diabetes Care* 2014;37(4):1048–1051

Williamson, T, Greene ME, Birtwhistle, R. et al. 2014. Validating the 8 CPCSSN case definitions for chronic disease surveillance in a primary care database of electronic health records. *Annals of Family Medicine 12: 367–372.*

Winthereik BR & Jensen CB (2017) Learning from experiments in optimization: post-critical perspectives on monitoring and evaluation. *Journal of Cultural Economy* (10)3: 251-264. DOI: 10.1080/17530350.2017.1289414

The call of data: A post-mortem of the Danish General Practice Database

John Mark Burnett

Abstract

This article conducts a post-mortem examination of a national database from Danish healthcare that was destroyed in 2015 following an excessive data acquisition from general practice. More specifically, I examine a conundrum that is central to the demise of the database and yet underexposed: How could personally identifiable data on three quarters of the Danish population end up in the Danish General Practice Database when its authorisation was limited to four chronic conditions? To make sense of this conundrum, I put forward the 'call of data' as an analytic by bringing a vital materialist understanding of 'hoarding' into conversation with critical data studies. Using this analytic, I explore how data, in its embedded materiality, exerts an agential and aesthetic imprint on social phenomena. The empirical corpus consists of document analysis and interviews collected between 2017 and 2019, which I use to scrutinise the implementation and governance process between 2007 and 2014. I argue that not only did actors fail to adequately recognise the problematics of the database, they exhibited a significant will - through action and inaction - to ensure the longevity of what in hindsight should have been negotiated. The case study demonstrates how failure to navigate legal and social thresholds acceptability can undermine the social sustainability of data initiatives. As a result, attempts to exploit wasted opportunities in data can themselves result in waste with long lasting effect for public trust. Analytically, the article proposes that the 'call of data' is theoretically useful analytic for engaging with data.

Keywords

Health data, hoarding, Denmark, aesthetics, vital materialism, sustainability

Introduction

In the contemporary policy landscape, investment in digital data is promoted as a central lever to modernise national healthcare systems. The OECD publication (2019) *Health in the 21st Century,* for example, takes the subtitle *Putting Data to Work for Stronger Health Systems*. According to the OECD, health systems are notoriously 'plagued by a significant waste of resources' (ibid., 11) with inefficiencies reported to take up a fifth of healthcare expenditure. 'Putting data to work' is elevated to the subtitle of the report as it conveys a key idea: 'data are a valuable resource, but have no intrinsic value unless put to work within an enabling institutional environment' (ibid., 19). Across sectors, active use is thought to transform idle data into new value constellations and since health systems are already 'data rich', such usage should be as broad as possible if the 'opportunity to save a significant number of lives and billions of dollars' is to be fully exploited (ibid., 11). In this respect, for *Health in the 21st Century* it is not enough to have data – nor is it enough to use them just once.

In this article, I conduct a post-mortem examination of a national database from Danish healthcare that was destroyed in 2015 following an excessive data acquisition from general practice. Designed with the (re)use value of data in mind, the Danish General Practice Database [*Dansk AlmenMedicinsk Database*] (hereafter 'DAMD'), as it was called, became an unsustainable project of data re-use. Implemented in 2007 with an authorisation for the collection of personally identifiable *diabetes* data in the interest of public health, the authorised scope of DAMD's data collection was later expanded with three other disease areas. In late September 2014, the DAMD became a matter of public concern when the Danish Broadcasting Corporation reported that it had (since its implementation) been used to acquire personally identifiable data related to 704 diagnoses (Tynell and Fischer, 2014). The suspected illegal status of the data collection - acquired without the explicit consent of three quarters of the national population - was confirmed by the end of the year. A half a year of fierce, political debate ensued, and the 'unique' database was eventually destroyed to protect patient confidentiality and public trust - despite its critical function and claims on its 'substantial reuse value' (see Author 1 and Author 2, forthcoming: 10).

How the DAMD came to be an unsustainable project has been the subject of prior studies (Langhoff et al., 2016; Wadmann and Hoeyer, 2018). Analysing the sociotechnical expansion of the infrastructure, these studies highlight how an accumulation of actors and agendas resulted in conflicting views on legitimate data use that was fatal for the infrastructure. Shifting the focus from use to acquisition, this article places the DAMD under further scrutiny to explore the following conundrum: *How could personally identifiable data on three-quarters of the Danish population end up in the DAMD when its approval was limited to four chronic conditions?* While this conundrum is engaged by the aforementioned scholarship (Langhoff et al., 2016, Wadmann and Hoeyer, 2018), there is more to say about *how* the DAMD came to exceed its legal remit with severe consequences to follow. To elucidate this question of 'how', I put forward the 'call of data' as an analytic by bringing a vital materialist understanding of 'hoarding' into conversation with critical data studies. Using this analytic, I explore how data, in its embedded materiality, exerts an agential and aesthetic imprint on social phenomena.

The empirical material for my post-mortem of a now defunct database was gathered between 2017 and 2020 in the context of a doctoral study on the data intensification of health in Denmark. I have built my analysis from a corpus of documents regarding the development and demise of the DAMD, ranging from official documents and policy papers to public media and professional journals' media outlets in biomedicine. The empirical base is additionally informed by two interviews from February 2017 and March 2018. Chronologically, the first interview is with the head of the Danish Health Data Authority (HDA) currently responsible for the administration of databases like the DAMD. The second interview was with a central figure in the development of the DAMD who formerly headed its legal data processor, the Danish Quality Unit of General Practice (DAK-E). I respectively refer to my interview informants under the pseudonyms of Jane Jensen and Poul Petersen.

As the article proceeds, I introduce the national setting of Danish General Practice, and make a case for the theoretical utility of a hoarding lens. I then move on to the empirical analysis to respectively scrutinise the implementation and governance of the DAMD between 2007 and 2014. First, I describe how the DAMD was envisaged, detailing the implementation process, and elucidate how the data infrastructure was introduced with a hoarding logic that was subsequently intensified through a collective action to achieve a particular *form*. In the second step of analysis, I demonstrate how the governance process is equally crucial for understanding how the database 'hoard' could grow uninterruptedly for seven years until it entered the limelight of public scrutiny. With the case laid out, I argue that not only did actors fail to adequately recognise the problematics of the DAMD, they exhibited a significant will - through action and inaction - to ensure the longevity of what in hindsight should have been negotiated. In closing, I discuss how the case study is relevant for the social sustainability of contemporary data initiatives and propose that the ‘call of data’ is theoretically useful analytic for engaging with data with a suggestion of its possible use.

Background

> In god we trust, all others bring data

I first encountered the DAMD case during my preliminary field mapping. Initially, I was looking for new, innovative initiatives that could be useful as an ethnographic case study to explore the emerging practices and politics of big data. At the various sites I attended in the Danish biomedical sector, I noticed how attempts to talk about the future of health data rarely failed to mention the DAMD, despite its destruction one or several years earlier. This was also true for my interview with Jensen from the HDA[1], which is where I found the above epigram as a print-out in my informant’s office. The process of establishing this dedicated authority for *health data* was, according to my informant, a manifestation of Denmark’s political ambition: to develop and maintain its leading position regarding digitalisation and documentation in times when decisions are becoming increasingly ‘data-driven’.

When I commented on the print-out, positioned in plain sight near her desk, Jensen told me that the author was William Edwards Deming (1900–1993), an American management consultant with a background in engineering. According to popular accounts, Deming was ‘the genius who revitalised Japanese industry’ (Walton, 1988: front page) after World War II by promoting the application of statistical techniques of quality control to the manufacturing of goods.[2] Prominent in the 1980s, Deming built on variation theory in statistics, seeking to wield information as a corrective 'feedback device' to optimise quality and productivity levels in organisations. As the placement of Jensen’s office print-out suggested, this scientific approach

---

[1] The HDA took over sectoral responsibilities for documentation and IT from Statens Serum Institut (SSI) in November 2015, less than two years before the interview.

[2] Some are however more critical of the personal contribution of Deming – the ‘quality guru’ – to the economy and management paradigm of quality control in Japan (Tsutsui, 1996).

to continual improvement has today made its way from industry to public administration, coinciding with New Public Management along the way. The conditions for making Danish healthcare more 'data-driven' have, however, been uneven across its sectors, where primary care specifically has been problematised as an 'information void' compared with the hospital sector (Forde et al., 2016: 33).

When I asked why primary care was singled out in the new policy objective for 'better knowledge about the entire healthcare system' (HDA, 2017: 4), Jensen turned the conversation to the DAMD*;* the database that accomplished an unprecedented - but short-lived - achievement of replacing the information void in General Practice with data on three quarters of the population:

> …there has been a long case about a database that was illegal and that, so it is a bit of a sensitive area. But you just have to realise that if we are to see what is wrong with people, what pathways exist in the healthcare system, whether we are diagnosing diseases at the right time or using resources appropriately, then we don't really know what is happening in general practice.

Twenty-one months after its destruction, the database was notably still a touchy topic in our interview, but Jensen bounced off of it, to focus on the responsibilities of the healthcare system as a whole: on diagnosis, resource use, and care pathways. At the root of her insistence ('you just have to realise') is the desire to 'really know' what is happening in general practice, and the worry that without access to necessary information for documenting and evaluating outcomes, decisions regarding treatment and resource allocation in the healthcare system are somewhat taken 'in the blind'.

Danish general practice

Other nations than Denmark have managed to develop infrastructural capabilities to automate the acquisition of computerised data on patient encounters from general practice in systematic fashion (e.g. New Zealand, Scotland, Ireland, England and Holland) (Schroll, 2009). Nevertheless, the DAMD was distinguishable in a number of other ways. For instance, many of the other initiatives only acquired anonymised data for the sole purpose of research and sometimes with funding from the medical industry. This was however not the case with the DAMD, which was developed in a tradition that is in several ways particular to the Danish context.

In Denmark, as in many countries with national healthcare systems, General Practice is ordinarily the patient's first access point. General practice thus functions as a 'gatekeeper' to other healthcare providers, whether specialist or hospital care. Similarly to the UK, general practice is a private enterprise that operates as a contractual extension to the Danish welfare state wherein healthcare is universal. Every three years, the terms of general practice are collectively negotiated by the Organisation of General Practitioners (PLO) and the regional authorities. The acquisition of data from Danish general practice was once driven by internal pioneers who, with the support of professional associations, sought to advance research but also the quality of clinical practice (Hartlev and Wadmann, 2018). The early development of the DAMD was similarly shaped by a small group of pioneers for both purposes. In 2006, the DAMD was written into the collective agreement, where it was financed by the negotiating parties to promote quality improvements in chronic care. With an initial focus on diabetes, the national implementation of the DAMD began 1 April 2007. Increasingly, the DAMD would become crucial to the state institutionalisation and centralisation of quality development for general practice, which had been underway in the hospital sector since the 1990s (Hartlev and Wadmann, 2018).

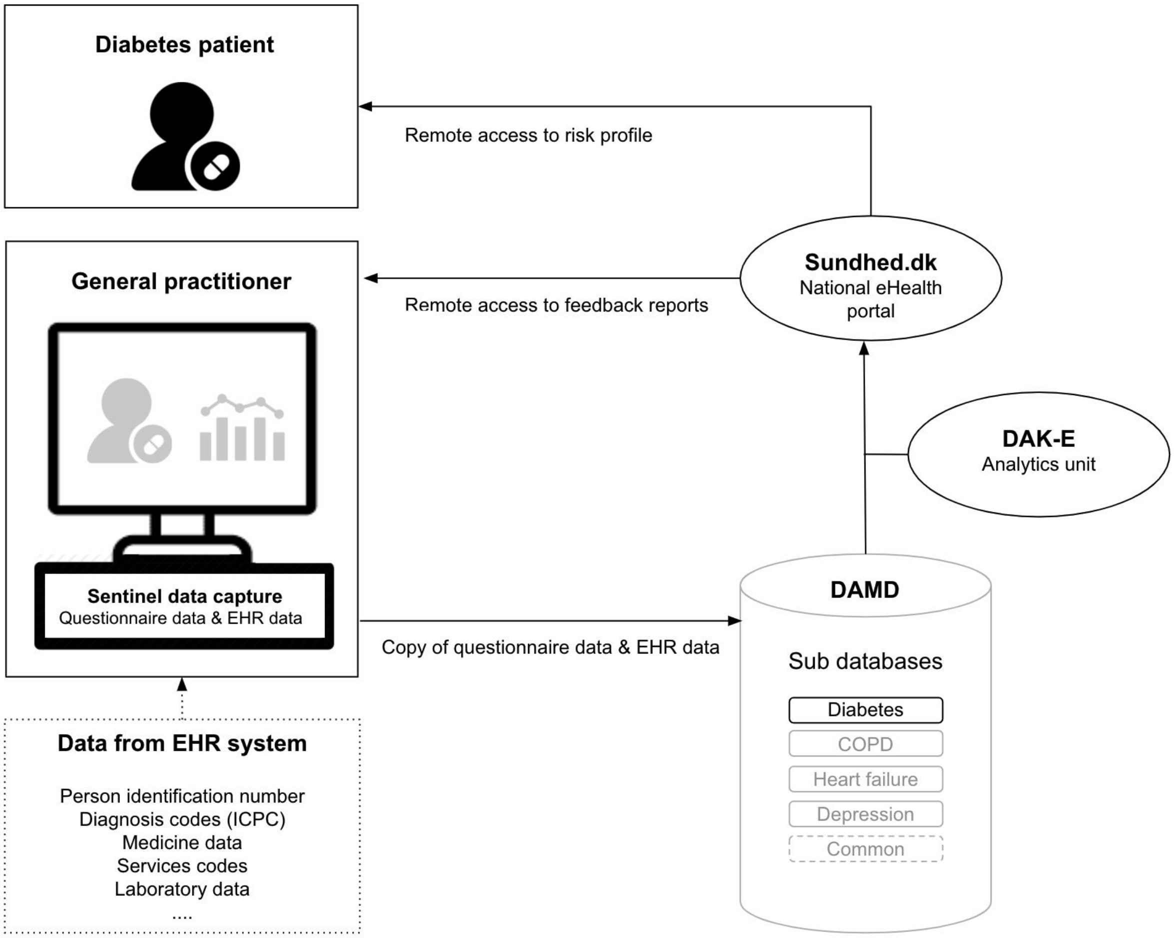
The national implementation of the DAMD occurred in a time when more than one fourth of the Danish population was estimated to suffer from chronic illness (Kronisk Sygdom, 2005). In terms of diabetes, regional authorities reported that the number of patients *diagnosed* with this illness was somewhere between 100.000 and 150.000 with concern[3] that the *actual* number was 'considerably' higher (SSI, 2014a: 273). It is against this backdrop that the DAMD was implemented to curb the socioeconomic burden of diabetes, as data-driven quality improvements in primary care was envisioned to reduce the risk of comorbidity and costly hospitalisation. To this end, the DAMD supported a new decision support tool called Sentinel Data Capture (Sentinel), designed to help general practices monitor the quality of diabetes care and thereby self-evaluate the effect of their work to target improvement efforts.

General practices that opted to adopt the tool were required to have a local installation of Sentinel on their computer to transfer relevant data to a disease-specific sub database in the

---

[3] The risk of comorbidity in the circulatory, nervous, urinary and eye systems are exacerbated when unregulated.

DAMD. Most data were automatically retrieved from the 12 electronic health record systems operating in the Danish market (e.g. diagnostic codes, prescribed drugs, laboratory results). Supplementary data on for instance lifestyle however needed to be registered manually during annual controls through a digital questionnaire. Centralised in the DAMD, the data would be cleaned and algorithmically processed to produce disease-specific quality 'indicator sets' developed and overseen by DAK-E.[4] Users (i.e. professionals in general practice using Sentinel) would gain access to data analytics in the form of weekly 'quality reports' at Sundhed.dk, the national eHealth portal. For the first time, general practice would be able to *self-evaluate their quality of care* based on a stratified overview of their own diabetes population that could be compared with colleagues (on an aggregate level of municipal, regional and national) and national benchmarks.



---

[4] Indicator sets were coordinated with the National Indicator Programme to support the production of anonymised public audits across the various sectors in Danish healthcare regarding specific diseases.

Figure 1. Author's diagram of the technical and organisational landscape regarding the provision of feedback reports on diabetes

The acquisition of diabetes data was initially approved by the National Board of Health (NBoH) under the legal framework for 'clinical quality databases' (CQD). With this classification comes a legal exemption from the usual requirement of explicit consent for the acquisition of personal health data. As a result, CQDs are legally obliged to operate under a set of limitations that notably confine the legal power of this classification to specific data subjects and purposes. For DAMD, the data subject limitation referred to a specific patient group/disease area while the purpose limitation related to the surveillance and quality development of patient treatment. By Autumn 2014, the approved scope of data acquisition was expanded to support quality reports for three additional disease areas: chronic obstructive pulmonary disease (COPD), heart failure and depression. Each disease area was defined according to the corresponding diagnosis code in the *International Classification of Primary Care* (ICPC) (e.g. 'K77' for heart failure) and stored in a disease-specific sub-database (see figure 1). The approved collection of data on one disease, diabetes, had become four.

Hoarding and the 'call of data'

It is the expansion of these four disease categories towards the collection of personally identifiable data on three-quarters of the Danish population that has brought my attention to the analytic potential of the hoard. While my analysis explores the extent to which the DAMD be usefully considered a 'hoard', I first provide some background to my choice of framing. The etymological origins of hoard (noun) is a Gothic derivative (*huzd*), which literally means 'hidden treasure' (Online Etymology Dictionary, n.d.). This metaphor appears in dominant discourses that speak to the political economy of health data. In the Nordic welfare states, and Denmark in particular, the metaphor appears in the guise of a 'goldmine' to convey the perceived value of national health data registries (Bauer, 2014) and, more recently, the prospect of a combined 'Nordic data gold mine' (Tupasela et al., 2020).

The metaphor however also has analytical value since the notion of concealment posits that the hoard can be visible to an observer but only treasured by the hoarder. As such, the hoard can be seen as something that should be treasured and trashed at once when competing views are present. This is illustrated in the empirical analysis where competing valuations are assigned to the DAMD when seen from a perspective of public health and legal protection of rights for individual data subjects. Hoards do not however appear from nothing, but through a particular kind of acquisition practice that adheres to the characteristics of 'hoarding'. Looking to cultural anthropologist Natasha Dow Schüll, we can understand 'digital hoarding' as 'the excessive saving, archiving, and storing of digital artefacts, along with a reluctance to delete, even when the content in question is redundant, irrelevant, or no longer valuable to the owner' (Schüll, 2018: 44). In short, digital hoarding can be viewed as a reluctance to discard an excessive acquisition of data. Although hoarding has a history in clinical medicine[5], my analytical use of the concept lies elsewhere. Building upon work from political theorist Jane Bennett, the following introduces hoarding through a vital materialist lens that decentres conventional understandings based on clinical medicine. Inspired by Bennett's advocacy of an approach that takes the 'call of things' seriously (Bennett, 2012: 247) I explore what it might entail to take the 'call of *data'* seriously.

Hoarding in vital materialism

In *Powers of the Hoard* (2012), Bennett engages the notion of hoarding from a perspective of vital materialism. In contrast with a psychopathological focus on subject behaviour, Bennett problematises the ontological bias of depriving objects the possibility to speak before the study even begins. Writing against anthropocentrism Bennett asks: what if the objects that come to form hoards also play an active role in the phenomenon of hoarding? Bracketing the explanatory powers of psychopathology, Bennett attunes her analysis to the possibility of 'thing-power and distributed agency' (2012: 241). Her primary interest is in *what the hoard - or object hereof - does to the hoarder* instead of vice versa. One implication that follows from this reframing is a depathologisation since the hoarder is seen to possess a 'special gift or a special ability' rather than a 'deficiency' (2012: 245). The gift or ability that Bennett speaks of is an 'extreme perception' to locate value in objects that non-hoarders fail to recognise. So, how hoarders come to see 'the space and the colour and the texture and the form' as value properties of what most would consider a 'useless' bottle cap (Bennett, 2012: 245). In this sense, Bennett argues,

---

[5] Whereas hoarding has been recognised as a clinical diagnosis since 2013 (Steketee and Bratiotis, 2020), digital hoarding has as of yet not attained the same level of recognition despite calls to class it as a subtype of mental disorder (e.g. van Bennekom et al., 2015).

hoarders share certain affinities with artists as they are both particularly susceptible to what she describes as ‘the enchantment-powers of things’ or the ‘call of things’ (Bennett, 2012: 247).

Although Bennett concedes that ‘It is not normal today to think of 'inanimate objects' as possessing a lively capacity to do things to us and with us’, she argues that ‘it is quite normal to experience them as such’ (2012: 263). When hoarders for instance try to explain ‘the insistent allure’ of things in a thrift shop or dumpster, they tend to describe a sensation of being ‘overwhelmed’ or a feeling that ‘things just took over’ (Bennett, 2012: 252). Bennett reads these statements as evidence for her claim that hoarders affirm the existence of material agency; ‘they experience the hoard as having its own momentum or drive to persist and grow’ (2012: 252). Although non-hoarders are generally not considered to possess a special gift, they may too experience the ‘call of things’ in their daily lives (e.g. possessions, tools, keepsakes or commodities) and become attached for reasons that are irreducible to what they project in them. My adoption of hoarding as an analytical lens is *not* intended to label the actors in the empirical case as hoarders - and certainly not in its clinical understanding. My intention is instead to comprise a sociomaterial approach that is open to the idea that the call of things might have played a role in the empirical acquisition of data that, according to the existing legal framework, was ultimately deemed excessive. Bennett’s engagement with hoarding is however also an eco-critique since the notion is suggested to convey ‘the madness appropriate to a political economy devoted to over-consumption, planned obsolescence, relentless extraction of natural resources ("Drill Baby Drill"), and vast mountains of disavowed waste’ (Bennett, 2012: 248). The call of things is thus for Bennett part of a posthuman project to illuminate consumption practices in a different light in an attempt to make them more sustainable. The analytic I propose seeks to extend this project by specifying the focus on *data* and the *social* sustainability of their consumption.

The ‘call of data’

When Bennett attends to the call of things it is then an analytical choice to take seriously what anthropocentric studies of hoarding tend to miss or ignore: How the agency and aesthetics of material objects might illuminate the phenomenon of hoarding. But how might Bennett’s work in vital materialism be understood when the item of acquisition is *digital* data? Responding to such curiosity, the following relates two aspects of her work - materiality and aesthetics - to digital data and work in critical data studies.

Although its aesthetics may not be appreciated by everyone, all can agree that a bottle cap has a particular feel and materiality to it. Such consensus is currently not present for digital data, which are often conceived as an immaterial object bereft of tangible substance. In his historical analysis of database technologies, Dourish (2014: 5) however argues that ‘the digital is always, *inherently*, and inescapably material’. Despite the elusive metaphor of 'the cloud', digital data *are* stored on hardware kept in physical location, as studies engaging the material form of clouds highlight (Amoore, 2016; Maguire and Winthereik, 2019). Dourish is however less concerned with the ‘brute materiality’ of large-scale infrastructures (2014: 31) than with the ‘existing and consequential materialities of digital systems and digital representations, from the flowchart to the algorithm, the data structure, and the virtual machine’ (2014: 5). In his analysis, Dourish arrives at the conclusion that it matters whether aspects of the world are encoded in a relational database or NoSQL database. Why? Because they are argued to offer distinct material specificities that shape how information can be recorded and consequently made amenable for processing. Inspired by this approach, I go on to examine the material properties of the information infrastructure that contributed to the specific form of the data stored in the DAMD.

One thing is describing the *form* of data, another is to get at its *call*. Here, the analysis will be sensitised to the aesthetics qualities of the data acquired in the DAMD. In *Beautiful Data* (2014), Orit Halpern offers an aesthetic history of data, observing how data and evidence are often embellished with the label of 'beautiful'. Increasingly, solutions are required to be 'smart' and 'elegant', tying the beauty of data to their utility (Halpern, 2014: 1) and highlighting ‘a new aesthetic and practice of truth’ emerging in the aftermath of World War II: ‘a valorization of analysis and pattern seeking’ (Halpern, 2014: 15). Although the example momentarily sidesteps the question of what data should be acquired in the first place, the initiative behind the FAIR Data Principles illustrates how the aesthetics of data are tied to infrastructural practices in contemporary science. In this example, Findability, Accessibility, Interoperability, and Reusability (FAIR) are championed as guiding principles for ‘good’ data management and stewardship to exploit the affordances of machines in the reuse of scholarly data (Wilkinson et al., 2016). While Halpern tells us how the current aesthetics of data may have come about, my approach explores the capacity of aesthetics to *affect* what eventually made its way in the DAMD. In this respect, my approach is similar to what Walford describes in *Data Aesthetics*, namely ‘the ways in which data's forms are charismatic, how they capture attention, or command action, in ways other than because of their objectivity or accuracy' (Walford, 2021: 206). Instead of imposing philosophical 'judgements about the beautiful and good', Walford (2021: 205) understands aesthetics to be 'about what constitutes a 'persuasiveness of form' in any cultural or social context (Strathern 1991: 10; cf. Riles 1998)'. Whereas Walford observes that the data's forms may command action, the present case study extends this observation by demonstrating how the forms of data can also command 'inaction'. What I put forward with the call of data is an analytic to explore how data, in its embedded materiality, exerts an agential and aesthetic imprint on social phenomena; which in this case concerns the specific call of the DAMD and its contribution to its illegal status.

The call of the Danish General Practice Database

With the conceptual and methodological underpinnings established, we now begin the analysis of the empirical conundrum: How could personally identifiable data on three-quarters of the Danish population end up in the DAMD when its approval was limited to four chronic conditions? Using the introduced analytic, I answer this question by scrutinising sociomaterial convergences in the implementation and governance process that contributed to this outcome.

In the first step of analysis, I detail how the DAMD was implemented with a hoarding logic that was subsequently intensified by stakeholders who attempted to marshal the database into a particular *form.* The intensification of this logic is traced to a collective action that in particular involves a reconfiguration of the material and contractual terms of implementation. In the second step analysis, I demonstrate how the action *and* inaction of stakeholders in the governance process is equally crucial for understanding how the database 'hoard' could grow uninterruptedly for seven years until it entered the limelight of public scrutiny. Throughout the analysis, I unfold how two competing 'conventions' were used to adjudicate the value of data that ended in the DAMD. The first convention concerns the legal protection of individual rights for human subjects. Moving from the individual to the collective, the second convention regards the matter of public health.

Implementation process

While DAMD eventually had a vast amount of data, its initial scope was quite modest. According to the official accounts, the approval for DAMD was initially issued for the specific disease area of diabetes (SSI, 2014a: 16), and subsequently the legal framework of CQDs was used to acquire data on four select disease areas between 2007 and 2014. As noted, CQDs are classed with the legal authority to waive individual rights in the interest of public health. In order to justify the use of this authority, CQDs are required to deliver a 90% national coverage of the specified patient group within the approved timeframe. The importance of this condition was articulated by the regional authorities in the funding process of 2006, where they stressed that 'the utility of the database is highly dependent on the data being near complete' (SSI, 2014a: 274).

CQDs are thus obliged to take on a certain *form* - at a particular pace - to exist; one that appeases a legal definition of 'good data' informed by scientific ideals of representation that adequately mitigates the risk of population bias. The DAMD was however granted dispensation from this condition. Instead, the DAMD was expected to deliver national coverage of diabetes between 70% and 90% by April 2010 (SSI, 2014a: 274). With an annual increase of 30 clinics, only 84 (10,7%) of the 2000 clinics were registered to transfer diabetes data to the DAMD by the end of 2008. Not only was this 'slower than desirable' (DAK-E, 2009: 16), it threatened the prospect of a renewed approval since it was significantly short of the agreed target. By the end of the three-year approval, only 17% (n=340) out of 2000 clinics were registered to transfer diabetes data.

Despite these shortcomings, the DAMD was written in the subsequent collective agreement beginning 1 April 2011. This second collective agreement put in motion a series of changes that had a profound effect on the accumulation of data from general practice. This time, the DAMD was expected to deliver a national coverage of 80% by the end of 2012 (SSI, 2014a: 75). These changes diminished the space for professional autonomy on the part of doctors by replacing user choice with contractual obligation. The first modification was to make the adoption of Sentinel mandatory for clinics by April 2013. This change proved an effective response to the underwhelming user uptake acquired with the opt-in approach used in the first collective agreement. In the new collective agreement, users were, secondly, contractually obligated to register questionnaire data for seven disease areas and ICPC code seven conditions along with non-psychotic mental disorders. Previously, the user benefits of Sentinel were reported to provide a sufficient incentive for its adoption and related labour of datafication (DAK-E, 2009: 16). A third modification entailed the mandatory transfer of the (prospective) data capture of Sentinel to the DAMD. This was implemented by removing a 'switch' in the Sentinel interface. Previously, users could access this switch to configure what could be copied from the clinic to the DAMD, if any, from three options: (1) 'Everything', (2) 'Pr. Project', or (3) 'Nothing'. Given the gravity of this reengineering, a brief elaboration is in order.

Option two would delimit the data transfer to support the production of feedback reports or approved research projects (e.g. on the pathological evolution of symptoms, such as breast pain). Option one would by contrast include data on *all* ICPC-coded patient encounters, be it diabetes type-2 (T90) or marital problems (Z12). The catch-all approach represented by option one was included as an intentional design choice from the beginning of the national implementation of the DAMD. Back then, this design choice was by DAK-E envisioned to '*future proof* registrations for other upcoming clinical NIP [National Indicator Project] databases and it simultaneously provides unique opportunities for retrieval of data for quality development in the practice sector and research in quality' (SSI, 2014a: 21, author's emphasis). This statement has not aged well since the design configuration would in the long-term prove to be anything but sustainable. In the short-term, however, the configuration did have its benefits.

According to my interview with Petersen, the former head of DAK-E, users 'had a lot of motivation' to choose option one for swift access to quality reports where access is conditioned upon six months of data. A pre-emptive transfer of 'everything' would in this respect cover the longitudinal requirements when indicator sets are developed for new disease areas. According to a user manual from 2008 (DAK-E: 3), selecting option one was also articulated as a means to deliver data for 'epidemiological purposes that can reveal [*vise*] encounter patterns in the general practice'. While the user benefit of immediacy and support of epidemiology may not be the sole reason,[6] 90% of all clinics (n=184) were reportedly configured to transfer 'everything' the year before this option became a contractual obligation (DAK-E, 2009: 23). Nevertheless, it would be a leap of faith to assume that the popularity of option one would remain stable considering the contractual changes that imposed additional demands on new and existing users alike (see Figure 2).[7] As Langhoff et al. argue (2016), the sociotechnical arrangement of the material infrastructure was eventually reconfigured to an extent that it took on a fundamentally different

---

[6] To which extent option one was a default setting upon installation is unresolved, but a number of users later claimed to be unaware of the scope of their contribution - a claim repudiated by DAK-E (Schroll and Friborg, n.d.: 1).

[7] The DAMD also acquired new purposes for data reuse (for an overview see Wadmann and Hoeyer, 2018: 6).

ontology. In the first two years of the second collective agreement, the adoption of Sentinel among the 2012 clinics in Denmark increased from 17% (n = 340) to 94.4% (n = 1900; DAK-e 2014). So, while the contractual intervention managed to extend the national coverage of Sentinel beyond 90%, it did so without restricting the data acquisition to the approved patient groups.[8] As a result, the DAMD simultaneously met and exceeded the legal requirements for a CQD.

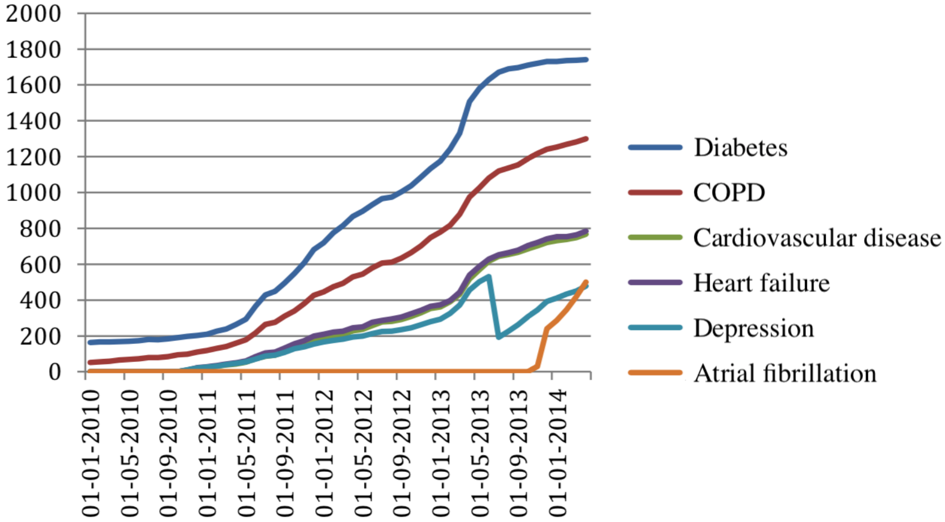


Figure 2. Development of the number of clinics subscribed to transfer disease-specific data to the DAMD between 2010 and 2013 (DAK-E, 2014: 10, translated by author)

What this first part of analysis highlights is how the acquisition of data was not due to the actions of one actor. What can be observed is in contrast a story of distributed accountability involving the collective action of several actors. The system designers configured the material conditions of possibility for users to transfer 'everything' as a means to anticipate future demands for clinical data. Manual writers helped communicate the user benefits of this catch-all option that 90% of users eventually applied as a default setting. Lastly, the negotiating parties in the second collective agreement made the catch-all option compulsory for all users, along with the datafication of several disease areas.

In each case, the resulting data acquisition occurred under the assumption that it could

---

[8] At one point, approval was not present for any of the disease areas recorded in this period (figure 3.)

deliver a future epistemic return of value to public health. This value was in turn premised upon that the data acquisition would adhere to a certain *form*, which also shaped the eventual content of the DAMD. This form was in part governed by the technoscientific (i.e. longitudinal requirement) and legal (i.e. national coverage requirement) ideals of 'good data' that had implications for the individual and collective acquisition of data. These requirements supplement a range of others that made the form of the data desirable to acquire and use. The orderly structure, centralised accessibility and machine-readability of the data were also among the properties valued by researchers and others with an interest to 'reveal' patterns through computational processing.

Revisiting digital hoarding, the concept can in simple terms be summarised as a *reluctance* to discard an excessive *acquisition* of data. This definition is closely tied to the problem of valuation introduced in the opening of this article - how to *know* what data should be acquired and when they should be discarded - but highlights the matter of conduct above epistemology. Schüll’s work on *Digital containment and its discontents* (2018: 44) does not only provide a useful definition of digital hoarding, it also highlights how the contemporary relevance of the outlined problem is shaped by technological advancement: ‘While the material world provides a natural check on stockpiling and hoarding by asserting itself as a problem to contend with when it swells in size, the swelling of informational stuff is invisible enough to render discarding more cost-intensive than retaining’. This has an impact on the incentive structure used to decide what to acquire and discard, as the logistical and economic cost of storing population-wide data is significantly reduced with digital technology compared with paper-based practices.

Another aspect of the aforementioned problem relates to the promise of epistemic return from pattern seeking that is argued to be ‘at the heart of the current enchantment of big data’ (Walford, 2021: 214). In this setting, impetus to hoard is added by an increasing difficulty to predict the value and purpose of data sources. The evolving nature of data sources - in which non-biomedical data can be reused for biomedical purposes and vice versa - does not only complicate conventional understanding of ‘health data’ (Vayena and Gasser, 2016). It also complicates the practice of knowing what to acquire and keep, against which hoarding may appear as a practical solution to appease the data demands of an uncertain future.

Governance process

So far, I have highlighted changes primarily imposed by a contractual intervention, which intensified the manifestation of a logic built into the infrastructure from the beginning. To make sense of the conundrum, I argue that the conduct of key actors in the governance process was equally vital to the accumulative achievements of the DAMD. Before I move on to the governance process, it is first helpful to set the scene of Autumn 2014, when public institutions were forced to address a public contestation of the legal status of the DAMD.

As stated in the introduction, the public life of the DAMD began in late September 2014, when the *Danish Broadcasting Corporation* reported claims of an unlawful data collection from general practice (Tynell and Fischer, 2014). Within a week, SSI had instigated an internal investigation and issued an instruction to DAK-E (the data processor) and Region Southern Denmark (the data controller) for the temporary suspension of data acquisition regarding all but four disease areas: diabetes, COPD, heart failure, and depression.[9] In late November, the investigation affirmed legal support for the acquisition of data regarding four of out five sub-databases, although each had issued a three-year approval by SSI two years earlier. This was reasoned in accordance with the Danish Health Care Act, in which paragraph 196 exempts the requirement of consent for CQD is limited to 'a defined group of patients' for purposes of surveillance and quality development; (SSI, 2014b: 1). Whereas the four sub-databases were respectively limited to the above disease areas exempted from the temporary suspension, the fifth sub-database was – with a coverage of 704 ICPC codes – effectively a catch-all database. The investigation concluded that the legal status of this 'common database', as it was called, would need to be tested in accordance with the Danish Act on Processing of Personal Data by the Danish Data Protection Agency. By the end of December 2014, the common database was found illegal in accordance with both acts, which raises the question of how it was approved in 2012.

A contested scope of approval

According to the official statement from SSI, their approval of the common database in 2012 was issued under the impression that its scope was limited to 'the sum of the content of the four

---

[9] Self-access (via Sundhed.dk) to patient records retrieved from the DAMD was also blocked in the same week.

disease-specific databases' (SSI, 2014b: 22). The regional authorities (SSI, 2014a: 299) and DAK-E, however, claimed that SSI were cognisant of the full extent of data acquisition, which is noted twice in the approved application from 2010: 'the DAMD is fed with data from patient encounters in general practice for all the different disease and health problems that patients address to general practice' (SSI, 2014: 230). This was additionally reflected in the letter of approval that, in contrast to the four other sub-databases, did not specify the scope of approval to a delimited patient group. If the official stance from SSI is to be accepted, it raises the question why the four disease-specific databases were needed if their aggregated content was included in the common database. Alternatively, why was there a need for the common database if its content was considered to constitute the sum of the four disease-specific sub-databases? Regardless, the scope of approval remains a contested matter. Similar to the implementation process, the question of accountability is neither straight-forward nor necessarily reducible to one actor. Reviewing the documents disclosed with the internal investigation, however, provides ample evidence to assert that central actors were aware of the practical scope of data acquisition. Furthermore, there are at least two episodes where the practical scope was flagged as a legal concern: one beginning in Spring 2007 and the other in Winter 2011. I review these in order to consider the temporal, aesthetic and material dimensions of what it meant to continue to collect data.

Articulation of legal concern - Episode I

In the first part of the analysis, I introduced how DAK-E described the catch-all approach to data acquisition as a 'future proof' tactic. This description was formulated in a memo sent to the NBoH[10] in April 2007 regarding the applied approval of the diabetes (type-2) sub-database issued a prior month. The undelimited scope of the practical data acquisition was thus known by the applicant *and* authorising entity in the approval process from the beginning; a point that the regional authorities claim SSI could have articulated more clearly in their draft of the investigation report (SSI, 2014b: 26). More troubling, however, is that the memo referenced a legal concern from a meeting held a month prior, flagging that 'the approval might cause problems' (SSI, 2014a: 21). Although the minutes of the meeting are undisclosed, one of the

---

[10] I.e. The National Board of Health

problems was indicated in an email from the Ministry of Health in response to a request for legal advice from the NBoH (SSI, 2014a: 18):

> Every time you (the database in general practice) want to expand the data type (area/purpose) […] the area requires approval from the NBoH (so the consent requirement can be ignored legally). This puts the brakes on the development of general practice significantly, given that 45 percent of all the doctors today code everything (and is thus retrieved automatically) to the database [DAMD].

The legal tensions regarding the DAMD are clearly articulated in the above summary, almost to the point of prophetic foresight. Yet, the wording of these tensions frame law as an obstacle to technoscientific progress (Hurlbert, 2015). The legal requirement of subject limitation meant that expansion of the data acquisition could only be approved incrementally, from one disease area to another. The NBoH is thus positioned as an obligatory point of passage that 'puts the brakes' on the full exploitation of what no other database had accomplished in Danish general practice; that is, the *centralisation* of data from an entire sector, which for technical, legal, and cultural reasons had traditionally been a hinterland of the national health data landscape.

Embedded in the above quotation is a logic of waste and recycling, that Thylstrup (2019) argues is the premise for the logic of datafication. This logic is for instance observable in the OECD (2019: 3) policy for *Health in the 21st Century* where national health systems are advised to mitigate 'wasted' opportunities in their 'data rich but information poor' landscapes. In the quoted problem formulation, the data from 45 percent of the general practitioners are articulated as low-hanging fruit - 'given' that it exists in a form that is *readily available* for recycling - to gain a surplus value for public health. Yet, it was endangered of being trashed as a wasted opportunity when seen in relation to the legal framework of CQD. While the coded data were retrievable through digital automation, their utility for quality development was generally limited to purposes where the ideal of national coverage was less important (e.g. quality reports).[11] If national

---

[11] Out of the 94,4% (n=1900/2012) clinics that adopted Sentinel by April 2013, the number of clinics that transferred everything, and ICPC coded 70% or more of face-to-face patient encounters, amounted to one third (n=613) (DAK-E, 2014: 11).

coverage is to be considered a decisive criteria for its value to public health, then one could argue that the data were primarily desired more for its technical availability than practical utility. This can then be viewed as an example of how data can 'call out' by means of its embedded technological arrangement.

The email from the Ministry of Health concluded with the discretion that the aforementioned problem was a matter of 'interpretation issues', which could be overcome with a meeting between the NBoH and The Danish College of General Practitioners (DSAM). According to the revised minutes from this meeting (SSI, 2014a: 33), held 23 April 2007, both actors acknowledged that the DAMD could only be approved as a CQD for data on specified disease areas. The NBoH, however, also noted that data registrations for CQDs should by law also be mandatory. This irregularity was duly noted by DSAM, who expressed an intention to abandon the implemented opt-in approach in the next collective agreement. The minutes stated that, given time, the DAMD could in principle come to function as a Danish National Patient Register – the registry used to record patient contacts from public hospitals since 1973 – for the private enterprise of general practice. In order to take on this imagined form, the minutes concluded that NBoH was to 'assess the possibility of approving DAMD as a *profession*-specific [CQD]' (SSI, 2014a: 33, author's emphasis). Concluded in July 2007, and confined to the jurisdiction of authority of the NBoH, the assessment found no 'barriers for DSAM's quality-promoting initiative to the extent that a [CQD] is sought approved' (SSI, 2014a: 36).[12]

Articulation of legal concerns – Episode II

The second episode begins in December 2011 when an EHR vendor flags the (newly implemented) contractual scope of data acquisition as a legal concern. In an email to the NBoH, the EHR vendor requested verification of the legitimacy of a claim put forward by DAK-E regarding a 'general approval' in specific reference to paragraph 196 of the Danish Health Act: 'DAK-e requires the registration of data to DAMD to include <u>all</u> citizens, regardless where a treatment is initiated[…] and without reference to a delimited group of patients' (SSI, 2014a: 101, *original emphasis).* The email highlighted that the claim from DAK-E concerned 500.000 patients

---

[12] The assessment also stated that the 'Use of data in DAMD, aside from approved [CQDs], belongs to the data protection regulation [*dataloven*] and is an area for the Minister of Justice' (SSI, 2014a: 36).

and noted that the request for verification was also sent on behalf of their clients in general practice.[13]

After two follow-up emails in 2012,[14] the EHR vendor's inquiry to the NBoH was eventually answered by SSI 15 months later in March 2013. Following the initial inquiry, the responsibility for handling approvals was reallocated from the NBoH to SSI after an organisational restructure in March 2012. In their email reply, SSI confirmed that the Danish Healthcare Act offered no legal support for a 'general approval' while conceding that, at the time of the initial inquiry, necessary approvals were exclusively present for the diabetes sub-database.[15] SSI however noted that the four aforementioned disease areas had since been approved—in practice, three-year approvals were issued in December 2012 for each of the four disease-specific sub-databases and the common database. For questions concerning the acquisition of additional data, SSI alluded to a 'data processing agreement' governed by the Ministry of Justice (SSI, 2014a: 242). The EHR vendor was eventually referred to the Danish Regions after making it clear that they, SSI, had 'no knowledge about it in any way' (SSI, 2014a: 242).

Given the severity of the EHR vendor's claim, one could ponder whether SSI should have taken a more proactive course of action to ameliorate the articulated knowledge deficit - even though it would exceed its jurisdiction. Staying within the jurisdiction of SSI, a few details regarding the approvals are worth highlighting. Aside from the fact that the approval applications were submitted and processed while the EHR vendor was awaiting a response, the approvals of December 2012 were issued with retrospective effect from February 2012. Furthermore, it is protocol that all CQDs are registered with the Danish Data Protection Agency. Yet, each approval application references the same registration document from 2008 – formulated more than three years before the application was written – which crucially only includes one of the five sub-databases approved in December 2012: the diabetes sub-database. The application noted diabetes to have had national coverage since May 2007, but this was also the case in the

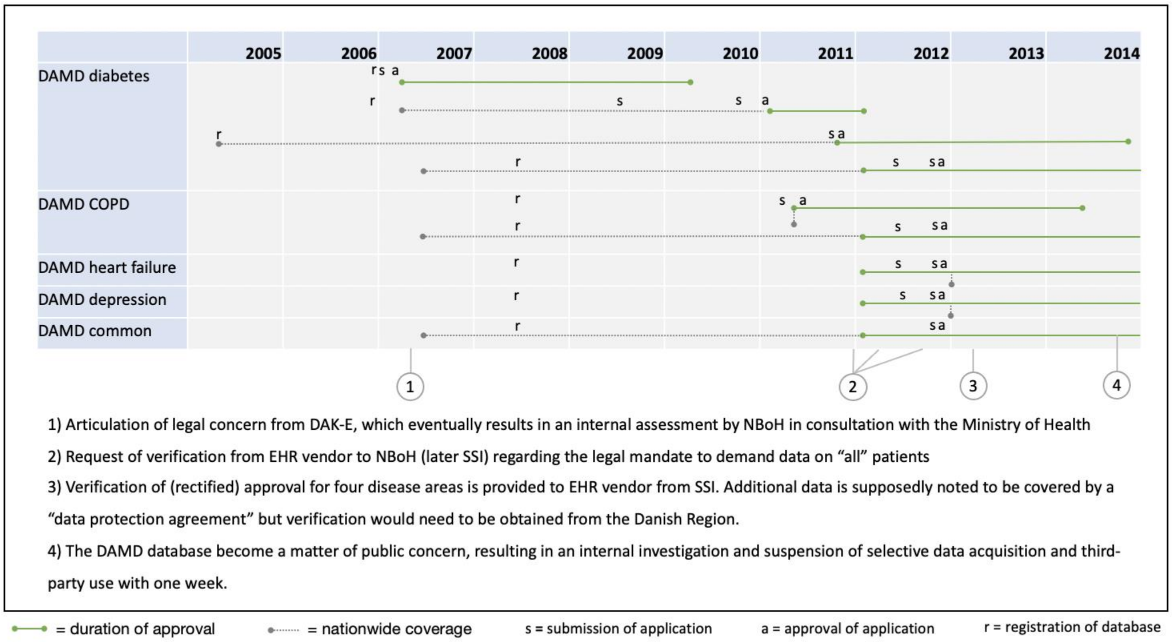
---

[13] NBoH was additionally requested to confirm whether the form of data needed to be personally-identifiable. In response, SSI asserted that anonymity 'would not be a possibility' (SSI, 2014a: 243) since civil registration numbers on citizens are traditionally used to pool data across registries (see also Author 1 and Author 2).

[14] One in May, including all stakeholders, and another in August, adopting a more forceful choice of words (SSI, 2014a).

[15] This is factually incorrect since a one-year approval was issued for the COPD sub-database in July 2011 with effect from May 2011 (SSI, 2014a: 86).

common database and COPD databases. So, what does this demonstrate?



1) Articulation of legal concern from DAK-E, which eventually results in an internal assessment by NBoH in consultation with the Ministry of Health

2) Request of verification from EHR vendor to NBoH (later SSI) regarding the legal mandate to demand data on "all" patients

3) Verification of (rectified) approval for four disease areas is provided to EHR vendor from SSI. Additional data is supposedly noted to be covered by a "data protection agreement" but verification would need to be obtained from the Danish Region.

4) The DAMD database become a matter of public concern, resulting in an internal investigation and suspension of selective data acquisition and third-party use with one week.

Figure 3. Author's illustration of the process for the issued approvals concerning the DAMD sub-databases based on applications and approval letters disclosed in SSI's (2014b) investigation

Whereas the previous section illustrated how data can command action, this section highlights how action *and* inaction in the governance of the DAMD is equally important to explain its excessive growth. With the technology in place, data from general practice had become coveted by central authorities since it existed in a *readily available form* that for the first time made *centralisation* a technological possibility. To this end, a significant will was exhibited through action and inaction to keep the data acquisition safe from the threat of discontinuation and deletion that is a hallmark feature of digital hoarding. The DAMD was riddled with legal and administrative irregularities that even predated its implementation in May 2007; Dispensations were granted from the legal requirement of national coverage for diabetes, and the mode of data acquisition from general practice was optional when it should have been obligatory (i.e. authorised diabetes data during the first collective agreement) and obligatory when it should have been absent (i.e. unauthorised data in general). All this was evidenced in the governance process, either in the episode of 2007 or that of 2012. Yet, decisive action to delimit the scope of data acquisition - either temporarily or permanently - remained absent until the evidence was placed under the light of public scrutiny.[16]

In his key study on the promotion of 'data-driven' health services in Denmark, Hoeyer observes how 'the policies and plans themselves use existing data and evidence in a very *selective* manner' (2019: 1, emphasis added). The DAMD case seems to suggest another instance of selective use of data and evidence in the promotion of data-driven promises. Extending Hoeyer's observations, the material manifestation of this selectivity is in this case not limited to *plans* and *policies*; it pertains to the *practices* too. These practices are specifically manifested in the administration of what should be acquired and discarded in the DAMD.

Concluding remarks

In this article, I have put forward a hoarding lens to explore the following conundrum: How could personally identifiable data on three-quarters of the Danish population end up in the DAMD when its approval was limited to four chronic conditions? With the case laid out, I argue that not only did actors fail to adequately recognise the problematics of the DAMD, they exhibited a significant will - through action and inaction - to ensure the longevity of what in hindsight should have been negotiated. The story I have presented is one of distributed accountability involving the collective (in)action of multiple stakeholders. In the implementation process, system designers, contractual negotiations and to an extent users contributed to the database 'hoard' as they sought to achieve a certain form. In the governance process, central authorities exhibited action and inaction to keep the data acquisition safe from the threat of discontinuation and deletion and thereby institutionalise the centralisation of general practice data. With this argument, I am not suggesting that the actors acted out of malice. Indeed, I have repeatedly highlighted how the action of various stakeholders can be traced to different value articulations of importance to public health. It is furthermore important to contextualise my findings within the brief history of institutionalisation of quality development. Contrary to what I have presented, the DAMD was preceded by several cases where the legal framework was adapted to existing data practices - rather than vice versa - in a relatively unproblematic fashion (Hartlev and Wadmann, 2018). What

---

[16] The regional authorities were by their own admission not approached by SSI (and its predecessor) before October 2014 when notified about the temporary suspension (i.e. scope delimitation) of the data acquisition. This account also states that Region Southern Denmark, the data controller, was not approached by DAK-E concerning 'problematics *that required a decision* regarding the change of the operation of DAMD' (SSI, 2014b: 300, author's emphasis).

the case study suggests is that the continuation of this tradition of legislation requires a particular practice of public reasoning and acceptance (Jasanoff, 2012) when the case is a matter of public concern (see also Author 1 and Author 2).

Analytically, I have explored what it might entail to take the ‘call of data’ seriously. In critical data studies, data are often described as if they possess a capacity or power to captivate attention in ways that sometimes overshadow the broader politics of data. While it is common practice to highlight the ‘enchantment’ (Borgman, 2015: 222), ‘allure’ (Trish, 2018: 34) or ‘seduction’ (Merry, 2016) of data, studies rarely take time to dwell on data’s aesthetic appeal and material agency. In the present case study, I have attempted to accommodate this gap by paying attention to the capacity of data to ‘call out’ and command action (Walford, 2021) as well as inaction. At times, it was almost as if the DAMD was seen to have ‘its own momentum or drive to persist and grow’ (Bennett, 2012: 252), damned with a curse of being ‘endlessly hungry’ (Slota et al., 2020: 4). This came at a profound cost, as what began as a shared desire for more data eventually resulted in ‘damnation’ for all. With the destruction of the database came the downfall of an otherwise important infrastructure for chronic care. The loss was broadly bemoaned (Wadmann and Hoeyer, 2018), with OECD (2016, 12) calling it a “major step backwards” for primary care in Denmark. Consequently, investments have since been funnelled to rebuild the data infrastructure. While some things can be replaced, not all can be erased. Even after the DAMD was destroyed to protect patient confidentiality and public trust, the case was in the introductory interview with Jensen still considered a ‘sensitive area’, as experienced in my fieldwork for years to come. The cost of socially unsustainable data initiatives is in this respect not necessarily limited to the immediate loss; they can also lead to a ripple effect in the health data ecosystem (Sharon and Lucivero, 2019) that in turn shapes the conditions of possibility for future data initiatives.

Part of what the present case engages can also be narrated with ‘function creep’ as an empirical descriptor. In this concluding paragraph, I take a close look at this notion to broaden the case study contribution and future possibilities of the analytic I propose. The concept of function creep is widely used in surveillance studies and is descriptive of phenomena when the use of technology acquires additional functions or is otherwise modified (Dahl and Sætnan, 2009). In a case study of a forensic DNA database in Norway, surveillance scholars have critically observed that information databases as a technology are especially susceptible to function creep since their use is ‘almost limitlessly interpretatively flexible’ (Dahl and Sætnan, 2009:100). This is suggestive of the contemporary relevance of the notion as an extension of the critical observation that ‘Once a technology is in place, it becomes *wasteful* not to use it to the fullest *acceptable* limit’ (Dahl and Sætnan, 2009: 89, author’s emphasis). Seen from a strict perspective of public health, one could argue with legitimacy that the state has a moral obligation to mitigate ‘waste’ given the technoscientific development and the stakes involved. If legitimacy is to be achieved, efforts must however navigate what constitutes the ‘fullest *acceptable* limit’ in consideration of legal *and* social interests. Overstepping this limit was proved a costly mistake in the present case study on data acquisition - as well as others on its reuse (Langhoff et al., 2018; Wadmann and Hoeyer, 2018) and retention (Author 1 and Author 2, forthcoming) - that demonstrates how efforts to exploit 'wasted' opportunities (OECD, 2019: 3) can result in waste themselves regardless of intentions. As a final remark, I propose that the call of data could be a useful analytic for engaging data that could for instance complement the lens of function creep - by bringing the role of aesthetics and materiality to fore - in the analysis of *how* change in the use of technology comes about.

References

Amoore L (2016) Cloud geographies: Computing, data, sovereignty. *Progress in Human Geography* 42(1): 4-24.

Author 1 and Author 2 (forthcoming) Refusing Retention: Health Data and its Deletion in the Danish Welfare State. *Social Sciences Studies.*

Bauer S (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.

van Bennekom MJ et al. (2015) A case of digital hoarding. *The BMJ.* Available at: doi:10.1136/bcr-2015-210814.

Bennett J (2012) Powers of the hoard: Further Notes on Material Agency. In: Cohen JJ (ed) *Animal, Vegetable, Mineral: Ethics and Objects.* Punctum Books, pp.237-269.

Borgman C (2015) *Big Data, Little Data, No Data.* Cambridge/London: MIT Press.

Dahl JY and Sætnan AR (2009) ''It all happened so slowly'': On controlling function creep in forensic DNA databases. *International Journal of Law, Crime and Justice.* 37: 83-103.

DAK-E (2008) Vejledning i brug af Sentinel Datafangst: Vejledning til Sentinel version 3.0. DAK-E.

DAK-E (2009) Dansk AlmenMedicinsk Database: DAMD Årsberetning 2008.

DAK-E (2014) Dansk AlmenMedicinsk Database: DAMD Årsberetning 2013.

Dourish P (2014) No SQL: The Shifting Materialities of Database Technology. *Computational Culture.* (4). Available at: http://computationalculture.net/no-sql-the-shifting-materialities-of-database-technology/ (accessed 12 august, 2021).

Forde I, Nader C, Dietrich KS et al. (2016) Primary Care Review of Denmark. Report, OECD.

Halpern O (2015) *Beautiful Data*: *A History of Vision and Reason since 1945.* Duke University Press.

Hartlev M and Wadmann S (2018) Sundhedsdata og kvalitetsudvikling – et retligt kludetæppe. *Juristen (4): 116-128.*

Hoeyer K (2019) Data as promise: Reconfiguring Danish public health through personalized medicine. *Social Studies of Science* 19(1): 1-25.

Hurlbut B (2015) Remembering the Future: Science, Law, and the Legacy of Asilomar. In: Jasanoff S and Sang-Hyun K (Eds) *Dreamscapes of Modernity.* The University of Chicago Press, pp.126-151.

Jasanoff S (2012) Science and Public Reason. *Routledge.*

Langhoff TO, Amstrup MH, Morck P, and Bjorn P (2016) Infrastructures for Healthcare: From Synergy to Reverse Synergy. *Health Informatics Journal*, (24)1: 1-11.

OECD (2019) Health in the 21st Century: Putting Data to Work for Stronger Health Systems.

In: *OECD Health Policy Studies*. Paris: OECD Publishing.

Online Etymology Dictionary. (n.d.) Hoard (n.). *Online Etymology Dictionary.* Available at: https://www.etymonline.com/word/hoard (accessed 6 April 2020).

Maguire J and Winthereik BR (2021) Digitalizing the State: Data Centres and the Power of Exchange. *Etnos 86(2): 530-551.*

Merry SE (2016) *The Seductions of Quantification: Measuring Human Rights, Gender Violence, and Sex Trafficking.* The University of Chicago Press.

Schroll H and Friborg S (n.d.) Overvejelser vedrørende DAMD debatten. Available at: https://www.dsam.dk/files/255/data_150408_overvejelser_vedroerende_dadm_debatten_hs_sf_final.pdf (accessed 14 January 2021).

Schüll ND (2018) Digital containment and its discontents. *History and Anthropology*, (29)1: 42-48.

Sharon T and Lucivero F (2019) Introduction to the Special Theme: The expansion of the health data ecosystem – Rethinking data ethics and governance. *Big Data & Society*: 1-5.

Statens Serum Institut [SSI] (2014a) Bilag til udredning vedr. ansøgninger om godkendelse af DAMD som klinisk kvalitetsdatabase, fra 2007 og frem. Available at: https://www.ssi.dk/-/media/arkiv/indhold/dk-dansk/aktuelt/nyheder/bilag-til-udredning-vedr-ansgninger-om-godkendelse-af-damd-som-klinisk-kvalitetsdatabase-fra-2007-og-frem.pdf?la=da (accessed 26 October 2020).

Statens Serum Institut [SSI] (2014b) Udredning vedr. ansøgninger om godkendelse af DAMD som klinisk kvalitetsdatabase, fra 2007 og frem. Available at: https://www.ssi.dk/-/media/arkiv/indhold/dk-dansk/aktuelt/nyheder/udredning-vedr-ansoegninger-om-damd-26nov2014.pdf?la=da (accessed 26 October 2020).

Steketee G & Bratiotis C (2020) *Hoarding: What Everyone Needs to Know®*. Oxford University Press.

The Danish Health Data Authority [HDA] (2017) *Strategi 2020: Helhed – Sammenhæng – Tryghed.*

The National Board of Health [NBoH] (2005) *Kronisk sygdom: Patient, sundhedsvæsen og samfund.* pp.1-106.

Thylstrup N (2019) Data out of place: Toxic traces and the politics of recycling. *Big Data & Society:* 1-9.

Trish B (2018) Big Data under Obama and Trump: The Data-Fueled U.S. Presidency. *Politics and Governance.* 6(4): 29-38.

Tsutsui (1996) W. Edwards Deming and the Origins of Quality Control in Japan. *The Journal of Japanese Studies.* 22(2): 295-325.

Tupasela A, Snell K and Tarkkala H (2020) The Nordic data imaginary. *Big Data & Society* 7(1): 1–13.

Tynell J and Fischer A (2014) Fortrolige sundhedsdata samles ulovligt i database. *Danish Broadcasting Corporation*, 29, September.

Vayena E and Gasser U (2016) Strictly biomedical? Sketching the ethics of the big data ecosystem in biomedicine. In Mittelstadt BD and Floridi L (eds) *The ethics of biomedical big data*. Springer, Cham, pp. 17-39.

Wadmann S, and Hoeyer K (2018) Dangers of the Digital Fit: Rethinking Seamlessness and Social Sustainability in Data-Intensive Healthcare. *Big Data & Society* 5(1): 1–13.

Walford A (2021) Data Aesthetics. In: Carroll T, Walford A and Walton (Eds) *Lineages and Advancements in Material Culture Studies.* London/New York: Routlegde, pp. 205-217.

Walton M (1988) *The Deming Management Method: The Bestselling Classic for Quality Management!*. Penguin.

Wilkinson et al. (2016) The FAIR Guiding Principles for scientific data management and stewardship. *Scientific data*, 3(1): 1-9.

Refusing retention: Health data and its deletion in the Danish welfare state

John Mark Burnett and **Rachel Douglas-Jones**

Abstract

This article explores data deletion as an analytical challenge for Science and Technology Studies. We ask how data becomes valued, not to be kept but because it should be erased. We follow a controversy, widely discussed in 2014-15, about medical data collected and held by the institutions of the Danish state. In contrast with well-established narratives of the value of data and its potential for producing future insight, the case we review demonstrates a moment in which data retention was challenged and variously delimited and defined. Following the controversy chronologically, our analysis documents uncertainties that arose when the deletion of data was requested. We first highlight the challenges of jurisdiction over deletion in public institutions before going on to examine the technicalities of practices of deletion and to explore the (im)possibility of certainty that it has been achieved. Through the case, we illuminate the political difficulties of deletion as a sociotechnical practice and as an emergent site of negotiation between states and citizens. We aim to develop an analytic for deletion that challenges the positivity of data retention as a default mode for valuing data, inviting STS scholars to attend to the close relationship between data deletion and data retention.

Keywords

health data, deletion, Denmark, state, valuation

In this article we explore the challenges in attempting to delete data. We tell our story through a data collection scandal that has reverberated in the Danish health technology scene for a number of years, shaping efforts to capitalize on the data-rich resources of the nation state (Bauer, 2014; Tupasela et al., 2020). To focus on deletion puts a longstanding promise of data accumulation in the spotlight; demands for deletion can challenge a largely inchoate positive stance around collection. We argue that STS scholars should attend to the sociotechnical challenges of deletion and the institutional complexities arising around challenges to retention. In our account, deletion is an achievement, and negotiations over the retention of data make evident its partibility and mutability: What data is depends on what it can be used to do.

The controversy through which we work took place between 2014 and 2015 and became known as the 'DAMD scandal'. DAMD is an acronym for the 'Danish General Practice Database' [*Dansk Almen Medicinsk Database*], phased into the Danish healthcare system in 2006. It's stated purpose was to collect data to improve general practice and reduce the burden of chronic illness through improved treatment (see Lippert et al., 2017). Eight years later, however, the database made the national headlines for all the wrong reasons: 'Confidential health data is illegally collected in a database*'* was, in late September 2014, the title of the first news article of what would quickly become a media storm (Tynell and Fischer, 2014a). The burning question that the authorities faced was, why had a national database for select chronic disease-relevant personal health data (e.g. medical diagnoses, lab results and medical prescriptions) expanded to cover three-quarters of the population. How could the data collection from general practices take place without the explicit consent of patients? What began as a database for quality development had now become the centre of a public controversy, which took in protests from civil society organizations and members of the professional community of general practitioners. The political conundrum for public institutions was what to do with the dubiously collected, likely illegal and yet potentially invaluable, 'unique' data. More importantly, who was to decide?

Why study deletion?

While this paper gives an account of how a specific database was fought over and how deletion was conceptualized within the long path to political resolution, some data can be deleted easily and quietly.[1] Skinner and Wienroth (2019) recount the 'wholesale destruction of [UK DNA] samples [and] 1,766,000 computer records', in a story spanning 'collection, curation and "ending" in an era of biobanking, digitalization and surveillance' (p. 2). Their interest was attracted by the deletion itself which, according to them, 'commanded relatively little public attention and took place without fanfare' (p. 2). The inarticulate potential in data is one starting point for why the 'destruction' of this data seemed remarkable. Beer writes that 'the power of Big Data is not just in the data themselves, it is in how those data and their *potential* is imagined and envisioned', and how the notion 'recraft[s] notions of value and worth' (Beer, 2016: 9). This sense of potential is about the relations data may yet come into: As Leonelli puts it, upon recontextualization, data is 'potential evidence', evolving and changing across contexts (Leonelli, 2016: 78, see also Buckland 1999). In Skinner and Wienroth's case, 'forward-looking imaginaries' were both a reason to keep the DNA archive and to delete it. Discussing forward looking imaginaries in the biosciences, anthropologists Taussig et al. observe the etymological roots of potential in Latin's *potens/potence*, connoting 'the powerful, possible and capable' (Onions, 1966, cited in Taussig et al., 2013: S5-6). Much of this careful unpicking of potential in the biomedical domain, from embryos to stem cells, (Taussig et al., 2013) holds for the way data is also held *in potentia*.

Despite this sense of inchoate potential, keeping 'everything' has both physical and financial consequences – as archivists familiar with both possibility and incompleteness could have told database managers (Bowker, 2005; Derrida, 1996). A site of decision-making, the act of deletion involves the 'making' explicit of norms around what should be remembered, and what forgotten. Since attention in privacy and information studies to increases in retention of data, scholars have attended to voluntary and involuntary information loss, studying the character of organizational memory (Foroughi et al., 2020) alongside strategic forgetting (de Holan and Phillips, 2004). With data, the question of 'who gets to retain or discard it' is at stake (Blanchette and Johnson, 2002: 34). After two decades of what Blanchette and Johnson term as an 'erosion of social forgetfulness' (p. 43), the lens of value has begun to turn towards the importance of forgetting (Meyer Schönberger, 2011).

Critical questions about the purpose and use of data collection have long been at the core of feminist critiques of archival theory and its intersection with information management (Cifor and Wood, 2017; Dever, 2014). These questions have concerned the imaginative as much as the technical. For Caswell (2014: 55-56), 'archival imaginaries' provide 'affective counterbalances and sometimes resistance to dominant legal, bureaucratic, historical and forensic notions of evidence that so often fall short in explaining the capacity of records and archives to motivate, inspire, anger and traumatize'. Gilliland and Caswell (2016: 56) argue that imagined records can function in ways similar to actual records, 'because of their aspirational nature (i.e. because an individual or community wants it to exist, or wills it into an imagined existence)'. For information managers and those who study their practices, deletion sits alongside other everyday, mundane activities that data workers conduct (Walford, 2021): the cleaning, curation and selection that produces data and its 'shadows' (Leonelli et al., 2016). However, the sheer labour involved in deciding what to keep, how to keep it and what to delete has spawned literature on practices of capture and storage, particularly concerning government data (Lappin et al., 2018). From mechanisms to 'automatically appraise, select, preserve or permanently delete data based on their "value"' (van Bussel et al., 2015: 187) to sociotechnical analyses of how such automated systems 'build a value for nothing (the deleted)' (Neyland, 2018: 22), deletion, in its various guises, is a lively topic in fields of policy and practice.

The DNA samples Skinner and Wienroth reported on would have cost the UK's Home Office 'ninety pence per annum to store' (2019: 22). This specificity puts a price on the 'memory epoch' within which the organization was working (Bowker, 2005).[2] Yet, in an analysis that centers the financial value of data within a framework of accumulation, Sadowski argues that 'for the increasing number of companies participating in the "data economy" or "digital economy", deleting data because of storage costs would be like burning piles of money or dumping barrels of oil down the drain because renting a warehouse was too much trouble' (Sadowski, 2019: 1). These warehouses for data do exist, contrary to the language that puts information storage out of sight and into 'clouds' (Hogan and Vonderau, 2019; Jørgensen and Gudmundson, 2019), seemingly out of the material realm.[3] Work in STS has repeatedly returned the physical locations of data into our view by attending to data centres and internet cables (Maguire and Winthereik, 2019; Starosielski, 2015). The digital is physical, somewhere. Yet, with the ease of storage and a growing imperative to collect and keep data for yet unspecified purposes, deletion (along with maintenance, durability and obsolescence) has taken the work of attempting to realize the unknown potentials in 'data' backstage.

Conceptualizing deletion

Regimes of data retention and deletion don't operate in isolation. Our approach here is inspired by growing analyses of how data is conceptualized, *in and with* the 'long history of the accumulation of data about individuals and populations' (Beer, 2016: 4), cognizant of the relationships that data have with histories of statistics, counting people and populations (Cline Cohen, 1999 [1982] ; Douglas-Jones et al., 2021; Hacking, 1991; Ruppert et al., 2018). In our analysis, what stands to be deleted depends on what the data is thought to be, something that, as we show, does not remain stable. We approach data deletion through a series of refusals, both refusals to accept retention and refusals to delete. We borrow from Benjamin's neologism 'informed refusal': Articulated as a negative corollary of the more familiar informed *consent,* Benjamin notes that a refusal to participate (in her case, in biomedical research), 'is seeded with a vision of what can and should be and not only a critique of what is' (Benjamin, 2016: 4). Thought through refusal (to accept or allow that data be retained) deletion can thus be recognized as creative as much as it is critical. The problem of deletion goes well beyond a moral dilemma of making 'ethical choices' (Floridi and Taddeo, 2016) about what to retain or a question of making and maintaining legal boundaries around that which may or may not be collected (Bachmann et al., 2017; Pfitzman and Hansen, 2010). It is also an enterprise inseparable from the social and political institutions and infrastructures that make data collection possible in the first place (Bowker, 2005; Jasanoff, 2004). Since our focus is on state data collection in Denmark, readers should understand that the collection of data in this state has longstanding politics oriented towards an explicit public good (Bauer, 2014; Taylor, 2016). Thus, refusal already stands against something particular: an environment where the potential of biomedical data has been articulated and largely accepted (Hoeyer, 2019). Both those who wish to see data deleted, those who want to see it retained ask what acceptable deletion looks like and what creative form it might take.

The empirical data was gathered between 2017 and 2020 as part of a doctoral study about the data-intensification of healthcare in Denmark. Initially, the aim was to conduct an ethnographic case study of infrastructure in the making to explore emerging practices and politics of big data. Instead, the collapsed infrastructure of the DAMD became the dominant case study. This change was primarily based on a preliminary field mapping of the Danish biomedical sector. Despite its destruction two years prior, the DAMD reappeared at various sites as an ethnographically rich object to study the tensions of data-intensification in Denmark. The preliminary field mapping consisted of an interview with the head of the Danish Health Data Authority,[4] participatory observations at various gatherings and online research of various sources (news media, professional journals and social media). Consequently, the research design was revised to incorporate the DAMD. An interview was conducted with the former head of the Danish Quality Unit of General Practice, who has been a central figure throughout the development of the DAMD.

Participatory observations were carried out at additional gatherings where the DAMD made an – sometimes scheduled and other times unexpected – appearance. The online research continued in a more targeted fashion. In terms of social media, the article draws on a Twitter dataset retrieved via a case-specific query[5] in the summer of 2017. Once cleaned, the dataset resulted in 1363 tweets, dating back to late September 2014.

Mandating deletion

Over a period of nine months, between September 2014 and May 2015, state institutions jostled to claim their stake in the DAMD database. There were five key actors: First was the data controller of the DAMD database, Region Southern Denmark (RSDK). RSDK was invested in the database for the purpose of quality development in healthcare, but not at all costs. Second were the government Ministries for Health, Justice and Culture. Each was forced to respond to the DAMD database in order to clarify both its legal status and possibilities for its retention. Then we have the National Archives (DKNA). It asserted its interest in the database by classifying it as 'worthy of preservation'. Last was the Danish Parliament. Here, the ultimate fate of the database was decided. As the story unfolds, we see what kinds of arguments and interests were put forward to prompt decisions about deletion and how these moves differentially assigned value to data. As such, the analysis makes a specific kind of contribution, extending arguments about individual considerations of (potential) use, values and norms in data sharing and privacy decisions (Bowser et al., 2017) to show how similar considerations are debated at institutional, governmental and societal levels.

From 'clinical data' to 'personal data'

The day after the national news media reported that the DAMD database contained illegally collected health data, Statens Serum Institut, a century-old health research institute, initiated an (internal) investigation under the auspices of the Ministry of Health. Nearly a month later, the investigation concluded that only four disease areas could be legitimately authorized for a 'clinical quality database' in accordance with the Danish Healthcare Act. These four areas were a fraction of the 704 diagnoses stored in the DAMD database (Statens Serum Institut, 2014b).[6] The following day, Region Southern Denmark (RSDK), the data controller, declared that it would delete the unauthorized[7] collected data in one week, beginning on 4 December 2014. In response, the Minister of Health cautioned that deleting the data would be like 'throwing the baby out with the bathwater' as the unauthorized data could still be of 'significant importance for our treatment options in Denmark' (Fischer, 2014a). The gathering in one place of a considerable amount of data on the activities of general practice was a remarkable achievement even for Denmark, which had had years of investment to address an 'information void' in primary care (Forde et at., 2016: 33).[8] The Minister of Health continued by stressing that, although the broad data collection could not be classified as 'clinical', it could still be *legal*, since it had not yet formally been considered through the Danish Act on Processing of Personal Data[9] by the Data Protection Authority. Now, the mandate to decide the fate of the broad collection of data was passed forward to the Ministry of Justice, which governed the legal jurisdiction for 'personal data'.

From 'personal data' to 'archive material'

With an official assessment from the Danish Data Protection Agency due, the database looked destined for obliteration on the eve of the declared deletion deadline. This was until the Danish National Archives (DKNA) made a move to classify the DAMD database as 'preservation-worthy' on grounds that the database would provide 'insight into the activity of the health system ..., which will be relevant for historical use and scientific use for many years to come' (Fischer, 2014b). As a result of being 'preservation-worthy', the chief consultant in DKNA concluded that RSDK would be left with little choice other than to submit *a full copy* of the database to DKNA, which was reported to apply for citizens, too: 'They *must* provide the information. In the case of preservation-worthy records, no one can oppose the transfer of data to public archives in accordance with the Archive Act.' (Fischer, 2014b, original emphasis). While the intervention from DKNA brought the deletion to halt, it also gave impetus to various acts of refusal: An online petition was launched, a patient advocacy group convened, and these were soon joined by strong statements from the Danish Medical Association and the Danish College of General Practitioners (Heissel, 2015).[10] On December 11, the Danish Data Protection Agency confirmed what RSDK had suspected: The broad data collection exceeded the legal confines of the Danish Act on Processing of Personal Data (National Archives, 2015b: 2). One-and-a-half weeks later, RSDK reiterated its stance against archiving the illegal data, insisting that the data would 'never be able to be used legitimately' (Region Southern Denmark, 2014: para 5). At the same time, the Danish Organization of General Practitioners (PLO) demanded action from the responsible minister to safeguard patient confidentiality by ensuring that the database is deleted (Tynell and Fischer, 2014b). With the database now considered potential archival material, the ministerial mantle of responsibility had shifted yet again, from the jurisdiction of the Ministry of Justice to that of the Ministry of Culture.

Worthy of preservation?

The response from the Minister of Culture (MoC) was swift and firm. The absence of proportionality between the approved and actual collection of data was unequivocally noted as 'entirely unacceptable' and unwelcome in a law-abiding Danish society (Nielsen and Tynell, 2014). 'If necessary', the Minister continued, 'then I am ready to make the rule that stops the large illegal collection and archiving of data.' However, nearly two months later, in mid-February, the MoC changed her stance. Now, the preservation of sensitive personal information *was* the public interest *due to its significance for democratic accountability* – despite its illegal acquisition (Boysen, 2015a). In a press release, the MoC stated that the material had 'historical value' and should be archived precisely because of illegal action on the part of authorities. Now, it was not the health of the nation at a given time that warranted archiving, but the actions of authorities (Ministry of Culture, 2015a):

> If illegally collected material were to be deleted, it would be impossible to scrutinize the authorities [*kigge myndighederne i kortene*] once they have committed illegalities. You have to be able to do that in a constitutional state.

Contrary to her early statements, the MoC added that it should not be the prerogative of a minister to overrule the 'impartial authority' of national archivists by 'deciding which traces of illegalities are kept' (Boysen, 2015a). Such conduct, the Minister concluded, 'would be what happens in totalitarian states'.

One month prior to this political u-turn, the MoC had gained access to a commissioned report from DKNA that articulated what was at stake. In the report, the database was noted to hold 'substantial reuse value' and seen as 'a unique opportunity to trace disease pathways through the health sector [that] has great historical value in relation to illuminating health and illness in today's welfare society' (National Archives, 2015a: 3). This kind of historical value was constructed as a way of looking back upon the present ('In 150 years the data from DAMD will still be a unique source to understand the welfare system of our time') as well as supporting future potential research techniques ('In that time it will be possible to link information to other registers and get an enhanced understanding'). So, while DKNA and the MoC both argued that a copy of the database should be preserved due to its 'historical value', it is noteworthy that they did so for different reasons. Whereas the public reasoning of the MoC foregrounds the *historical* value of democratic accountability, the value articulated by DKNA emphasizes its 'unique' potential for health-related knowledge production.

Meanwhile, the head of the department for health cooperation and quality in RSDK, acting as spokesperson for the data controller, rejected both proposed arguments for classifying the database as 'preservation-worthy' (Boysen, 2015b). Challenging the public reasoning of the MoC, the spokesperson, whom we will call Ole Olsen,[11] asserted that an investigation of possible illegalities was not a sufficient cause to preserve the data: 'You do not need person-identifiable data for that'. Furthermore, RSDK disputed DKNA's argument that DAMD was 'unique', a uniqueness upon which the archivists' classification depended (National Archives, 2015b).

> We can procure the information prospectively and get an impression of today's health condition in Denmark … the price in our optic is too high to archive illegally collected information. It exists legally elsewhere. (Boysen, 2015b)

The legal existence of the data 'elsewhere' refers to, among other places, twelve electronic health record systems from which most data was acquired. What was remarkable, then, was not so much the datafication of activities in general practice,[12] but the centralization of siloed data from general practice that, for technical and cultural historical reasons, was a hinterland to the Danish health data landscape before DAMD. Uncertainty shifted to the reconstruction or prospective legitimate gathering of granular data on a population level from primary healthcare, otherwise characterized – unusually for the Danish healthcare landscape – as an 'information void' (Forde et al., 2016).

Last stop: The Danish Parliament

In order to establish an unequivocal legal mandate to archive the illegally collected data, the MoC tabled a bill in the Danish Parliament to amend the Danish Archiving Act where access to the archival material would be fully available to the public after 230 years (Ministry of Culture, 2015b). RSDK expressed its discontent: This seemed an 'unreasonable' use of authority on the Ministry of Culture's part, which ironically was what the archive material was also reasoned to record: 'Law and reason are not necessarily the same, and we believe, it is an unreasonable way to use authority, to demand an archived version, but we cannot argue against the law' (Boysen, 2015b). Regardless, the fate of DAMD was now to be decided by the Parliament, where the bill needed to be processed.

The debate on the first reading of the bill lasted an intense two hours (Danish Parliament, 2015a). In acknowledgement of the absence of parliamentary support, the MoC was forced to amend the bill to explicitly prohibit the preservation of the DAMD database. After more than half a year of political jostling, RSDK obtained a parliamentary mandate to destroy the illegally collected data on 12 May, 2015, with the support of all votes (Danish Parliament, 2015b). The eventual rationale for this resolution was the need to safeguard patient confidentiality and trust in public institutions in acknowledgement of their vital role in the continued collection of personal health data. On 28 May, 2015, the deletion of DAMD was confirmed by RSDK (Borggreen, 2015).

Clearly, what is valuable about the data is not agreed upon, and what it *is* also shifts. As the database traverses the regulatory realms of jurisdictions, it is noteworthy how its content (in addition to its status of legality) is variously negotiated as 'clinical data', 'personal data' and 'archival data'. Following how meanings changed shows why it can be more helpful to ask '*when* are data' as opposed to '*what* are data' when tasked to pin down data ontologies (Borgman, 2015: 4-5). The political proceedings show how classificatory practices have ontological, practical and political implications for shaping the conditions of possibility for data. While DKNA and the MoC both took proactive steps of refusal to preserve the illegally collected data, they did so by highlighting its future historical value for (respectively) health-related research and accountability. RSDK, by contrast, did not accept the justification for classifying the illegally collected data preservation, insisting that the preservation would lead to detrimental effects for public health. The various trials for deciding the prefix of data engendered the competing enactments of 'public value' that respectively highlight a potential for public health, harm to data protection rights and cultural history. In the end, the parliament was invoked, as the democratic *agora* and mechanism for public legitimacy, to decide what the public value should be based on law.

Technologies of deletion

Once the majority of the data collection within DAMD was found to exceed the Danish Healthcare Act, a series of interventions within the space of data deletion gradually negotiated the database away from an imagined 'total' deletion. While technologies might bring to mind commercial hard disk crushers and data wipes, in this case we see competing efforts to shape deletion into a form of acceptable retention. We review three salient social technologies: ‘archiving’, ‘opting out’ and ‘anonymization’. Central to each is the negotiation of the database’s value relative to what it was seen to be able to offer.

Archiving

The first technology of deletion can be traced to early December 2014, when DKNA classified the DAMD database worthy of preservation. Crucially, the intervention came one day before the deletion deadline set by RSDK. Archiving was, in this case, invoked as a technology of deletion that would allow the *operational* version of the DAMD database to be destroyed, but only after an *archival* copy had been secured.[13] A question that arises is whether it makes sense to view archiving as a form of deletion, if this act presupposes the retention of a copy that effectively saves what would (and perhaps should) otherwise be lost? Contrary to The Danish College of General Practitioners who rejected the proposition as ‘really weird’ (Boysen, 2014),[14] the answer noted in the investigatory report from DKNA was ‘yes’: ‘Archiving of data is … in the context of data protection *equated with deletion*. Data is removed from administrative use’ (2015b: 5, emphasis added). How this equation – archiving as deletion – holds is largely to do with access: The data is considered deleted if it is unavailable. However, the question becomes to whom it is unavailable and over what time periods. Following this empirical definition of archiving as a technology of deletion, we now turn to the regimes of *access* governing the DAMD database[15] to understand how archiving and deletion came to be equated here.

Once personal archive material enters the custody of DKNA, it becomes governed by the regime of access control defined by the Danish Archives Act §23 (2007). Ordinarily, access would be blocked for 75 years. When applied in relation to deletion, expiration dates are more commonly associated with principles of storage limitation, to ensure that personal data is not kept for a longer period than necessary. DKNA’s definition of archiving in this case, however, inverts the applications of expiration dates in that it is deletion, not retention, that is temporary and finite. However, a chief consultant in DKNA also made it known publicly that, following the Danish Archives Act, ‘anyone can apply for access to not immediately accessible records’ (Larsen, 2015). This dispensation not only provided a window for continued research access, it also opened space for uncertainty regarding the efficacy of archiving as a deletion technology (e.g. Boye, 2014; Ritzau, 2014). When the MoC eventually tabled her bill in the Danish Parliament, she made use of an exception in the Danish Archives Act to tighten the regime of access, increasing the blocked period from the usual 75 years to 120 years. Furthermore, the materials would be held *without* possibility of third-party dispensatory access. After 120 years, third-party access would be permitted in an *anonymous* format for 110 further years. Full access to the archive material would then be blocked for 230 years in total, three times longer than the 75 years reported in the beginning. As both the periods of embargo and restrictions on access grew, the rationale given was, in part, to safeguard the privacy and integrity of data registered illegally (Ministry of Culture, 2015b) and also to convince the public that ‘it is not possible to cheat your way to person-identifiable information’ (Larsen, 2015), cheating here referring to an unlawful exemption of explicit consent.

Returning to archiving as a form of deletion, taking the operational version out of ‘administrative use’ would implement corrective measures to ensure legal compliance regarding the use of personal health data in the Danish Healthcare system. What was said during patient encounters in the general practice would no longer be accessible to 100,000 health professionals or public administrators, thereby mitigating the concern of breached patient confidentiality and individual control. The proposal to remove the possibility for dispensation would additionally mitigate immediate concerns regarding the possible reuse of data for purposes that invoked a contentious version of the ‘common good’. One of such purposes was commercialization, which is a common cause of scepticism in Europe (Skovgaard et al., 2019). Commercialization was registered as a cause of anxiety in this particular political climate, where health data was valued as a means to achieve economic growth (Wadmann and Hoeyer, 2018).

Opt-out rights

A second technology of deletion in the case appeared in December 2014, when RSDK was informed of its duty to process *opt-out requests* from citizens. In Denmark, citizens generally do not have the choice to opt out when it comes to participation in registry-based research (Nordfalk and Hoeyer, 2020). However, the illegally collected data could not be classified within the legal regime of ‘clinical data’ that supersedes individual rights enshrined in the Danish Act on Processing of Personal Data (2000), such as the right to opt-out. Exercising this right would make it possible for citizens to have their data erased from the *operational* version of the DAMD database – and therefore be excluded from any future *archived* version. According to DKNA (2015b: 2), RSDK was initially ‘positive’ about the prospect of preserving a digital copy of the DAMD database in DKNA until notified of its duty to process opt-out requests. A couple of weeks after the initial archive request, RSDK published a press release in which the council chairman in RSDK said that the requirements from DKNA simply did ‘not make sense’ (Region Southern Denmark, 2014):

> They are bizarre and contradictory. We are already against delivering anything to the National Archives [DKNA] that will never be able to be used legitimately. If we then additionally are handed the obligation to delete information from citizens that do not wish to be included in the archive version, then it makes no sense at all. If the archive version was useless before, then it would become entirely worthless if we begin to fiddle with it.

One of the reasons why DAMD was so highly valued was precisely because it was considered to contain data that was considered representative of the ‘entire population’ of Denmark from 2011 and onwards (National Archives, 2015a: 3). This representativeness was largely due to the collective agreement of 2010 between Danish Organization of General Practitioners (PLO) and regional government (Danish Regions), where data-intensification was a high priority (see Wadmann and Hoeyer, 2018). From a research perspective, the subtraction of individuals on the basis of opt-out requests could potentially introduce a ‘bias’ that would ‘fiddle with’ or compromise the achievement of population representativeness desired for clinical quality databases. Here, the value of rights on an individual level (opting out) and participation on a population level (being in a representative dataset) are conflicted.

On 27 March 2015, two days after the proposed bill, citizens encountered a similar trade-off between individual and public interests when DKNA exercised its authority (*arkiveringsbestemmelse* in Danish) to prevent RSDK from processing opt-out requests, beginning at midnight the same day (Flyvbjerg and Larsen, 2015). Once an archive version is made, it enters DKNA, where the individual right to opt-out or be forgotten is superseded by the collective interest vested in the *preservation* of the records as a ‘whole’. Aside from its coverage in the national newspapers, the imposed deadline for opting out caused significant consternation and public discussion. Within the Twitter dataset collected to study the moment,[16] the most retweeted tweet (n=83) came from a tech journalist who is publicly known for his appearances on the Danish Broadcasting Cooperation. His tweet poses the rhetorical question of whether archiving is a fitting solution for upholding the integrity of patient confidentiality (Sonne, 2015):

> Think of everything you have told your doctor – well then. You have 6 hours left to prevent it from ending up in the National Archives #damd rsyd.dk/wm461183

Aside from the caution, the tech journalist linked to practical instructions on how to opt out provided by RSDK on their website.[17] On the same day, citizens were encouraged to visit the homepage by Ole Olsen from RSDK, who estimated that between 4000 and 6000 citizens had requested to opt out at that stage (Flyvbjerg and Larsen, 2015). Citizens then have two technologies of deletion to choose from: The default option that temporarily blocks access to personal health data at the price of preservation for the foreseeable future or the proactive option that destroys the central database without further duplication. In total, RSDK estimated that it spent two to three man-years of labour to delete 23,000 citizens from the operational version of the database that was destroyed altogether months later (Boysen, 2015c).

Anonymization

A third technology of deletion encountered in the case is 'anonymization'. Since 1995, anonymization of personal data has, in EU data protection law, entailed that the data subject is 'no longer identifiable' (Council Directive 95/46/EC and European Union, 2016). Legally, then, anonymization is premised upon the condition of 'irreversibility', meaning that it should not be possible to reverse *non*-personal data to its former *personal* state.[18] In the case of DAMD, anonymization would entail an effective decoupling of the CPR-number, which is a person-unique identification number assigned to the residents by the Danish state upon birth or immigration for administrative reasons.[19] While the CPR-number would enable direct identification, other attributes provide a less direct path to identification by means of inference, for instance, through a combination of name, age, weight and affiliated general practice. If opting out is the deletion of individual data, anonymization is, in this instance at least, deletion of population-level data *limited to person-identifiable attributes*. Anonymization was not widely discussed during the case, but its timing was potentially highly significant. We now contrast the difference between anonymized *access to* and anonymized *storage of* the archive material and unfold the stake involved for those who reject retention and deletion alike.

In the bill tabled by MoC in March 2015, anonymization was included as a safeguard to protect the 'descendants' of the data subjects recorded in the archive material once the 120 years of blocked access expired (Ministry of Culture, 2015b).[20] Unlike the a priori processing of opt-out requests, the bill proposed that the shield of anonymization would apply once the archive material had been preserved at the point of access for approved applicants. Whereas opting out was a technology to prevent the identity and data of individuals from *entering* the DKNA, anonymization was a technology to prevent the identity from *leaving* DKNA. The archive version would thus remain intact as a preserved duplication of the original without the infliction of further 'fiddling' beyond the processed opt-out requests.

The possibility of archiving an *anonymized* version, rather than moderating its availability to third parties, was considered in DKNA's report for the MoC, submitted a month prior to her U-turn to preserve the database (National Archives, 2015a). From a data protection perspective, this option would be a more forceful safeguard for descendants since it, in principle, would eradicate the risk of a leak or breach, however small this might be. Fundamentally, the same option would, however, eradicate the possibility of full archival access to personally identifiable information after 230 years as well, since anonymization is defined by 'irreversibility'. Expiration is simply irreconcilable with anonymization. According to the report from DKNA (2015a: 6), anonymizing the database before its submission to DKNA would tarnish its research value 'significantly' for historical and social science:

> [An anonymized] Database will continue to be valuable in historical and social science studies, but to a significantly reduced extent compared to the full database. Furthermore, it is estimated that data loses its value in relation to contemporary medical research if the database is anonymized before submission.

The solution that DKNA was 'prepared' (National Archives, 2015a: 7) to implement was, therefore, the alternative of anonymizing *accessed* data rather than *stored* data, which would preserve an archive version enriched with personal-identifiable attributes (as eventually proposed by the MoC):

> [DKNA] Draw attention to the fact that the archive is prepared to only make data available for users in an anonymous form for a longer period. [DKNA] Can thus carry out linkage of records relevant for the concerned health science research projects.

At stake in the distinction between anonymization *within* the archive and anonymization of data *accessed* was not only the enrichment of person-identifiable attributes but also the possibility of linking records on select individuals with other data registries. Elsewhere, the notion of assigning numerical marks to residents is considered controversial (e.g. England, USA, Australia, India and Japan), unconstitutional (e.g. Hungary and Germany) or even prohibited (e.g. Portugal) (Krogness, 2011: 89). With the introduction of the Central Population Registry in 1967, the CPR has become integral to the routine administration of the Danish welfare society in stark contrast (Krogness, 2011: 89). This is also common practice in the Nordic region more generally, where the CPR-number is registered when residents encounter the 'state' in healthcare or elsewhere, enriching the record registries in the process. Denmark is described as having a comparatively 'lenient' legal framework (Hoeyer, 2016: 79), that provides exemption rules for informed consent in regard to – but not necessarily limited to – certain kinds of registry-based research. Alternatives for linking data exist for nations that do not possess the tradition of marking residents; however, these practices are generally more cumbersome and costly at scale (Bauer, 2014; Schmidt et al., 2014). The CPR-number and legal regime of data collection is thus what differentiates Denmark from most nation states, but also part of the reason why their comprehensive health registries have been described as an 'epidemiologist's dream' and a 'goldmine' in the scientific community (Bauer, 2014).

By preserving the CPR-number, the future historical value of the archive material would – according to DKNA – be multiplied indefinitely, given the wealth of combinations by which health determinants can be explored in relation to social and economic domains of life:

> In 150 years, the data from DAMD will still be a unique source to understand the welfare system of our time: What was wrong with people, how often they encountered the physician, how much medicine was used, when did stress become a public disorder, etc ... In that time, it will be possible to link information to other registers and get an enhanced understanding of, e.g., the significance of education for physical attrition, links between disease and income, etc. The researchers will be able to ask an endless amount of interesting questions to this data material.

Implementing anonymization at the point of access was put forward as a way of protecting descendants while providing third-party use. This form of anonymization would, however, also protect Denmark's status as an 'epidemiologist's dream' or 'goldmine' within the scientific discourse by retaining a competitive edge for data linkage while enriching the national health data pool.[22] It can be questioned whether the parliamentary decision to destroy the DAMD database would be different had the MoC bill proposed anonymized *storage* rather than *access*, which would have then resulted in a compromised resolution of more data protection and less reuse value.

Degrees of retention

Each of these technologies modify the proposed deletion of DAMD with different degrees of retention. Each works as a technology of retention where the threshold between what is kept and erased is socially negotiated – both now and for the future. Indeed, retention can, in certain instances, be viewed as the primary *telos* of a deletion technology, particularly explicit with archiving, so that the 'ending' of certain data legitimizes the survival for others (Skinner and Wienroth, 2019).

As the discussion about DAMD progressed, archiving was invoked as a last resort of state action to retain what was destined for 'total' obliteration. Archiving *duplicated* the data so the original could be deleted and the *copy* could be retained. The defining condition for equating archiving with deletion was that the illegally collected data would be removed from its 'administrative use'. Although such removal would meet immediate concerns of administrative control among general practitioners, archiving would – in contrast to the pre-existing schedule of deletion – retain a *copy* for preservation that would only be inaccessible to third-party actors for a *temporary* period of 120 years. Preserving an archive version was thus an attempt to reconcile two competing versions of public value in the form of data protection and research potential. However, removing the illegally collected data from its 'administrative use' proved an insufficient proposition of deletion, as the parliament asserted that it was 'out of place' (Thylstrup, 2019) in the state altogether.

When 23,000 people successfully exercised their right to have their personal data removed from the operational database before an eventual duplication would take place for preservation, they opted out. Here, opt-out rights can be seen as 'empowering' citizens to refuse the retention of their personal data. However, while the opt-out right is enshrined in the Danish Act for Processing Personal Data, it provides no guarantees. In the DAMD case, the right to opt out was superseded by the Danish Archive Act, assigning primacy to the claimed interest of the nation over that of the individual. In other cases, the approval of opt-out requests has been assigned the caveat that the number of requests does not exceed the threshold of what the authority defines as acceptable, leaving the final say with the authority rather than data subject.[23] For those who wish to retain the database for registry research in public health, the opt-out right was considered a concern, since it could potentially undermine the achievement of population-wide representation and thus introduce bias if approved at scale. Given the alternative of 'total' deletion, opt-out rights can, on the other hand, be seen as politically desirable for those interested in data retention. Drawing public attention to the *possibility* of opting out can result in this right simultaneously functioning as a calculative device for reconfiguring the responsibility – and thereby stakes – of deletion from a default state obligation to individual choice. This way, the right to opt out can be used to contain collective issues of deletion at the level of the individual to alleviate the conditions of possibility for retention. In terms of the third technology of deletion, the demarcation of what is retained varied depending on *when* anonymization would be implemented in the proposed archiving process. From the two enacted in the case, anonymizing the database prior to its submission to DKNA would, in principle, constitute the most forceful safeguard to protect descendants. This option of anonymized storage would, however, come at the expense of the perceived research value, which was a particular concern for registry research in health and medicine, since it relied upon the CPR-number for data linkage. Since anonymization is defined by irreversibility, anonymizing storage would, however, also eradicate the option of providing public access to the full database once the 110 years of anonymization access control expires. The option that DKNA and MoC proposed was, therefore, to preserve the database enriched with person-identifiable attributes, including the CPR-number and the provision of anonymized access to third-parties instead. Anonymized access was proposed to retain the research value of the future archive and to strengthen Denmark's status in the scientific community.

Certainty of deletion

Let us pick up the story again at the point when the data controller (RSDK) publicly confirmed deletion of the DAMD database (Borggreen, 2015). While this should have been the end of the story, on the same day, an article from a national news media outlet covered the event under the headline 'Illegally collected health data is nearly deleted' (Ritzau, 2015). Our account of the DAMD case closes with the phrase 'nearly deleted' and the question of proof. Calls that reject retention seek proof that data is really gone. But how does one know that data is no longer retained? By what measures of near or far, total or partial, was the data 'nearly' deleted? Just as deletion was made relative through anonymization, opting out and archiving, certainty about deletion is an ideal state: In practice, it is not always easily achieved. It leads straight back to social questions and into the world of checking, evidence or audit (Power, 1997; Strathern 2000). Must deletion itself be observed? Supervised? Documented? If so, by whom? As Neyland (2018: 21) observes, 'it might seem that accountability could provide the means to transform nothing (the deleted) into something (proof of deletion)', but both in his case and ours, creating a 'certain accountable nothing, a notable absence' is challenging. In our opening section we explored uncertainty about which institutions held a mandate to delete; here we examine the *not knowing* that haunts a deletion process.

Time horizons

In contestations over the immediate future of data, the far future – 120 years to research access, 230 years to public access – was invoked. The Ministry of Culture intended to offer public reassurance that the Danish welfare state would not allow third-parties to 'cheat' their way to person-identifiable data access (Larsen, 2015). Segments of the public, however, remained sceptical of the claims made on behalf of the protective efficacy of using *time* as a mechanism of deletion, including internet activists (Loiborg, 2015), legal experts (Larsen and Foldberg, 2015) and not least the chairman of the Patient Data Association. The latter, also a practicing general practitioner to whom we will refer as Georg Paulsen, remarked that time horizons gave:

> Many years to change your mind. A law is a law until a new law is introduced. And you can change the law in, for example, four years. ... I can't help thinking whether the meaning of this is to gain some time. When people get tired of shouting that 'it's illegal', you go in and look at the data anyway. (Boysen, 2015a)

Throwing access to the data into the future was seen, by some, as a means of removing it from the present, but to others, this removal was insufficiently partial. The mechanism of certainty is in doubt in the quote above: A law can be changed, people 'get tired'. The data is actually still there. We should, perhaps, not be surprised by this form of retention. As Hoeyer has recently pointed out for Denmark and beyond, a 'rarely acknowledged effect' of the promise of data, he writes, 'is postponement'. He is writing about a logic of accumulation, wherein embedded in potential is the presumption that more data is better data, 'allow[ing] the prospects of future knowledge to overrule what we know already' (2019: 548). In this case, postponement is not waiting for more data, but the introduction of hardly graspable time horizons as a tactic of retention designed to tame the present media storm. At stake, in this moment, is how far in the future the future stays.

Copies

A key function of the DAMD database was to supply research material for health science. After RSDK received the parliamentary mandate to delete the illegally collected data, it became apparent that subsets of anonymized data material had been delivered to at least 50 research projects. This complicated the imagined construct of 'total' deletion that was assumed to follow from the destruction of the central database, as partial copies of the illegally collected data existed in local research environments. According to Ole Olsen from RSDK, part of the labour of deleting the illegally collected data included the dispatchment of letters and emails to 'encourage' the researchers in charge of the projects to make the 'right' ethical choice and delete the data. Ole Olsen added that RSDK were not in a position to apply further pressure beyond encouragement, since the data subsets were no longer under their authority (Boysen, 2015c):

> We are not the data controlling authority for the researchers, that is themselves, nor do we have instructional powers, but all good practice and data ethics must tell them that it must be put in order.

The continued existence of the data (albeit as copies, not in the central database) led to uncertainty around deletion as a practice. While the public case had resulted in an agreement on the illegally collected data, the execution of the political mandate for its deletion left room for uncertainty. Was it really gone? The decentralized storage of the data meant that, even after the central database was destroyed, the data itself remained, split across or 'returned' to different sites. Most of it would also continue to exist in the Electronic Health Record systems. This draws out a broader point for STS scholarship on deletion: Copies proliferate, actors must ask *where* contentious data exists, which locations count as data being 'out of place' and, returning to our earlier theme of the political mandate to delete, who gets to decide what will count as 'out of place' to the extent where the only solution is deletion (Douglas, 2001[1966]; Thylstrup 2019). While the data content from a technical standpoint may be identical, *where* it is located shapes the conditions of possibility for the enrolment of stakes that can inform what that data can become in the future, which then impacts its valuation in the present.

Since the time of DAMD, The Research Machine (*Forskermaskinen)* has been implemented by The Danish Health Data Authority (2017) to modernize the entire service delivery regarding health data access. The Research Machine has a number of in-built security features to retain control, in part, over where data may travel, somewhat like to the Welsh databank studied by Tempini and Leoneilli (2018).[24] With the new system, researchers must use a remote access solution through which they may view and analyse pseudonymized registry content approved for the project. In this way, researchers can usually only request a local copy of the anonymized results, and not the dataset that was previously delivered on a physical memory device. The vulnerability – and digital immaturity – of the previous practice was exposed in a public report from 2016 on events from the prior year (Hecklen, 2016): Two CDs from Statens Serum Institut were mistakenly delivered to a Chinese visa company instead of the national authority on Danish statistics via the national postal service. The CDs contained unencrypted personal health data (e.g. disease pathways, hospital admissions and psychological disorders) from various registers linked to more than five million CPR-numbers.

Residual traces

The original DAMD database was populated with data from general practitioners. Technically, data transfer was accomplished through a data capture model on GPs' computers, which would store structured data (excluding working notes) from the operating electronic health record system and supplementary entries in a local SQL database. The discussion of deletion took place against a backdrop that anything that had ever been transferred to DAMD *still existed* in these local SQL databases. Thus, a month after the official announcement that the DAMD dataset had been deleted, Georg Paulsen, the chairman of the Patient Data Association, devoted considerable efforts to make it known in public, professional and political circles that residual traces of the deleted data still existed locally in the infrastructure through which it had originally been gathered (Bundgaard, 2015; Kristiansen, 2015; Lundström, 2015). This sense of a residuality was captured by the Member of Danish Parliament Özlem Sara Cekic, who had taken a vocal stance against the retention of DAMD during the political proceedings. She tweeted: 'Does anybody know whether [the] #DAMD the illegal health data is deleted? #dkpol #dksund' (Cekic, 2015).

In professional journals, Georg Paulsen advised general practitioners to follow his lead and destroy the local SQL database by uninstalling the data capture module, as this would eradicate the possibility for its reconstruction, should those in power decide to do so (Bundgaard, 2015). However, RSKD made it clear that it was the responsibility of individual general practitioners, and not the data controller, to decide the fate of the local SQL database that they were contractually obligated to install with the collective agreement of April 2010. Cekic's question ('has DAMD been deleted?') was not easy to answer, in part because there was no agreement on what constituted deletion. Centralized storage in the DAMD database and state archives was established as unlawful, but there was (unlike research environments) no legal obstructions against the *decentralized* existence of copies in the data capture mechanism on the computers of general practitioners.

By foregrounding different obstacles to certainty on deletion, we have brought out ways in which it was imagined and put into practice. For the DAMD database, achieving deletion was not so much a question of technological destruction, but of settling agreement on what *kind* of deletion would be sufficient. Different methods for data destruction involve different 'trade-offs': The more certain the data destruction, the longer the process takes to complete' (Mooy et al., 2017: 6, see also Diesburg and Wang, 2010). Evidencing destruction has become a new, forensic area, deserving considerable research attention, especially if deletion is contested. In the words of the International Data Sanitization Consortium (2019), a device, once sanitized, 'has no usable residual data, and even with the assistance of advanced forensic tools, the data will not ever be recovered'. The kinds of technologies (and terminologies) that go along with such 'data sanitization' include physical destruction, cryptographic erasure and data erasure.

What the DAMD case demonstrates is that, whilst certainty of deletion remains an ideal and a commercial concern, the conditions under which deletion is achieved remain subject to uncertainty. Institutions are never far from sight, particularly when decentralized data is split across locations and differential regimes of accountability. In her recent article on toxic data traces, Thylstrup observes that 'waste theories show us that datafication is ... always haunted by [digital] traces and the threat that they will resurface' (2019: 2). This haunting is, in part, addressed by new rights granted to EU citizens under GDPR, which allows them to not only ask for their data to be deleted, but for proof that it indeed has been (European Union, 2016). In the run-up to the implementation of this right, John Rose from the Boston Consulting Group expressed concerns: 'I am not sure how many organizations will be able to execute deletions in the context of the infrastructure they have now, and at a cost that is manageable', he said (Palmer, 2013). His reasoning was that most data about an individual is distributed, held in different formats and potentially painstaking to locate. In a time of copies and decentralized databases, technological and institutional work provide different mechanisms for arguing how deletion is socially achieved. If one can locate data, its deletion must then be evidenced and irretrievability certified. Through new legal regimes, both verification and certification have been bound into the legal technologies of deletion. The burgeoning data sanitization industry sells not just data deletion, but *evidence* thereof. To have acted on data – deleted it – one must produce data on that act of deletion. Unlike the data that makes other data usable (the ever neglected achievements of metadata (Boyd, 2018; Edwards, 2011; Mayernik, 2019), or the work that makes data actionable (Tempini and Leonelli, 2021), this metadata offers closure. Its evidentiary power as data, then, is that (other) data deletion has occurred (Neyland, 2018).

Conclusion

Through the twists and turns of a complex database case, we have framed deletion as the outcome of a rejection of retention and as a negotiated socio-technical achievement. We have argued that what becomes available for deletion emerges from intertwined institutional, technical and temporal processes, from both those who reject retention and those who seek to make deletion happen. Returning to our assertion at the opening of this paper that deletion must also be recognized as creative, as much as it is critical, we can now reflect on the creative and interpretive work within the story. As the mandate to delete shifted between the actors involved, there were competing valuations of the dataset, temporalities in which it was being viewed, and technologies that sought as much to delete as to retain.

First, the mandate to delete the DAMD database was not easily obtained. As responsibility was passed between actors in the story, public institutions came forward with an interest to see it kept. In addition to illustrating the importance of *process* around deletion (who decides?), these interests can remind scholars to remain focused when attending the data environments in which deletion takes place. As we outlined in our introduction, the Scandinavian states not only have a long history of collecting centralized data, but also of putting it to use within epidemiological studies and biomedical research projects (Bauer, 2014; Hoeyer, 2019). Thus, the calls to deletion in the DAMD case were taking place in a promissory data environment. We might extend Bauer's

(2014) use of the term ‘laboratory’, initially used to describe traits of the emergent data collecting nation state in a more literal direction, by observing the quasi-experimental nature of what data is kept and deleted in these cases. The negotiations illustrate that case outcomes cannot always be known in advance, and that the national and historic setting in which data deletion takes place shapes the conversations that occur. For example, work reviewing the increased attention to ‘data sovereignty’ has pointed out that in ‘cases of conflicts between claims from different putative data sovereigns, e.g. when national data sovereignty undercuts individual citizen data sovereignty’, the concept of data sovereignty in itself offers no mechanism of resolution (Hummel et al., 2021: 14). By studying deletion as a site of negotiation between states and citizens, we can see these conflicting claims. When deletion is mandated by a state, these negotiations take on a different character, with retention acting as a means to manage state power, also discussed by Dillon et al. (2019: 5) as a key practice of ‘situating data’. Thus, in addition to questions of context and control, attending to deletion necessitates the identification of its opposite: What kind of retention is keeping data in place? Is it passive retention, or active? Is it deliberate or accidental? As environments of data promise differ, so do the futures to which they articulate a relationship.

Second, our analysis demonstrates that, while language may tempt us towards a binary, deletion and retention do not operate as clean opposites. As different technologies of deletion played out, archiving, opting out and anonymization were shown to be tied to postponement – a sidestepping of deletion that allowed the continued existence of data. Building on our point about deletion in promissory environments, presented as options to counter the strongest forms of deletion, opting out is a technique that retains the norm of data’s value and future promise.

These should, we argue, be seen alongside and in conversation with other technologies of information erasure, from shredding (Blacker, 2020) and burning (Ovenden, 2020) to data sanitization. As the International Data Sanitization Consortium has noted, a more precise vocabulary around data deletion is needed. Deletion is perhaps too singular a term, yet its legibility in public debates, especially after the 2018 implementation of the GDPR in Europe, means assumptions about data being ‘gone’ after deletion continues (Neyland, 2018). Within DAMD, the state management of information means an interrogation of the work of the digital state and the accountabilities within which calls for deletion are enmeshed. Deletion – in its many forms – can become politically desirable for the (paradoxical) purpose of retention.

Calls to delete data demand that actors articulate the basis of their desire to keep that data. Refusals make the deletion processes an explicit negotiation on whether what is contested is to be retained or not. Refusal became available as a position once media scrutiny brought the data collection to the fore. Refusal denoted a series of public reasonings, which in turn made data infrastructure and motivations visible. Refusals prompted changes to deletion practices within different parts of the state. To return to Benjamin's argument, refusals come 'seeded with a vision of what can and should be, and not only a critique of what is' (2016: 4). Her call marks the role of the state in data retention and, read back against our cases, exposes tensions over the tendency of data protections to deal with *individual* rights over collective investments (Taylor et al., 2017): New conversations are needed for what can and what should be. Without such public conversations that bring to light how data is valued, by whom and in what ways, it becomes increasingly possible that retention becomes a site of refusal.

---

[1] Some cannot be. On the 11th of December 2016, just a month after the election of the US president Donald Trump, concerns about deletions of state-held environmental data led to initiatives to harvest, duplicate and save millions of files. The Environmental Data and Governance Initiative was led by STS scholar Michelle Murphy (Dillon et al., 2019), and an excellent analysis of public initiatives to prevent data deletion is covered in Lamdan (2018). The example highlights the inseparability of information and practices of valuation. As Cohen (1999[1982]: 2060) writes in another context, 'what was counted was what counted'.

[2] Skinner and Wienroth cite a 2011 UK Home Office impact assessment that details highly precise financial calculations of the cost of *reprogramming* the computer software to delete DNA profiles, the cost of *deleting orphaned profiles* and destroying DNA samples and paper records of fingerprints (Skinner and Wienroth, 2019: 16).

[3] Celebrating International Libraries Week, the official Twitter account of Oxford's Bodleian library suggested that e-books might be a 'green alternative to printed books', arguing that they 'don't use paper for their production nor require physical storage' (BodleianLibs, 2019). The appended image was a vast warehouse of thousands of filed, printed academic journals. The comment attracted the mocking attention not only of librarians but also science and technology scholars for its failure to recognise that e-books also require storage with physical components – servers, reader devices and so on.

[4] The Danish Health Data Authority was established in November 2015, which is now responsible for the approval of 'clinical quality databases' such as the DAMD and research access to Danish health registers.

[5] The query included '#DAMD', 'Dansk AlmenMedicinsk Database' and 'DAMD'.

[6] Under the Danish Healthcare Act, exemption from explicit consent regarding the registration of health information can only be approved when limited to 'defined groups of patients' (Statens Serum Institut, 2014b: 1). This approval was established for four chronic disease areas (diabetes, chronic obstructive pulmonary disease, depression and heart failure), each one of which existed as a copy in a dedicated sub-database (or 'clinical quality database').

[7] The extent to which the authorizations were issued to include the broad collection is, however, subject to dispute between, among others, the applicant and the authorizing body (i.e. Statens Serum Institut) (see Statens Serum Institut, 2014a: 299).

---

[8] A similar sentiment was, among others, articulated by The Danish College of General Practitioners (DSAM) and the umbrella organization for 83 patient associations in Denmark, Danish Patients (Nielsen, 2014).

[9] The mentioned act was replaced by the Danish Data Protection Act in 2018.

[10] Two general practitioners had previously filed the data controller and data processor to the police (Heissel, 2014).

[11] All contributors to the debate are pseudonymized when named. Given the public nature of the debate, those familiar with the case will be able to identify the relevant spokespersons.

[12] The sociotechnical infrastructuring of the DAMD database did, however, play a key role in the intensification of this process (Wadmann and Hoeyer, 2018).

[13] This is not a move unique to the DAMD case. In 2018, a hitherto unknown archive in the Danish Security and Intelligence Service (PET) entered the public spotlight because it contained expired data that was supposed to be archived under restricted control in DKNA. Instead of sending the data to DKNA, however, an undisclosed volume of expired data was, for 'practical and safety reasons', moved to an internal archive under the custody of top-level management where access was governed under opaque conditions (Kjeldtoft, 2018). This was termed 'logical deletion' (Kjeldtoft, 2018), which critics argued was 'nor deletion nor logical' (Kongstad, 2018).

[14] Others were more forthright in their criticism, such as the chairman of the IT-Political Association of Denmark, Jesper Lund (2015), who equated archiving with 'whitewashing', not deletion.

[15] DAMD was created to improve healthcare quality, but access to the database for research was of interest from the beginning. Anonymized data for research required an application approved by a quality and research committee, the Danish Data Protection Agency and (if relevant) a regional ethics council. By the time DAMD was hitting headlines in 2014, research approval had been granted for 50 projects (National Archives, 2015a). The prospect of data reuse, both for administrative control and economic growth, created particular concern among general practitioners and citizens (see Langhoff et al., 2016, Wadmann and Hoeyer, 2018).

[16] The 27th of March accounted for more than a fifth (n=260) of the tweets in the Twitter dataset (n=1363).

[17] Media outlets occupied with welfarism (Henriksen, 2015) and digitization (Bernth, 2015) also circulated the practical instructions.

[18] In practice, however, 'acceptable level of risk' (Esayas, 2015) is compared to 'irreversibility', a more pragmatic evaluation criteria for anonymization, which acknowledges the risk of re-identification involved with practices of data linkage and inference (see also Bernal, 2011).

[19] According to an interview informant involved in the development of the DAMD, CPR-numbers were kept in a dedicated data table separate from the patient data stored elsewhere in the database (accessible to a select few with logged access). The linkage between CPR-numbers and patient data was facilitated by a shared unique code or proxy applied both places.

[20] As years of work in the field of social genetics has shown, data about one person can be personal to others too (Konrad, 2003). For DAMD this could be inherited or stigmatized diagnosis codes within a family.

[22] The establishment of health registries in Denmark has, according to Hallas (2011: 621), generally been 'uneventful'. The prescription database from 1993 is, however, an exception to this observation. due to concerns (by the Danish Medical Association, among others) over the scope of data registration, possibility of state surveillance of prescription practices and risk of misuse (Hallas, 2011). Responding to the political climate, the Minister of Health, Torben Lund, abandoned the previously stated intentions to append a CPR-number for the registration of prescription information, in part by implementing an irreversible encryption of person identifiers.

In the DAMD case, the bill then proposed to reverse the 'irreversible' by means of duplication, as DAMD entailed person-identifiable prescription records tied to a CPR-numbers.

[23] A recent example concerns personal well-being information on children collected through a national survey without the necessary consent. After denying the right to opt out entirely, the Ministry of Education (2018) issued a four month window within which parents could request anonymization on behalf of their children *insofar* that the number of requests did not exceed an unspecified amount that would compromise the perceived value of the dataset.

[24] The modernization of health data access has become an object of investment in Europe more broadly with concrete initiatives in the Nordic countries, Italy, Spain, France, Germany, and the UK (Burgun et al., 2017).

References

Bachmann G, Bialski P and Hansen M (2017) The terms of anonymity: An interview with Marit Hansen, German data protection expert. *Ephemera* 17(2): 421-430.

Bauer S (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.

Beer D (2016) How should we do the history of Big Data? *Big Data & Society* 3(1): 1-10.

Benjamin R (2016) Informed refusal: Towards a justice-based bioethics. *Science. Technology, & Human Values* 41(6): 967-990.

Bernal PA (2011) A right to delete? *European Journal of Law and Technology* 2(2): 1-18.

Bernth M (2015) Sådan sletter du dine data mens myndighederne slås om omstridt database. *Version2*, 19 March, 2015. Available at: https://www.version2.dk/artikel/omstridt-patientdatabase-kan-ikke-gemmes-under-gaeldende-lovgivning-106088.

Blacker S (2020) Analogue privacy: The paper shredder as a technology for knowledge destruction. In: Bauer S, Schlünder M, and Rentetzi M (eds) *Boxes: A Field Guide.* Manchester: Mattering Press, 365-378.

Blanchette JF and Johnson D (2002) Data retention and the panoptic society: The social benefits of forgetfulness. *The Information Society* 18(1): 33-45.

BodleianLibs (@bodleianlibs) (2019) Are e-books & e-journals the green alternative to printed books? They don't use paper for their production nor require physical storage. And there are an incredible 5 million e-books and e-journal articles available to readers through #UKLegalDeposit. #librariesweek. *Twitter*, Oct 9, 2019. Available at: https://twitter.com/bodleianlibs/status/1181854717594484736.

Borggreen M (2015) Ulovlige DAMD-oplysninger slettet. *Region Syddanmark* May 28. Available at: https://www.regionsyddanmark.dk/wm467297 (accessed 8 Oct 2019).

Borgman CL (2015) *Big Data, Little Data, No Data: Scholarship in the Networked World.* Cambridge/London: MIT Press.

Bowker G (2005) *Memory Practices in the Sciences*. Cambridge: MIT Press.

Bowser A, Shilton K, Preece J et al. (2017) Accounting for privacy in citizen science: Ethical research in a context of openness. In: Lee C and Poltrock S *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing.* New York: Association for Computing Machinery, 2124-2136.

Boyd NM (2018). Evidence enriched. *Philosophy of Science* 85(3): 403–421.

Boye M (2014) Rigsarkivet vil gemme ulovlig patientdatabase. *Version2.* 18 December. Available at: https://www.version2.dk/artikel/rigsarkivet-vil-gemme-ulovlig-patientdatabase-75682.

Boysen M (2014) Trine Jeppesen: Katastrofe at DAMD skal slettes. *Dagens Medicin*, 17 December. Available at: https://dagensmedicin.dk/trine-jeppesen-katastrofe-at-damd-skal-slettes/.

Boysen M (2015a) Kulturminister vil alligevel bevare DAMD-data i Rigsarkivet. *Dagens Medicin,* 18 February. Available at: https://dagensmedicin.dk/kulturminister-vil-alligevel-bevare-damd-data-i-rigsarkivet/.

Boysen M (2015b) DAMD-data bevares foreløbigt. *Dagens Medicin*, 25 March. Available at: https://dagensmedicin.dk/damd-data-bevares-forelobigt/.

Boysen M (2015c) Syddanmark har slettet 23.000 i DAMD: »Vi har vel brugt 2-3 mandeår på denne sag«. *Dagens Medicin*, 13 May. Available at: https://dagensmedicin.dk/syddanmark-har-slettet-23000-i-damd-vi-har-vel-brugt-2-3-mandear-pa-denne-sag/.

Buckland M (1999) Information as thing. *Journal of the American Society of Information Science* 42(5): 351-360.

Bundgaard B (2015) Kopier af følsomme patientdata ligger stadig i cyberspace. *Ugeskrift for læger*, July 23. Available at: https://ugeskriftet.dk/nyhed/kopier-af-ulovligt-indsamlede-patientdata-flyder-rundt-i-cyberspace (accessed 10 Oct 2019).

Burgun A, Bernal-Delgado E, Kuchinke W et al. (2017) Health aata for public health: Towards new ways of combining data sources to support research efforts in Europe. *Yearbook of Medical Informatics* 26(1): 235–240.

Caswell M (2014) Inventing new archival imaginaries: Theoretical foundations for identity-based community archives. In Daniel D and Levi A (eds) *Identity Palimpsests: Ethnic Archiving in the U.S. and Canada.* Sacramento: Litwin Books, 35-55.

Cekic ÖS (@OzlemCekic) (2015) Er der nogen der ved om #DAMD den ulovlige sundhedsdata er slettet? #dkpol #dksund. *Twitter*, Aug 19. Available at: https://twitter.com/cekicozlem/status/634255535202824192. (accessed 24 Jan 2018).

Cifor M and Wood S (2017) Critical feminism in the archives. *Journal of Critical Library and Information Studies* 1(2): 1–27.

Cohen PC (1999) *A Calculating People: The Spread of Numeracy in Early America*. New York: Routledge.

The European Parliament and the Council of the European Union (1995) *Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data.* Available at: https://eur-lex.europa.eu/eli/dir/1995/46/2003-11-20 (accessed 10 November 2020).

Danish Parliament (2015a) L 168 Forslag til lov om ændring af arkivloven: 1. behandling den 08-04-2015. Available at:https://www.ft.dk/samling/20141/lovforslag/L168/BEH1-73/forhandling.htm. (accessed 6 November 2020).

Danish Parliament (2015b) L 168 Forslag til lov om ændring af arkivloven: Afstemninger. Available at: https://www.ft.dk/samling/20141/lovforslag/L168/90/afstemninger.htm (accessed 6 November 2020).

de Holan PM and Phillips N (2004) Organizational forgetting as strategy. *Strategic Organization* 2(4): 423–433.

De Mooy M, Jerome J and Kasschau V (2017) Should it stay or should it go? The legal, policy and technical landscape around data deletion. Report, Center for Democracy & Technology, US, February. Available at: https://cdt.org/files/2017/02/2017-02-23-Data-Deletion-FNL2.pdf.

Derrida J (1996). *Archive Fever: A Freudian Impression*. Chicago: University of Chicago Press.

Dever M (2015). Archiving feminism: Papers, politics, posterity. *Archivaria* 77: 25–42.

Diesburg SM and Wang AA (2010) A survey of confidential data storage and deletion methods. *ACM Computing Surveys* 43(1): 1–37.

Dillon L, Lave R, Mansfield B et al. (2019) Situating data in a Trumpian era: The environmental data and governance initiative. *Annals of the American Association of Geographers* 109(2): 545-555.

Douglas M (2001) [1966] *Purity and Danger: An Analysis of Concepts of Pollution and Taboo*. London/New York: Routledge.

Douglas-Jones R, Walford A, and Seaver N (2021) Introduction: Towards an anthropology of data. *Journal of the Royal Anthropological Institute* 27(S1): 9–25.

Edwards PN, Mayernik MS, Batcheller AL et al. (2011) Science friction: Data, metadata, and collaboration. *Social Studies of Science* 41(5): 667–690.

Esayas S (2015) The role of anonymisation and pseudonymisation under the EU data privacy rules: Beyond the 'all or nothing' approach. *European Journal of Law and Technology* 6(2): 1-28.

European Union (2016) General Data Protection Regulation (GDPR). Available at: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32016R0679 (accessed 10 November 2020).

Fischer A (2014a) Region vil slette millioner af fortrolige patientdata. *Danish Broadcasting Corporation*, November 28. Available at: https://www.dr.dk/nyheder/indland/region-vil-slette-millioner-af-fortrolige-patientdata (accessed 10 October 2019).

Fischer A (2014b) Rigsarkivet mener det SKAL gemme ulovlige patientdata. Danish Broadcasting Corporation, December 4. Available at: https://www.dr.dk/nyheder/indland/rigsarkivet-mener-det-skal-gemme-ulovlige-patientdata (accessed 10 October 2019).

Floridi L and Taddeo M (2016) What is data ethics? *Philosophical Transactions of the Royal Society A: Mathematical Physical and Engineering Sciences* 374(2083): 1-5.

Flyvbjerg C and Larsen D (2015) Sidste chance hvis du vil fjernes fra omstridt database. *Danish Broadcasting Corporation*, 27 March. Available at: https://www.dr.dk/nyheder/regionale/fyn/sidste-chance-hvis-du-vil-fjernes-fra-omstridt-database.

Forde I, Nader C, Dietrich KS et al. (2016) Primary care review of Denmark. Report, OECD, France, December. Available at: https://www.oecd.org/health/health-systems/Primary-Care-Review-of-Denmark-OECD-report-December-2016.pdf

Foroughi H, Coraiola D, Rintamäki J et al. (2020) Organizational memory studies. *Organization Studies* 41(12): 1725–1748.

Gilliland AJ and Caswell M (2016) Records and their imaginaries: Imagining the impossible, making possible the imagined. *Archival Science* 16: 53-75.

Hacking I (1991) How should we do the history of statistics? In: Burchell G, Gordon C and Miller P (eds) *The Foucault Effect: Studies in Governmentality*. Chicago: The University of Chicago Press,181-195.

Hallas J (2001) Conducting pharmacoepidemiologic research in Denmark. *Pharmacoepidemiology and Drug Safety* 10(7): 619–623.

Hecklen A (2016) Overblik: Det ved vi om de fem millioner fejlsendte cpr-numre. *Danish Broadcasting Corporation*, July 20. Available at: https://www.dr.dk/nyheder/indland/overblik-det-ved-vi-om-de-fem-millioner-fejlsendte-cpr-numre (accessed 8 Jul 2021).

Heissel A (2014) To praktiserende læger politianmelder region og DAK-E for ulovlig datahøst. *Ugeskrift for læger*, December 12. Available at: https://ugeskriftet.dk/nyhed/praktiserende-laeger-politianmelder-region-og-dak-e-ulovlig-datahoest (accessed 10 October 2019).

Heissel A (2015) Minister: Rigsarkivet får alligevel en kopi af DAMD. *Ugeskrift for læger*, February 18. Available at: https://ugeskriftet.dk/nyhed/minister-rigsarkivet-faar-alligevel-en-kopi-af-damd (accessed 10 October 2019).

Henriksen LR (2015) Uacceptabelt, at Kulturministeren vil give Rigsarkivet ulovlige patientdata. *DenOffentlige.* Available at: https://www.denoffentlige.dk/uacceptabelt-kulturministeren-vil-give-rigsarkivet-ulovlige-patientdata

Hoeyer K (2016) Denmark at a crossroad? Intensified data sourcing in a research radical country. In: Mittelstadt BD and Floridi *L* (eds) *The Ethics of Biomedical Big Data.* Cham: Springer, 73-93.

Hoeyer K (2019) Data as promise: Reconfiguring Danish public health through personalized medicine. *Social Studies of Science* 49(4): 531-555.

Hogan M and Vonderau A (eds) (2019) The Nature of Data Centres. *Culture Machine* 18.

Hummel P, Braun M, Tretter M et al. (2021) Data sovereignty: A review. *Big Data & Society* 8(1): 1–17.

International Data Sanitization Consortium (2019) Data Sanitization Terminology and Definitions. Available at: https://www.datasanitization.org/data-sanitization-terminology/#physical-destruct (accessed 2 August 2019).

Jasanoff S (2004) *States of Knowledge: The Coproduction of Science and Social Order.* London: Routledge.

Jørgensen D and Gudmundson IML (2019) In the Clouds: An ArtScience Workshop. The Greenhouse, Stavanger University, Norway. 24-26 January 2019.

Kjeldtoft SS (2018) PET har skjult database takket være historikere. *Politiken.* Available at: https://politiken.dk/indland/art6413709/PET-har-skjult-database-takket-v%C3%A6re-historikere (accessed 12 April 2019).

Kongstad J (2018) Pape: PET skal forklare Folketinget om slettede data i hemmeligt lager. *Jyllands-Posten.* Available at: https://jyllands-posten.dk/indland/ECE10481279/pape-pet-skal-forklare-folketinget-om-slettede-data-i-ukendt-lager/ (accessed 12 Apr 2019).

Konrad M (2003) Predictive genetic testing and the making of the pre-symptomatic person: Prognostic moralities amongst Huntington's-affected families. *Anthropology & Medicine* 10(1): 23-49.

Kristiansen TB (2015) Kulturudvalget (KUU) Alm. Del: Bilag 40. *Folketinget,* 3 November. Available at: https://www.ft.dk/samling/20151/almdel/kuu/bilag/40/index.htm (accessed 9 October 2019).

Krogness KJ (2011) Numbered individuals, digital traditions, and individual rights: Civil status registration in Denmark 1645-2010. *Ritsumeikan Law Review* 28: 87-126.

Lamdan S (2018) Lessons from Datarescue: The limits of grassroots climate change data preservation and the need for federal records law reform. *University of Pennsylvania Law Review Online* 166: 231-248.

Langhoff TO, Amstrup MH, Morck P et al. (2016) Infrastructures for healthcare: From synergy to reverse synergy. *Health Informatics Journal* (24)1: 1-11.

Lappin J, Jackson T, Matthews G et al. (2018) The defensible deletion of government email. *Records Management Journal* 20(1-2): 42–56.

Larsen K (2015) P1 Orientering: Ulovligt indsamlede patientoplysninger havner alligevel i Rigsarkivet. *Danish Broadcast Corporation*, 17, February.

Larsen RP and Foldberg A (2015) Venstremand åbner op for salg af patientoplysninger. *Danish Broadcasting Corporation*, 24, March. Available at: https://www.dr.dk/nyheder/regionale/syd/venstremand-aabner-op-salg-af-patientoplysninger

Leonelli S (2016) *Data-Centric Biology: A Philosophical Study.* Chicago: University of Chicago Press.

Leonelli S, Rappert B and Davies G (2016) Data shadows: Knowledge, openness and absence. *Science, Technology, & Human Values* 42(3): 191–202.

Lippert ML, Reventlow S and Kousgaard MB (2017) The uses and implications of standards in general practice consultations. *Health* 21(1): 3–20.

Loiborg C (2015) Kulturministeriet lyn-ændrer høringsdato for at bevare ulovligt indsamlede personfølsomme oplysninger. *Version2*, *24, March.* Available at: https://www.version2.dk/artikel/kulturministeriet-lyn-aendrer-hoeringsdato-bevare-ulovligt-indsamlede-personfoelsomme

Lund J (@je5perl) (2015) Legal, illegal data collection? It means NOTHING for the public sector. The National Archives are whitewashing the illegal data with the archive act #dkpol #DAMD. *Twitter*, 26 March. Available at: https://twitter.com/je5perl/status/581238836686110720 (accessed 24 January 2018).

Lundström EC (2015) Følsomme patientdata, der skulle have været slettet, ligger stadig gemt i SQL-databaser. *Version2,* July 28. Available at: https://www.version2.dk/artikel/foelsomme-patientdata-der-skulle-have-vaeret-slettet-ligger-stadig-gemt-i-sql-databaser (accessed 9 October 2019).

Maguire J and Winthereik BR (2019) Digitalizing the state: Data centres and the power of exchange. *Ethnos 86(3)*: 530-551.

Mayernik M (2019) Metadata accounts: Achieving data and evidence in scientific research. *Social Studies of Science* 49(5): 732-757.

Meyer-Schönberger V (2011) *Delete: The Virtue of Forgetting in the Digital Age.* Princeton: Princeton University Press.

Ministry of Culture (2015a) Jelved vil ændre arkivloven efter DAMD-sag. 24 February, Available at: https://kum.dk/nyheder-og-presse/pressemeddelelser/nyheder/jelved-vil-aendre-arkivloven-efter-damd-sag/1/1/ (accessed 25 August 2020).

Ministry of Culture (2015b) Forslag til lov om ændring af arkivloven (L 168 af 25/03/2015). Available at: https://www.ft.dk/ripdf/samling/20141/lovforslag/l168/20141_l168_som_fremsat.pdf (accessed 6 Nov 2020).

Ministry of Education (2018) Cirkulæreskrivelse om bemyndigelse til direktøren for Styrelsen for It og Læring om særlig adgang for anonymisering af oplysninger i styrelsen om elevers besvarelser ved trivselsmålinger i folkeskolen i 2015 – 2017 (CIS nr 9198 af 23/03/2018). *Retsinformation.* Available at: https://www.retsinformation.dk/Forms/R0710.aspx?id=200365 (accessed 9 October 2019).

National Archives (2015a) Redegørelse, del 1: Indsamling og Bevaring. January, 15. Available at: https://ugeskriftet.dk/files/redegoerelse_del_1_dok2516643.pdf.

National Archives (2015b) Redegørelse, del 2: Indsamling og Bevaring. January, 15. Available at: https://ugeskriftet.dk/files/redegoerelse_del_2_dok2516644.pdf.

Neyland D (2018) Something and nothing: On algorithmic deletion, accountability and value. *Science and Technology Studies* 31(4): 13–29.

Nielsen HF (2014) Forsøg på at redde ulovlige data. *Jyllands-Posten.* Available at: https://jyllands-posten.dk/indland/ECE7258865/forsoeg-paa-at-redde-ulovlige-data/ (accessed 12 April 2019).

Nielsen LB and Tynell J (2014) Jelved stopper ulovligt arkiv med danskeres sundhedsoplysninger. *Danish Broadcasting Corporation,* December 22. Available at: https://www.dr.dk/nyheder/indland/jelved-stopper-ulovligt-arkiv-med-danskeres-sundhedsoplysninger (accessed 7 October 2019).

Nordfalk F and Hoeyer K (2020) The rise and fall of an opt-out system. *Scandinavian Journal of Public Health* 48(4): 400–404.

Ovenden R (2020) *Burning the Books: A History of the Deliberate Destruction of Knowledge*. Cambridge: Harvard University Press.

Palmer M (2013) Cloud computing hinders data deletion. *The Financial Times.* Available at: https://www.ft.com/content/0e8aad72-7444-11e2-80a7-00144feabdc0 (accessed 3 June 2019).

Pfitzman A and Hansen M (2010) A terminology for talking about privacy by data minimization: Anonymity, unlinkability, undetectability, inobservability, pseudonymity and identity management. Available at: http://www.maroki.de/pub/dphistory/2010_Anon_Terminology_v0.34.pdf.

Power M (1997) *The Audit Society: Rituals of Verification*. Oxford: Oxford University Press.

Region Southern Denmark (2014) Carl Holst: Meningsløs besked fra Rigsarkivet, December 21. Available at: https://www.regionsyddanmark.dk/wm458845 (accessed 9 October 2019).

Ritzau (2014) SF: Jelved må stoppe arkivering af ulovlig sundhedsdata. *Berlingske*, 22 December. Available at: https://borsen.dk/nyheder/politik/sf-jelved-maa-stoppe-arkivering-af-ulovlig-sundhedsdata-2makl

Ritzau (2015) Ulovligt indsamlet sundhedsdata er næsten slettet. *TV2,* 28 May. Available at: https://nyheder.tv2.dk/samfund/2015-05-28-ulovligt-indsamlet-sundhedsdata-er-naesten-slettet (accessed 7 October 2019).

Ruppert E, Grommé F, Ustek-Spilda F et al. (2018). Citizen data and trust in official statistics. *Economie et Statistique / Economics and Statistics* 505-506: 171–184.

Sadowski J (2019) When data is capital: Datafication, accumulation and extraction. *Big Data & Society* 6(1): 1-12.

Schmidt M, Pedersen L and Sørensen HT (2014) The Danish Civil Registration System as a tool in epidemiology. *European Journal of Epidemiology* 29(8): 541–549.

Skinner D and Wienroth M (2019) Was this an ending? The destruction of samples and deletion of records from the UK Police National DNA Database. *British Society for the History of Science* 4: 99-121.

Skovgaard LL, Wadmann S and Hoeyer K (2019) A review of attitudes towards the reuse of health data among people in the European Union: The primacy of purpose and the common good. *Health Policy* 123(6): 564–571.

Sonne N (@nikolajsonne) (2015) Think of everything you have told your doctor - well then. You have 6 hours left to prevent it from ending up in the National Archives #damd rsyd.dk/wm461183. *Twitter*, 27 March. Available at: https://twitter.com/nikolajsonne/status/581493927137660928 (accessed 24 January 2018).

Starosielski N (2015) *The Undersea Network: Sign, Storage, Transmission.* Durham: Duke University Press.

Statens Serum Institut (2014a) Bilag til udredning vedr. ansøgninger om godkendelse af DAMD som klinisk kvalitetsdatabase, fra 2007 og frem. Available at: https://www.ssi.dk/-/media/arkiv/indhold/dk-dansk/aktuelt/nyheder/bilag-til-udredning-vedr-ansgninger-om-godkendelse-af-damd-som-klinisk-kvalitetsdatabase-fra-2007-og-frem.pdf?la=da (accessed 26 October 2020).

Statens Serum Institut (2014b) Udredning vedr. ansøgninger om godkendelse af DAMD som klinisk kvalitetsdatabase, fra 2007 og frem. Available at: https://www.ssi.dk/-/media/arkiv/indhold/dk-dansk/aktuelt/nyheder/udredning-vedr-ansoegninger-om-damd-26nov2014.pdf?la=da (accessed 23 Oct 2019).

Strathern M (ed) (2000) *Audit Cultures: Anthropological Studies in Accountability Ethics and the Academy*. London: Taylor and Francis.

Tempini N and Leonelli S (2018) Concealment and discovery: The role of information security in biomedical data re-use. *Social Studies of Science* 48(5): 663–690.

Tempini N and Leonelli S (2021) Actionable data for precision oncology: Framing trustworthy evidence for exploratory research and clinical diagnostics. *Social Science & Medicine* 272(113760): 1-10.

The Danish Act on Processing of Personal Data (2000) The Danish Act on Processing of Personal Data (*LOV nr 429 af 31/05/2000*). *Retsinformation*. Available at: https://www.retsinformation.dk/eli/lta/2000/429.

The Danish Archives Act (2007) Consolidation Act of the Danish Archives Act (*LBK nr 1035 af 21/08/2007*). *Retsinformation.* Available at: https://www.retsinformation.dk/eli/lta/2007/1035.

The Danish Health Data Authority (2017) Forskerservice Klassisk udfases i 2018. Available at: https://sundhedsdatastyrelsen.dk/da/forskerservice/om-forskerservice/nyheder_forskerservice/udfases_2018 (accessed 8 Jul 2021).

Taussig K, Hoeyer K and Helmreich K (2013) The anthropology of potentiality in biomedicine. *Current Anthropology* 54(S7): S3–S14.

Taylor L (2016) The ethics of big data as a public good: Which public? Whose good? *Philosophical Transactions of the Royal Society A: Mathematical Physics and Engineering Sciences* 374(2083): 1-13.

Taylor L, Floridi L and Van der Sloot B (eds) (2017) *Group Privacy: New Challenges of Data Technologies*. Dordrecht: Springer.

Thylstrup NB (2019) Data out of place: Toxic traces and the politics of recycling. *Big Data & Society* 6(2): 1-10.

Tupasela A, Snell K and Tarkkala H (2020) The Nordic data imaginary. *Big Data & Society* 7(1): 1–13.

Tynell J and Fischer A (2014a) Fortrolige sundhedsdata samles ulovligt i database. *Danish Broadcasting Corporation*. 29 September. Available at: https://www.dr.dk/nyheder/indland/fortrolige-sundhedsdata-samles-ulovligt-i-database

Tynell J and Fischer A (2014b) Kun kulturministeren kan hindre at ulovlige patientdata gemmes. *Danish Broadcasting Corporation.* 19 December. Available at: https://www.dr.dk/nyheder/indland/kun-kulturministeren-kan-hindre-ulovlige-patientdata-gemmes

van Bussel G, Smit N and van de Pas J (2015) Digital archiving, green IT and environment: Deleting data to manage critical effects of the data deluge. *Electronic Journal of Information Systems Evaluation* 18(2): 187–198.

Wadmann S and Hoeyer K (2018) Dangers of the digital fit: Rethinking seamlessness and social sustainability in data-intensive healthcare. *Big Data & Society* 5(1): 1–13.

Walford A (2021) Data - ova - gene - data. *Journal of the Royal Anthropological Institute* 27(S1): 127–141.

Infrastructures in the promotion of ethical tech: The rise of data intermediaries in the Nordic market for digital health

John Mark Burnett

Abstract

In recent years, ‘ethics’ has become a trend in the tech industry. While an increased focus on ethics is timely, critics caution that ethical claims advanced by practitioners do not necessarily transcend the level of discourse to deliver a practical impact. This is somewhat exacerbated by the capacious nature of the term ethics, as it lends itself to ambiguity and thus questioning of what ethical change should look like. Responding to this ambiguity, this article explores emerging understandings of ethical data sharing through an ethnographic case study of two data intermediaries - Digi.Me and Data for Good Foundation - in their attempts to enter the Nordic market for digital health. The ideas of ethical data sharing enacted by both startups are intimately tied to ideas of individual control that they promote through competing platform models. The question I ask is this: What role does infrastructure play as Digi.Me and DfG promote themselves as ‘the ethical choice’? I examine this question in relation to three different audiences: the user, the market and the communities they engage in. What I present is a combined analysis attentive to how ethics is narrated, but crucially also how it is technologically infrastructured. The article demonstrates how infrastructures are also used in the promotion of ethical tech and argues that more ethnographic research is needed to illuminate the infrastructural practices that underpin the ethical claims put forward by tech actors in the data economy.

Keywords

Infrastructure, personal health data, individual control, ethics, data intermediaries, nordic welfare states

Introduction

Ethics has become a trend in the tech industry, with scholars describing the phenomenon as ‘the hottest product in Silicon Valley’s hype cycle today’ (Metcalf, Moss and Boyd, 2019: 449). In recent years, big tech companies have been observed to establish boards, working groups, positions and strategic relationships with a dedicated focus on ethics (e.g. Metcalf, Moss and Boyd, 2019; Phan et al., 2021). The context for such initiatives necessarily includes the recent ‘tech lash’ where public scandals have done little to strengthen the public relations of these actors. On the contrary, such events add fuel to a perceived disparity between reported data practices and the public perception of ethical conduct (Moss and Metcalf, 2019). On the one hand, an increased focus on ethics is then timely, and may even help improve their practices and salvage their reputation. On the other hand, critics caution that the capacious nature of the term ‘ethics’ introduces a risk of ambiguity that can ultimately be its own detriment (Metcalf, Moss and Boyd, 2019). This article aims to cut through the ambiguity of ethical tech by grounding it through an ethnographic case study that both situates the trend of ethics while exploring its manifestation through infrastructures.

At the centre of the case study are two platform startups that I follow in their attempt to enter the Nordic market for digital health: Data for Good Foundation and Digi.Me. Both startups are enrolled in a movement for ‘individual control’, which following STS scholar Barbara Prainsack can be loosely defined to consist of actors ‘who want individual citizens to have more effective control over their own data’ (Prainsack, 2019b: 2).[1] As Prainsack observes, individual control is often seen as an approach to counteract contemporary power asymmetries between the corporations as data users and consumers as data subjects. This counts for *MyData Global*, which since its origin in Finnish activism has become a central community for the specific kind of platform builders I study (Lehtiniemi and Ruckenstein, 2019). In practice, the two startups propose competing models of individuals that both offer services to make it possible for users to aggregate, monitor and share personal data held by corporate and state actors. While similar platforms exist within the region, DfG and Digi.Me are distinguishable on account of the explicit nature by which they narrate themselves as “the ethical choice”. The main research questions I ask are the following: What role does infrastructure play as Digi.Me and DfG promote themselves as ‘the ethical choice’? I examine this question in relation to three different audiences: The user, the market and the communities they engage in. This research design is chosen to acknowledge that the

---

[1] Others have employed the term ‘data sovereignty’ (Hummel et al., 2021) to elements of what Prainsack refers to with individual control.

promotion of ethical tech and enrollment of infrastructure might take on various roles and forms in different contexts.

By focusing on these platform startups, the article engages burgeoning developments in a field that in Science and Technology Studies (STS) is increasingly described as the ‘evolving health data ecosystem’ (e.g. Vayena and Blasimme, 2018; see also Sharon and Lucivero, 2019). With big tech vying to stake their claim in this ecosystem, critical scholars have encouraged further research on the values enacted by the contemporary platformisation of health (e.g. Sharon, 2018; van Dijck, 2018; Zuboff, 2019). This research gap has also been formulated for the type of platforms dealt with in this article (Riso et al., 2017; Lehtiniemi and Kortesniemi, 2017), which is increasingly recognised as “data intermediaries" in European policy (European Commision, 2020a: 10; 2020b: 1) and their communities (‘Event: Data intermediaries launch’). Data intermediaries are for instance championed in *The European Data Strategy* as the European Commission (2020a: 3) seeks a regional solution to the problem of ‘balancing the flow and wide use of data, while preserving high privacy, security, safety and ethical standards’. To this end, the data intermediaries are envisioned to deliver the tools and standards for ‘Empowering individuals to exercise their rights’ (European Commission, 2020a: 10).[2] These rights are chiefly enshrined in the European General Data Protection Regulation (GDPR) from May 2018 that with article 20 introduced a new right: *the right to data portability* that entitles individuals to request the portability of personal data from one data service to another in a structured, commonly used and machine-readable format.[3] Article 20 is thus considered as key to the ambition of European regulators to mitigate ‘lock-in’ effects that stifle European innovation and competition in the personal data economy (European Commission, 2020a: 10).

In order to answer the question, I draw on a theoretical framework that facilitates a combined analysis of how ethics is narrated with how ethics is technologically infrastructured. As with the frame, the analysis brings insights from infrastructure studies into conversation with critical data studies that provide descriptive accounts of how certain values are ascribed ethical importance through infrastructures. On a descriptive level, the article provides a contribution that situates the trend of ethics in the Nordic region and a demonstration of its manifestation through the promotion *and* infrastructural of the studied data intermediaries.

---

[2] The Smart4Health project is one example hereof that has received €21.781.120 of funding from the EU to “enable the citizen-centred EU EHR exchange for personalised health” (European Commission, 2021).

[3] To be precise, article 20 prescribes that data subjects are entitled to ‘receive the personal data concerning him or her, which he or she has provided to a controller, in a structured, commonly used and machine-readable format and have the right to transmit those data to another controller without hindrance from the controller to which the personal data have been provided’ (European Commission, 2016: 45).

The article provides insights into how advances in the name of ethics introduce new questions of ethical, societal and political importance that concern the role and responsibilities for state, market and citizen. The article argues that more ethnographic research is needed to illuminate the infrastructural practices that underpin the ethical claims put forward by tech actors in the data economy. The article demonstrates how infrastructures are also used in the promotion of ethical tech and argues that more ethnographic research is needed to illuminate the infrastructural practices that underpin the ethical claims put forward by tech actors in the data economy.

In terms of the article's progression, the next section lays out the conceptual terrain where the relation between infrastructure and ethics is established. The subsequent section accounts for the ethnographic data collection and introduces the startups in this case study. With the conceptual and empirical ground laid out, the following section unfolds the analysis concerning the self-promotion of 'the ethical choice' to user, market and peers. The article concludes with a summary of the case study and establishes the contribution and relevance of the article.

Infrastructure and the narration of ethics

As mentioned, the article proceeds to answer the main research question by adopting a dual focus on how ethics is narrated and technologically infrastructured. This approach has been chosen to transcend ethics as only discourse to also examine how technology is practically enrolled in the ethical claims encountered in the case study. As we shall see, such enrollment ranges from consent solutions through quasi-market safeguards to an app for "ethical data monetization".

There are two leading frames of analysis relevant for the study of values regarding the advance of platforms in digital health: modes of justification and a more normative framework of assessment. The former, discussed by Sharon (2018) in regard to the 'Googlization of health research', foregrounds an ethical terrain that allows for an analysis of which orders of worth are appealed to by different actors beyond the dichotomy of public and private gain. In contrast, the analysis of Riso et al. (2017), from a more philosophical perspective, puts forward a framework for which values *should* be considered in relation to ethical sharing of health data in online platforms. In order to explore what values Digi.Me and DfG advance, this paper draws on STS and social anthropology scholarship on infrastructures to scrutinise how the "ethical choice" is being built and narrated for different audiences. In addition, I explore how these infrastructures are being narrated and put forward for different audiences - a kind of infrastructural inversion of its own. Before proceeding to the analysis, I will first introduce the concept of 'infrastructure' and explain its relevance for this empirical analysis of values.

The sociotechnical perspective I propose here is in part informed by critical scholarship on infrastructure studies. Over the past four or five decades, infrastructure has become an important site of theoretical development and empirical investigation for STS, and more recently anthropology. From bridges (Winner, 1980) and roads (Knox and Harvey, 2015) to electricity systems (Hughes, 1983) and water pipes (Anand, Gupta and Appel, 2018), this scholarship demonstrates how technological constructs are always social and never neutral. These classic objects of infrastructures shape the conditions of possibility for where people, goods, power and basic rights can travel. In this sense, infrastructures can broadly be defined as 'matter that enable the movement of other matter' (Larkin, 2013: 329). Given the vital role of infrastructure, it is perhaps not surprising that the history of these technological objects are sometimes narrated to mark the coming of modernity itself (Edwards, 2002). Pushing back against the lure of technological determinism, the above-mentioned scholars however provide alternative accounts that illuminate how the promise of infrastructure is contingent upon social processes that include design, use and maintenance. It is however equally important to consider what lies beyond the promise of infrastructure, as 'One person's infrastructure is another's brick wall' (Star, 2002, 116). In this way, I explore infrastructures as a potential site of 'politics pursued by other means' (Larkin, 2013: 10, citing Latour, 2012: 38) to understand how ethics is enacted intentionally and otherwise (e.g. Hogle, 2016).[4]

The political dimension entailed in the infrastructuring data intermediaries is explicitly addressed as a discussion point in the closing of each section of analysis. Motivating this discussion is the paradoxical observation of how both startups describe themselves as 'neutral' and 'ethical' at once on the basis of distinguishable platform models. Concrete examples of what politics by other means may look like include accounts where claims to ethics are proposed, contested and defended among potential investors and peers to negotiate the cultural acceptance of their infrastructural platforms. Yet, infrastructures have - with notable exceptions (Larkin, 2013: 336) - a tendency to blend in the haste of everyday life until they cease to function, masking the politics they enkindle in the process (Star,

---

[4] This should already be evident considering some of the introduced shifts in what some term the 'health data ecosystem' (e.g. Sharon and Lucivero, 2019), for instance how the advance of GAFA enkindles questions of social and political importance.

1999). Responding to this observation, Bowker (1995) famously proposed the notion of ‘infrastructural inversion’ to bring what is commonly considered as background (or ‘second nature’) to the fore of analysis. Questions of how infrastructures gain form, value and authority are from this perspective to be answered on the basis of empiricism by assigning analytical primacy to the material manifestations. Throughout the analysis I look at how the infrastructure becomes foregrounded and enrolled by companies to illustrate their ethical credentials.

Fieldwork and case description

The empirical material was gathered between 2017 and 2020 in the context of a doctoral study. Data intermediaries has been one strand of this study, which I have explored to elucidate emergent practices and politics of personal health data in Denmark. Digi.Me and DfG were initially part of a provisional list of data intermediaries that caught my attention. Through participatory observation, interview and document analysis, the scope was eventually narrowed through three characteristics that led my attention to Digi.Me and DfG. First, both startups were community members of the MyData initiative. With *A Nordic Model for human-centered personal data management and processing* (Poikola, Kuikkaniemi and Honko, 2015), the MyData initiative spearheaded the movement for individual control on behalf of Denmark and its neighbouring welfare states. Second, Digi.Me and DfG were taking active steps to become operational in the Nordic region. The third and final characteristic regarded the manner by which they employed explicit claims to narrate themselves as “the ethical choice” for data sharing. Applying all three characteristics as an inclusion criteria, Digi.Me and DfG were singled out from the provisional list of data intermediaries as a unique phenomenon in the entrepreneurial space for data intermediary startups.

During the course of my fieldwork I have conducted empirical research at various events across the Nordic region. Most events were attended physically and studied through ethnographic observation at varying degrees of participation. Some events have been studied online through video recordings (synchronous and asynchronous) and other documentation (e.g. social media content, websites and email invitations), particularly those carried out in the wake of the covid-19 pandemic. Events attended by DfG (n=15) were typically limited to a single day and primarily hosted in the capital of Denmark by stakeholders in Danish health tech. Events attended by Digi.Me (n=5) were typically held over the course of three days, which include two *Nordic Health hackathons* in March 2019; One in Iceland, another in Finland. One place where both actors could be observed was at the annual MyData conferences where the community gathered to advance the idea of individual control through debate, technology and startups. Three annual conferences were attended between 2017 and 2020, which were primarily held in Finland with one a detour to Estonia. Semi-structured interviews (n=8) were conducted among the provisional list of data intermediaries and related actors in the startup community. One of these interviews was with the founder of DfG, who I call Sandra, which was conducted in her office in Denmark. Another interview was with a senior officer from Digi.Me, who I will refer to as Timothy, and was carried out during the hackathon in Iceland.[5] Informal interviews (n=23) have additionally been carried out among provisional data intermediaries and related stakeholders. These interviews have been conducted at various events where a total of nine can be traced to DfG (n=4) and Digi.Me (n=5) that include Sandra, Timothy and other affiliates. Finally, document analysis of promotional material, internal documents and grey papers also inform the empirical material. With the empirical material outline, it is time to introduce the two data intermediaries and articulate why they captivated my interest during my initial field encounter with them in 2017.

Briefly introduced, Digi.Me is a UK based company founded in 2012 by Julian Ranger, who is usually presented as a specialist in interoperability and military internet. In 2017, Digi.Me acquired the US based company Personal, which had since 2009 similarly worked to put individual consumers in control of their own personal data. The same year, Digi.Me announced that citizens in Iceland could via their platform - allegedly as a world's first - not only view but download and share their electronic health records by automated means (Firth, 2017).[6] It is this expansion across the North Atlantic ocean that brought Digi.Me into my study, as Iceland had become a testbed for entering the market of digital health in the Nordic region where self-access and sharing of patient records have previously enkindled ethical debate.[7] It is against this backdrop interesting to critically examine what Digi.Me means when they claim to deliver 'A new ethical and sustainable way for individuals to take control

---

[5] The job title has been revised following the examination committee's assessment due to GDPR considerations.

[6] Aside from Iceland, this service is currently available in the UK, US and Dutch market ('1,000s of sources of data. 1 SDK.', n.d.).

[7] In Denmark, self-access to view patient records was introduced in 2010 via a national portal, Sundhed.dk. Officially, the opportunity of self-access was to promote ideals of patient empowerment - and unofficially tend to needs of public accountability (Wadmann and Hoeyer, 2018). Others have however cautioned that self-access could potentially put patients with or predisposed to severe conditions in a disempowered position when insurance providers request a copy of their patient records. As a result, health professionals for such conditions have been reported to postpone screening appointments for uninsured patients and manipulate or censor patient records to protect patient rights (Rasmussen, 2010).

of their data and privately share with data-driven apps and services' ('What is digi.me?', n.d.).

Whereas Digi.Me is working to expand their operation to the Nordic Market, **DfG** was incubated *within* the region with a base in Denmark. After years of concept development, 2018 marked the year when the Danish startup was formally established as an industrial foundation. My first ethnographic observation of DfG was at a seminar in May 2017. The event took place in Copenhagen, Denmark, and was attended by various stakeholders interested in the exploitation of artificial intelligence and big data in Danish healthcare. Among those who sought to claim a stake in this agenda was the founder of DfG. As Sandra pitched her vision for the DfG platform to the audience, I found myself captivated by the assumed linkage between individual control and her ideas of personal and public value - and its explicit narration as something 'good'. As noted in a UK study of *Public perceptions of good data management* (Hartmann et al., 2020:3), 'good' can be a placeholder for any number of values and can even be 'seen to depoliticize "data relations" (Kennedy, 2016)'. The task of the ethnographer must then be to penetrate the *prima facie* veil of such positivity that 'good' connotes to critically explore how the notion is assigned meaning.

The two platform startups then suggested a confluence of values as ideas of individual control seemed to gain traction in the Nordic welfare states with a strong collective tradition. I soon discovered that these ideas were spearheaded by the MyData initiative that grew out of Finnish data activist community for open data.[8] With the establishment of MyData, personal data became a new frontier of data liberation, which in a white paper from 2015 was formulated as *A Nordic Model for human-centered personal data management and processing* (Poikola, Kuikkaniemi and Honko, 2015). DfG and Digi.Me pledged their commitment to the MyData declaration in 2017 that promotes individual empowerment as a means to build a 'fair, sustainable, and prosperous digital society' (MyData, n.d.).[9] As Prainsack observes, individual control is often by its proponents viewed as an approach to counteract contemporary power asymmetries between the corporations as data users and consumers as data subjects (Prainsack, 2019b). This observation rings true for a considerable segment of the MyData community (Lehtiniemi and Ruckenstein, 2019) where

---

[8] Other initiatives that promote individual control regarding personal health data includes the Blue Button Initiative in the US. Individual control is also promoted in the *The European Data Strategy*, as the EU looks for a regional response to the problem of 'balancing the flow and wide use of data, while preserving high privacy, security, safety and ethical standards' (EU commission, 2020: 3). This ambition has in part materialised in the Horizon2020 project *Smart4Health* that received a €21.781.120 funding form the EU to 'enable the citizen-centred EU EHR exchange for personalised health' (European Commission, 2021).

[9] Digi.Me signed Aug 24, 2017 while the DfG signed Sep 06, 2017 ('Declaration', n.d.)

Google, Apple, Facebook and Amazon - colloquially called “GAFA” - have somewhat become a ‘figure of the monster’ (Jasanoff, 2015; etc). This monster figure was especially pronounced in the first two annual MyData conferences that I attended in 2017 and 2018 as exemplified by figure below (see figure 1). At these events, GAFA was enrolled to epitomise the premise for the social movement of MyData Global and thereby articulate some of the monstrous characteristics of the status quo (Douglas-Jones et al., 2018) formulated in the MyData declaration: ‘the balance of power is massively tilted towards organisations, who alone have the power to collect, trade and make decisions based on personal data, whereas individuals can only hope, if they work hard, to gain some control over what happens with their data’ (MyData, n.d.).

@aureliepols

Some of you might remember

20

Figure 1. *Presentation from MyData Conference 2017 in the track GDPR - The new black*

This monster figure is also woven into the stories and promotional material produced by the platforms in the MyData community, often to make claims of moral and ethical superiority. During the annual conference of 2017, the French platform initiative Cozy.IO for instance delivered corporate cards with the motto ‘We can’t be evil’ written in large - a play on Google's former corporate code of conduct and motto ‘Don’t be evil’ - as a reference to their open source software. It is however more usual that self-narrated stories of platforms draw on the founder’s experiences to demonstrate trustworthiness, as the case for Digi.Me and DfG.

An infrastructural narration of ‘the ethical choice’

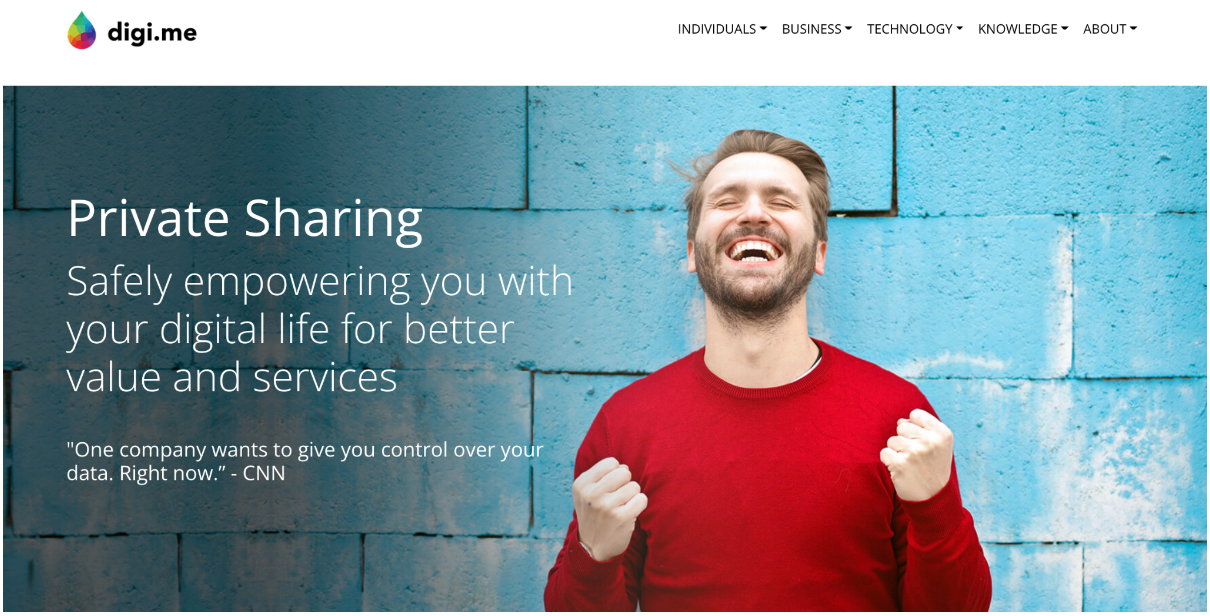
In this section, I analyse how Digi.Me and DfG narrate themselves as the ethical choice to three distinct audiences: individuals, the market and communities. For each audience, I analyse how infrastructure takes the form of ethics and discuss the social and political questions they engender.

The user: Ethical values in “private sharing” infrastructure

The aim of this section is to explore how Digi.Me and DfG promote themselves as the ‘ethical choice’ to the user. The subject in question is thus the ‘imagined user’ (Akrich, 1992) as the analysis draws promotional material and interviews from those who build the infrastructures. Underlying this notion is the basic tenet that it matters how infrastructure builders imagine who the user is imagined to be and act since this shapes what they build in terms of human-machine interaction. My use of the concept is however not limited to how the user is imagined for the purpose of design, but also promotion that has implications for what ethics is taken to be. In the empirical case, the imagined user is the data subject which both platforms varyingly refer to as consumers and citizens.[10] Although both startups share a number of similarities, there are observable differences in how they present themselves as ethical to users. One of the central messages that both platforms promote - and which I will use as my departure - is that personal data can be shared without compromising privacy if done correctly. Digi.Me for instance employs the term ‘private sharing’ as a prefix when talking about their user app and platform more generally. For those who conceptualise sharing and privacy in a zero-sum relationship the prefix will seem as an oxymoron. Digi.Me however insists that sharing and privacy are not mutually exclusive and even assert that it should be possible to ‘benefit from your personal data while *improving* your privacy’ (‘Video’, n.d., author’s emphasis). On the website of DfG, the Danish startup promotes a similar conceptual relationship between sharing and privacy with the assertion that ‘It must not become a question of EITHER / OR because we as a community need to capitalise on the enormous value created in using data in the right way, for the right purpose!’ (‘How do we secure trust in handling personal data?’, n.d.). Moving forward, the analysis illustrates how ethical values of user protection and user agency are being built and narrated to reconcile data sharing with privacy preservation. This is illustrated through a mutual enrollment ethics and infrastructure communication towards the user through 1) technical architecture, 2) governance model and 3) consent solutions.

---

[10] Third-party users will be specified accordingly when mentioned.



Private Sharing

Safely empowering you with your digital life for better value and services

"One company wants to give you control over your data. Right now." - CNN

Figure 2. Digi.Me promotion of empowerment[11]

Technical architecture: “We are never able to touch, hold or see user data”

In the preliminary case description, I introduced how Cozy.IO claimed that they ‘can’t be evil’. In this example, the open source software is a way to support this ethical claim while the claim itself draws attention to the infrastructure. The Digi.Me platform is also promoted with a pledge of an ethical code of conduct: ‘we are never able to touch, hold or see user data – or tell people how or where to use their data’ (https://digi.me/mission/). In contrast to Cozy.IO, the privacy-by-design principles highlighted by Digi.Me’s pledge is not one of transparency to enable public scrutiny. Instead, the pledge of Digi.Me highlights a principle of internal restriction where privacy compromising conduct is considered to be eradicated by the capabilities of its technological infrastructure. In order to get an idea of how this arm's-length principle works, it is helpful to introduce how users are imagined to engage with the platform.

First, the user must download Digi.Me’s “private sharing app” on their mobile device, free of charge. Once downloaded, the next step is to select a personal cloud storage from the following choices (see Figure 3): Dropbox, Google Drive or Microsoft OneDrive. This is one aspect of the arm’s-length principle, as the personal data is outsourced to a third-party for safekeeping. After choosing a storage solution, the user decides which data to import and normalise from a list of API integrated consumer services (e.g. social media, wearables, music, banking and medical records). Drawing on the professional background of the

---

[11] https://digi.me/ (Accessed 11 Nov, 2021)

founder, Digi.Me ensures the user that the imported data is protected with 'military grade encryption' (https://digi.me/get-data/) where it can only be browsed or queried by the user. The user can however also opt to 'privately share all or parts of that data for service, convenience or reward', but only when an explicit consent is provided from the user (https://digi.me/privacy-policy/). The consent service of Digi.Me is thus another measure of user protection designed to keep the circulation of personal data aligned with personal preferences. This mechanism and related idea of alignment is at the core of the model of ethical data sharing that Digi.Me puts forward, which will be elaborated on soon. Until then, it is sufficient to highlight how consent, as a mechanism to respect personal preferences, is closely linked to the last part of Digi.Me's pledge to 'never[...] tell people how or where to use their data'.

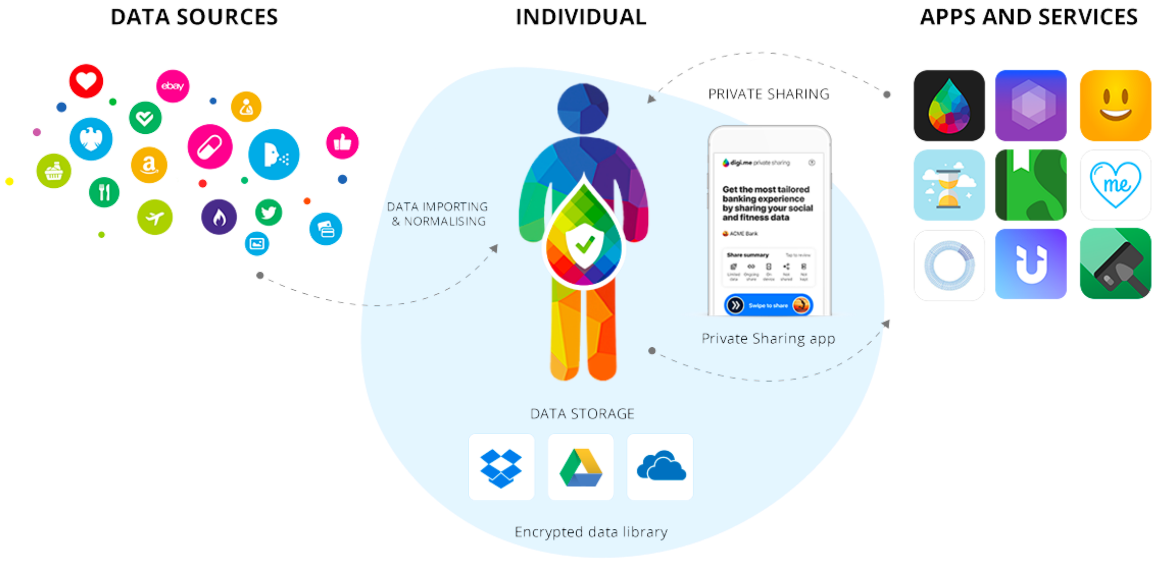


Figure 3. Image from a presentation by Digi.Me used to introduce their platform concept at a Hackathon event in Helsinki, Finland, from March 2019.

Governance model: "A unique legal structure"

The basic idea of DfG is quite similar to Digi.Me: 'to create a safe "harbour" and digital space (Personal DfG), where we as individual citizens have the opportunity, not only to gain visibility and control over our data, but also the ability to share the data when it makes sense' (https://dataforgoodfoundation.com/en/what-why). Whereas Digi.Me promotes the ethical value of user protection by highlighting *technological* elements of its infrastructure, DfG promotes the same value by foregrounding its *legal* elements. Included in the latter, is the governance model inherited from the industrial foundation constellation, which is on the company website of DfG described as a ‘unique legal structure’ with ‘the ability to safeguard the individual's privacy’ (Digi.Me What Why). Before I elaborate how the legal infrastructure of DfG relates to user protection, a brief detour is helpful to outline a few preliminary differences between the models of DfG and Digi.Me in their shared pursuit of individual control.

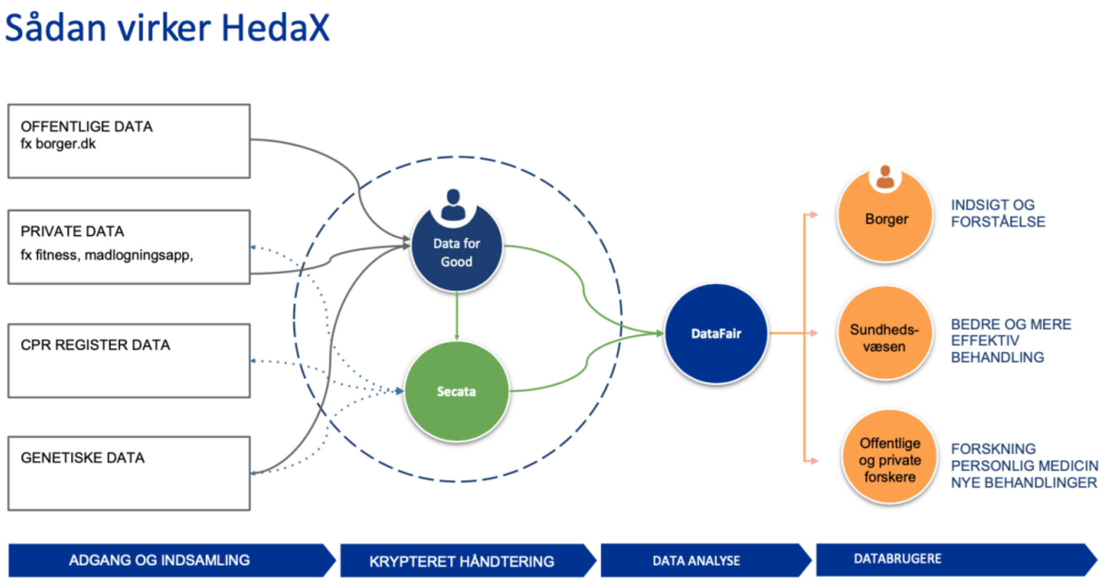


Figure 4. DfG “ecosystem” diagram for the HedaX project (Broch, 2020, translated by author)

Whereas the Digi.Me platform has been operational for several years, DfG received its first grants to develop the technological infrastructure in 2019 and 2020 from the Danish Innovation Fund amounting to €3 million. More than two thirds of this state funding was allocated to a personalised medicine project for cancer called HeDaX (abbreviation for Health Data Exchange) that was funded to ‘put Denmark in the Yellow Jersey[12] regarding the Collection of Health Data’ (The Danish Innovation Fund, 2020). The HeDaX project illustrates how the platform model of DfG differs from the one proposed by Digi.Me on a strategic and operational level. The first difference can be located in the route to market strategy where DfG emphasises the importance of cross-sectoral *collaboration*. HeDaX is for

---

[12] This is a reference to the sport of cycling where the yellow jersey is worn by the overall leader.

instance a public-private partnership of 10 actors that include a patient organisation, research institution, medical company and the capital region. The second difference is visible in data infrastructure for this 'ecosystem' (see figure 4) as core services are distributed among three Danish startups. The technical operation of DfG is limited to the personal DfG, which includes services for data integration, internal storage and consent management. Personal data imported to the Personal DfG are then handled by *Secata* and *DataFair.* The former provides "secrete sharing" encryption technology using Secure Multiparty Computation and Blockchain technology while the latter produces analyses for the health system, researchers and the medical industry. With the introduction of this preliminary outline, it is now time to resume the focus on the "unique legal structure". In her attempts to specify the value proposition of DfG from the roles now occupied by Secata and DataFair, Sandra also articulates what it is about the industrial foundation that sets DfG apart from competing initiatives that operate as regular corporations:

> Well, what we offer is in essence two things. Looking at DfG in isolation, we offer the citizen protection [...and] the right to control their own data. This is also why we are nonprofit. And if we in the future are founded as a cooperative or a foundation is not so important[...] What is crucial is that it is the citizen that in effect owns us.

During the time of the interview, it was undecided whether DfG should be established as a cooperative or a foundation. Regardless, Sandra had identified that the startup needed to be architectured with a different model of profit distribution and ownership than those prevalent among some competing initiatives. A first step to elaborate the differences between industrial foundations and regular corporations, is to note that the former can minimally be understood as 'an organization created to administer a large ownership stake in a particular company' (Heydemann, 2006: 237). One hallmark feature of industrial foundations is that they have no owners, hence why they sometimes are referred to as 'self-owning institutions' (Heydemann, 2006: 238). The citizen ownership model articulated in the quote from Sandra is in this respect technically more relevant for cooperatives (e.g. Hafen et al., 2014). Industrial foundations are however nonprofit entities that are in the case of DfG steered by a self-elected board of directors who are aside from the law governed by a formulated charter of the foundation. The charter is a governing instrument intended to represent the will of the founder, which defines the organisation and its social purpose. In the case of DFG, the defined purpose is 'to promote knowledge, dissemination and use of data-based development, health promotion, prevention and disease management, thereby contributing to development, growth and public health locally, nationally and globally' (PricewaterhouseCoopers, 2021: 11). As an industrial foundation, it is technically possible for

DfG to earn profits but they can only be distributed for charitable purposes (Heydemann, 238).

Now that we understand *what* Sandra considers important about the “unique legal structure”, it is time to address *why* they are afforded such status in regard to user protections. In our interview, Sandra describes the industrial foundation structure as a legal precaution and assurance to protect citizens against adverse market dynamics. In her own words, she explains her thinking while contrasting DfG to competing initiatives such as PatientsLikeMe, HealthBank and Digi.Me:

> [...]We just want to invert it, so it is not the commercial[...] that is the actual driver. Because I think it is dangerous what they are doing. [...]Not to say that it will happen, but there is a high risk that it can develop unethically. And I think that would be a great shame, because I consider the initiatives and ideas, and what they can do, to be super great.

One company highlighted to illustrate such ‘danger’ is the patient experience sharing platform, PatientsLikeMe. What makes this PatientsLikeMe 'totally commercial’ from Sandra’s perspective is not merely that they are a for-profit company, but more precisely the terms of consent provided to users. According to Sandra, once patients share their experiences and data with the platform they effectively enter a corporate regime of ownership that waives the users’ right to exercise meaningful control. This case also features in scholarship on the political economy of hope where it is used to highlight ‘a potential conflict between commercial imperatives and patient welfare’ (Petersen, 2015: 92). The tension in question is specifically articulated in relation to the possibility of information being sold to ‘partners (i.e., companies that are developing or selling products to patients)’, as stated on the website of PatientsLikeMe (see Petersen, 2015: 92). As elaborated in the following quote, establishing DfG as an industrial foundation is for Sandra a way to prevent a scenario where the “wrong values rule” (23andMe is an infamous example hereof, which is introduced in the third section of analysis):

> So if we were a private limited company then we would be able to go out and sell this data. But because we are going for a foundation construction, then it is a part of our charter to safeguard the citizen. The citizen’s rights. And there it is specifically noted […]what our profit should go to, […]distribution principles and all that.

Whereas a private limited company might be tempted to ‘go out and sell this data’ to appease shareholders, DfG are structurally architectured to be driven by a social purpose where eventual profits are earmarked for charitable purposes. The ethics of DfG are in this instance materialised in a legally binding charter as principles to ensure that ethical values can endure the test of time without being compromised by market pressures.

Consent solutions: “complete transparency” and “full control” (Digi.Me)

In the two previous examples, I have highlighted how technological and legal elements of infrastructure are used to promote the ethical value of user protection. In this last example, I review the consent solution of Digi.Me that combines both elements to promote user protection and user value as two sides of the same coin: the self-management of privacy and self-disclosure of information (REFS). In my formal interview with Timothy, the topic of consent emerged when I questioned how ‘ethically sourced data personal data’ was to be understood, since it is used to promote the ‘private sharing SDK’ (Software development kit) on their Digi.Me’s company website (https://digi.me/sources/). This statement, I am told, is also part of the company mission statement after which my informant explains how its motivation can be found in the current regime of consent that is described as a problematic element in the current personal data economy:

> Even when consent is obtained, it's not always that transparent; it's buried in the small print. You don't actually see what data you're sharing or what people are holding. And half the time you are just passing through cyberspace and people are scraping it as you go through and sort of processing and analysing what you're doing without you being very aware of what's going on.

For Timothy, a key problem with the status quo lies in the practices by which consumers are tracked, profiled and targeted in the dark where consent is either absent or obtained on an ill-informed basis. The practices described by Timothy have of course also been problematised in critical scholarship, leading some scholars to question the legitimacy of online consent as a general framework for the exchange of personal data (Zuboff, 2019). The quarrels that Digi.Me has with consent lies not in its legitimacy as a framework, but its current implementation. As Timothy explains, Digi.Me strives to create an alternative to the status quo by improving the terms of consent for their users:

> And what we're trying to do is make it [consent] very explicit and transparent. So, because you hold all the data, if you have a relationship with a business or an organisation, our consent access system allows that organisation to register and say; who they are, what data they want, for what purpose, so what's the value proposition, whether the data will leave the device, uhm, whether they will share it with third-parties, if it does leave the device […]And whether they’ll implement the right to be forgotten. And so then, when the user interacts with that service, it pops up on the

> phone, and then this summary of those six questions is then displayed to the individual and actually because they [individual users] hold the data before they share it, they can actually browse all that data and decide whether they want to share it and if they explicitly consent. Only then can that [third-party] app have access to the data. So, the individual has got complete transparency and control, and the organisations are handed into a very explicit agreement with the individual, which is very transparent and it's consent-based. […]we're trying to be more ethical in terms of how we do… it is... We're trying to simplify that process and make it as transparent and obvious to the individual, and make it easier for businesses to have that transparent relationship so that it is a more ethical process.

What makes the private sharing services of Digi.Me ‘ethical’ in Timothy’s view premised upon the assumption that consumers have ‘complete transparency and control’. Regarding transparency, the terms of consent are displayed as a summary of ‘six questions’ with the additional opportunity to browse the data in question before “swiping” consent.

While the summary notice may be more digestible than the ‘small print’, critical scholars will contend that it is a stretch to claim that it delivers ‘complete transparency’. Although the achievement of simplicity and clarity are commendable goals, they are in principle competing values and therefore inherently bound to result in a loss of fidelity. What this ‘transparency paradox’ (Barocas and Nissenbaum, 2014: 58) highlights is the critical insight that transparency (e.g. Tsoukas, 2003) and informed consent (Hoeyer, 2009) must both be considered as ideals. Seen as such, it is an overstatement to claim that the consumer is in ‘full control’ of their personal data, which additionally assumes that consumers understand the data in question and how they might be repurposed through aggregation and analysis (e.g. Solove, 2013). The problematic term of ‘full control’ will be elaborated in due course, but the ability to browse data prior third-party sharing does add another dimension of transparency. In any case, from Timothy' quote it is worth explaining that when Digi.Me claims to be the ethical choice, they do so from a realist perspective more than one of philosophical abstraction: they are ethical in comparison with the dominant practices in the current personal data economy.

Interim discussion: Ethics as empowering users

In the foregoing analysis, I explored how the startups draw on infrastructural credentials in the narrations of ethical values - user protection and user agency - to reconcile data sharing with privacy preservation. In terms of user protection, the ethical claims put forward by Digi.Me highlight the technological affordances of their privacy-by-design infrastructure whereas DfG foreground the affordances of their ‘unique legal structure’ as means of quasi-market regulation. A combined promotion of user agency and user protection is for both startups advanced in relation to their consent solution, but examined in relation to Digi.Me. A digestible notice and the option to browse data before sharing were articulated to appeal to ideals of transparency and control to facilitate an ethical consent agreement between users and third-parts. Each example highlights the role of infrastructure in the narration of an imagined user empowered to regain control over their personal data.

While both startups promote a narrative of empowerment it is a particular version of control that they offer to users. The suggestion that users can strengthen their control over personal data through data intermediaries is quite counterintuitive if the main objective is user protection. Sure, data intermediaries hold promise in terms of providing the infrastructural tools for users to exercise their data rights. The majority of these rights are however only supported once data is imported to the platform. While the right to data portability can help import data from other services, the data will in many cases be ‘copied’ (as opposed to ‘moved’) since the right to data erasure cannot be assumed to follow when a legitimate cause of retention is present (more on this later). In order to (re)gain ‘control’ over personal data, users would then be moving what will often be a copy to *another* digital space owned by Dropbox, Google, Microsoft or self-hosted storage solution - thereby exacerbating the risk of a security breach.[13]

Another concern for risk-aversive users can be found in the implications of “private sharing” or “secret sharing”. Autumn 2018 marked the end of the Swedish platform *Hälsa for Mig* (‘Health for Me’) that aspired to provide citizens a storage space for personal health records with option of third-party data sharing. The discontinuation of the state-funded initiative was in part due legal concerns regarding the data protection regime in the offshored data storage location (Ny Teknik, 2018), but the platform was also a concern for what it aspired to achieve. This latter point was voiced by the Swedish Data Protection Agency in relation to the unfathomable consequences of sharing sensitive personal data to a potentially endless array of third-party services (Ny Teknik, 2018). These consequences can be identified by the dispersion of copies to additional spaces, but also the uncertainty regarding the future use of such data (Lehtiniemi, 2017; Solove, 2013). While the startups claim to be neutral intermediaries, the infrastructures are explicitly built to reconfigure the balance of agency between data holding organisations and data subjects in favour of the latter. In doing so, it is

---

[13] One general observation worth mentioning from my fieldwork at MyData conferences regards the manner in which risks are articulated in different tracks. In business oriented tracks startups were on multiple occasions observed to describe the security credentials of their infrastructures in extreme terms (e.g. “100 percent anonymous”). The technical tracks were by contrast more acceptant of a reality in which risk will always be a factor.

however not only the agency that is delegated to the user, but the responsibility too (Akrich, 1992) that notably includes the self-management of privacy.

The Market: Ethics as a competitive advantage

Moving on from the user, this second part of analysis explores how infrastructures are enrolled as the startups narrate themselves as ‘the ethical choice’ to the market. This question is answered in three steps. The first step situates how ethics is being championed and embraced in the Danish tech industry with a narrative of ethics as competitive advantage. The second step establishes how Digi.Me and DfG embrace this corporate narrative of ethics, and locates the market they seek to enter. The third step analyses how both startups draw on their infrastructures to make them distinguishable in the located market. The analysis is concluded with an interim discussion that critically engages the claim to neutrality put forward by both startups on the basis of the foregoing analysis.

Situating ethics

In the introduction of this article, I highlighted how ethics has been suggested as ‘the hottest product in Silicon Valley’s hype cycle today’ (Metcalf, Moss and Boyd, 2019: 449). The corporate embrace of ethics in tech is also reflected in *Harvard Business Review* with recent publications such as *Thinking Through the Ethics of New Tech… Before There’s a Problem* (Ammanath, 2021), *Building an Ethical Company* (IH Smith and Kouchaki, 2021) and *How to Manipulate Customers … Ethically* (Sanyal, 2021). While it might not be the ‘hottest product’ in Denmark, I have since my fieldwork began in 2017 observed a growing interest in ethics from actors in the Danish tech industry. Part of this observed growth may perhaps be attributed to my gradual immersion in the field I study (you find what you seek) where the politics of data was a central part of my research design. Be this as it may, there is a particular narrative that reappears in my fieldwork that provides a market-oriented argument for the corporate embrace of ethics as a social responsibility.

One actor that has been particularly vocal in their advocacy of this narrative is *DataEthics*. Founded in 2015, DataEthics is a not for profit politically independent organisation that describes itself as a ‘ThinkDoTank’ based in Denmark with an international outreach. I encountered this organisation during the first year of my fieldwork at which time they were affiliated with DfG. DataEthics is also a trusted member of the MyData community that I have encountered during participatory observations at various events, including MyData annual conferences. An early articulation of the corporate narrative I spoke of can be found in a book publication from the ThinkDoTank entitled *Data Ethics – The New Competitive*

Advantage (Hasselbalch and Tranberg, 2016). Under the heading *The data ethical paradigm shift,* the authors pencil a parallel between the contemporary demand for ‘eco-friendly’ corporations in their prediction for data ethics: ‘Being eco-friendly has become an investor demand, a legal requirement, a thriving market and a clear competitive advantage. Data ethics will develop similarly – just much faster’ (Hasselbalch and Tranberg, 2016: 9). The narrative is that an ethical paradigm shift is afoot that rewards those who embrace ethics.

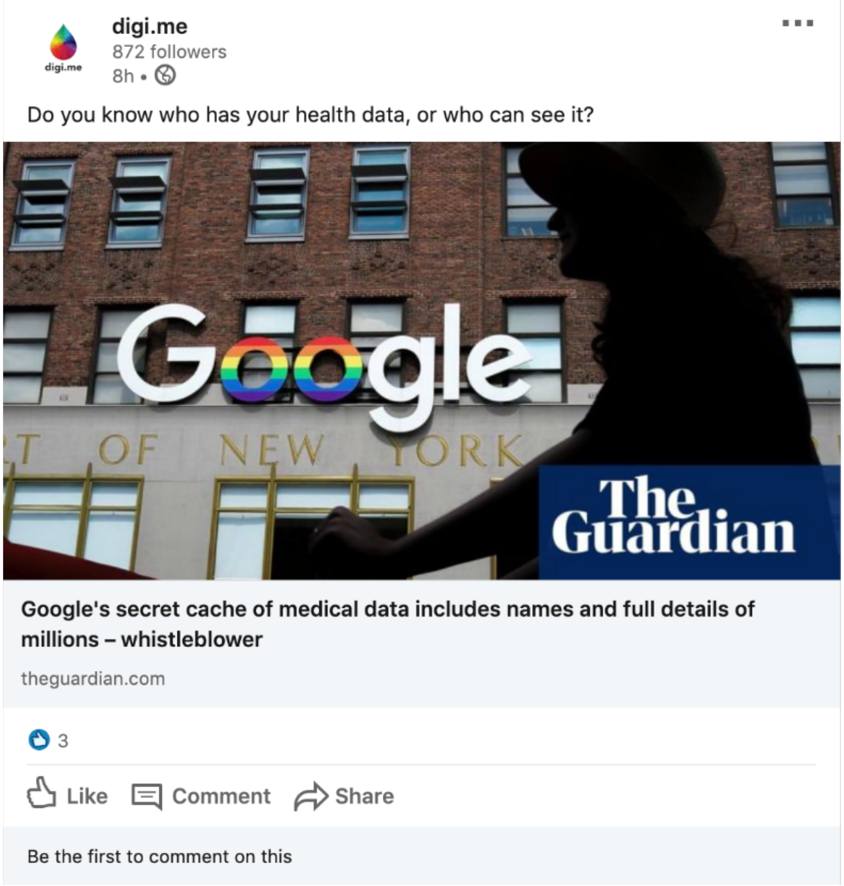
This corporate framing of ethics as a ‘competitive advantage’ has in recent years been taken up by national regulators in Denmark. In February 2019, The Ministry of Industry, Business and Financial Affairs (2019) announced eight initiatives of ‘digital responsibility’ to make corporate Denmark more trustworthy and thus ‘give them a competitive advantage internationally’ (Ministry of Industry, Business and Financial Affairs, 2019). Central initiatives include the establishment of an independent data ethics council, mandatory reporting on data ethics policies from large firms and a data ethics labelling scheme to strengthen consumer awareness and incentivise responsible corporate conduct. The various fundings granted to DfG from state bodies is also testament to the Danish political climate on ethics, along with the collaboration with Confederation of Danish Industry for co-hosting the first *Data for Good Summit* in 2020. While these initiatives do not validate the prediction of an ethical paradigm shift, they do demonstrate political buy-in to the ideas promoted in the narrative. But how do Digi.Me and DfG engage the narrative?

Locating a niche market

Exploring Digi.Me’s online presence reveals how both platform initiatives subscribe to the narrative that an ethical paradigm shift is afoot in tech. This is for instance articulated in the Digi.Me blog post *Is personal data ethics the new environmental concern? Yes it is!* (Firth, 2016b), which is a direct commentary on the aforementioned publication from DataEthics (Hasselbalch and Tranberg, 2016). On Digi.Me’s corporate website, startups and growing businesses are invited to ‘join the race to the top’ by becoming a part of the private sharing app ecosystem (https://digi.me/startups-and-growing-businesses/). Powered by Digi.Me, these actors are promised to have a chance to ‘build the next unicorn[...]’ and be a part of an alternative since ‘the current “race to the bottom” based on surveillance and exclusive corporate ownership of data must change’. Together, they can become ‘The ultimate weapon of mass disruption’. With these statements, there is a lot to suggest that the rhetoric of ethics is a strategic component in the marketing of Digi.Me's platform. Speaking to my interview informant, Timothy explains how this was not always the case. During early startup, back in 2009, Digi.Me was called *SocialSafe*, which is today the name of one the products available in Digi.Me’s ‘app ecosystem’*.* ‘The value proposition then’, he recalls, ‘was essentially that you can connect to your social media accounts and download and backup all your data'. Over time, however, they came to the realisation that the setup could be expanded to include other types of data and moreover made available for third-party sharing. It is during this journey of creating a 'private sharing SDK', I am told, that ethics arises out of a series of 'fortunate and coincidental decisions'. Ethics was then an addendum that coincided with an infrastructural and market expansion to claim a stake in the personal data sharing economy.

As mentioned, DfG and DataEthics are mutual acquaintances. Both have a history of attending the same events, including those hosted by each other. In September 2016, Sandra pitched the early ideas of DfG at the annual DataEthics conference entitled *Data control - monopoly or individual control?* The conference was notably attended by Margrethe Vestager, acting as European Commissioner for Competition, who championed the forthcoming right to data portability as a 'a precondition for competition being a thing' (REF video link). In September 2020, DfG hosted its first annual *Data for Good Summit* in collaboration with the Confederation of Danish Industry. The ability to attract Denmark's leading business organisation as a collaboration partner is an indication of the gradual solidification of DfG as well as the corporate narrative of ethics. The declared aim of the event was 'getting Danish firms, authorities and organisations to use data responsibly as a long-term strategy to give Denmark an International competitive advantage' (Danish Industry, n.d.). One of the ways responsible data conduct is promoted by issuing a "Data for Good Award" to one company and one person. Lego won the company award for its digital responsibility, data protection and privacy for customers and partners and transparency regarding the use and flow of data. The remaining award was handed to an individual from DataEthics due to her promotional and educational work in regard to responsible data usage, and contribution to the activation of digital data for the benefit of the individual and the public.

Through blog-posts or awards ceremonies, both actors then take active steps to affirm the narrative of an ethical paradigm shift in which they seek to participate. But where, more precisely, do the platforms locate a market for ethics in regard to the infrastructures they propose? To answer this question, I take a closer look at how the two platforms position themselves in relation to established infrastructures governed by state and corporate actors.


digi.me

872 followers

8h •

Do you know who has your health data, or who can see it?

Google's secret cache of medical data includes names and full details of millions – whistleblower

theguardian.com

3

Figure 5. Screenshot of Digi.Me's online presence on LinkedIn

Paying attention to how positional work is done through narration reveals how both actors describe themselves as a way to unleash the potentials of data that is currently held back by public scepticism and archaic systems. The public scepticism rooted in a perceived crisis of ethics induced by the conduct of corporate actors *and* public actors alike. This instance is articulated by Sandra in the media article *From Big Tech Data Control to Individual Data Control* (Data for Good Science, 2021) in reference to a health data scandal - implicating a Danish authority (Statens Serum Institut), Facebook and 263 pregnant women who wondered how data voluntarily obtained for public health research ended up in the US for commercialisation interest: '"How can we trust public authorities, when they can't live up to data laws or plain data ethics?[...] We need a safe, trustworthy and transparent platform, where everybody can control their data and decide what data should be used for, and that is what DfG can deliver"'. In our interview, Timothy similarly noted how the British NHS system had been involved in 'some real catastrophes' with damaging implications for public trust. In the interview, it is the care.data programme that is provided as an example (e.g. Vezyridis and Timmons, 2017) - a longer list of other examples can be found by reviewing Digi.Me's social media activity, which recycles the news of such catastrophes to make it visible to market actors why change is needed - and how Digi.Me might be the ethical choice (see figure #). The problem with the erosion of public trust, Timothy continues, is that it makes individuals less inclined to share their data. This was noted to be particularly problematic for ‘one size fits all’ systems - what scholars understand as ‘blanket’ and to an extent also ‘broad’ models of consent - where individual autonomy is largely limited to opt-out rights exercised with ‘one tick’. Why? Because people who might be willing to share data for a subset of purposes, or with a subset of actors, may opt-out altogether. The niche market identified by the startups then hinges upon two main shortcomings. One, an inadequacy of trustworthy data sharing infrastructures in the public and private sector alike. Two, a lack of flexibility in the traditional models of consent that in effect produce an untapped potential of otherwise willing data sharers. With cross-sectoral data sources and dynamic consent management services, the two startups then look to assert themselves as a new infrastructural actor that can be trusted to deliver ethical data access.

Ethics ‘as a service’

In narrating what they take ethics to be, both startups highlight regulatory compliance with the GDPR as a central frame of understanding. Taking regulatory compliance as the yardstick for ethical conduct however raises a relevant question when held against the corporate narrative of ethics. If ethics is a competitive advantage but mandatory for everyone, how does one distinguish oneself in the market? In what follows in this section, I show how the two startups draw on their infrastructural capabilities to tap into a particular market.

In the run up to its implementionation on May 25, 2018, the GDPR had already forced the hand of many corporations to assess and improve the organisational state of data management. Included in the assessment are regulatory principles of data minimisation, storage limitation, purpose limitation, which stipulate that personal data should only be collected, stored and processed insofar it is necessary for a specific purpose. Impetus to take such action was notably applied with the newly introduced threat of fiscal sanction where regulatory violations could amount to €20 million or 4% of the global annual turnover. To understand the implications of the GDPR for organisations in Denmark, I attended a number of conferences throughout the course of my fieldwork. At one conference in late 2017, a presenter from a large bank actor in Denmark and parts of Europe described the GDPR as a “monster” due to the scale of work required to ensure regulatory compliance. Compiling a gap analysis to get started was in itself accounted to require 3500 interviews in different countries and respectively 20000 and 15000 hours in internal and external resources. The price for cleaning up what should no longer be present in the IT systems was estimated to be north of €26 million. Given the scale of the project, the presenter shared how the company worked according to a “risk-based approach” so resources could be prioritised for critical areas most likely to damage the corporation if exposed. What this example conveys is two things. First, although companies may recognise the need for better data management, its regulatory enforcement has for many companies been considered a *burden*. Second, even when efforts are allocated to improve the organisational state of data management *full* compliance is in the short-term not always easily achieved or necessarily the target.

As it often plays out in the market, one company’s problem is another company’s opportunity. Among those who readied themselves to profit from the strengthened enforcement of data protection and privacy (e.g. legal consultancies) were data intermediary startups like Digi.Me and DfG. At the annual MyData conference of 2017, an entire track was assigned to the GDPR throughout the three-day event that in its description heralded the impending regulation as “the new black” and a “commercial opportunity” (REF). Turning the attention to Digi.Me, we can begin to see what such an opportunity looks like. With potential access to thousands of data sources from one software development kit (SDK), the value proposition for third-party developers is according to Digi.Me’s website summarised as: ‘Let digi.me do the hard work of integrating, normalising and maintaining data sources while you innovate’ (https://digi.me/sources/). Aside from the ‘1000s of sources of data’ readily available for reuse, Digi.Me’s service is additionally promoted as 'Simply the best way to comply with GDPR' (https://digi.me/give-data/). GDPR compliance is in this sense branded “as a service”, if you will, as third-parties are offered relief from the burden that they may encounter elsewhere (e.g. from some data brokers).

This way of enrolling the GDPR for branding purposes is rife among numerous data intermediaries in the MyData community, including DfG. Visiting the LinkedIn company profile of DfG, reveals ethics is explicitly connected to regulatory compliance: ‘DATA for GOOD is to ensure privacy and at the same time enable commercial and societal value creation by the use of intelligent data mining (big data) in an ethical responsible set-up (compliance)’ (https://dk.linkedin.com/company/data-for-good-foundation). In our interview from 2017, Sandra elaborates how the value proposition DfG can offer to third-parties is first and foremost “legitimate access to data, insight or knowledge - for commercial actors, public players and everyone”. The strength of this legitimacy is according to Sandra grounded in the legal structure and purposes of the foundation as well as the infrastructural mechanisms that support the securement of explicit consent from data subjects. For Sandra, the stakes are high “because the future social economy relies on how good we will be to extract value out of the data”. Done in the right manner, she predicts that health data can become Denmark’s next “North sea oil” venture. Part of what both startups offer to the market is a service of data access that relieves the burden of compliance for third-party actors, but also legitimate access that is a point I return to in the interim discussion.

What both startups propose to third-parties is however not merely an easy way out to comply with the GDPR and secure legitimate access. Digi.Me and DfG look to persuade the third-parties that they can offer a superior product to what the established players can deliver in the current market for data exchange; not only from an ethical perspective, but also a utilitarian one. In 2019, the founder of Digi.Me, Ranger, explained the thought process behind this claim to the readers of the business magazine *Forbes*. In outlining his grand vision for a ‘decentralized individual-owned personal data grid’, Ranger explained (Sherriff, 2019: para 8-9):

> The prevailing narratives around personal data security today all miss the key point – we can and should do more with personal data to enable innovation. [...]However, the problem is that the current personal data internet architecture – even with illegal tracking and violations of privacy – still cannot gather the right data together to support all the new opportunities available to us. The only person who knows all about me is me – only I can gather my data from my health sources, my banks, my social media, and other sources.

What Ranger wants to build is an infrastructure that offers third-parties self-curated data profiles on individuals from all sectors of society. Through this model of individual data sourcing, Digi.Me promotes the idea that they can deliver ‘better data than Google et al because it can be wider and deeper and should remain a lot more accurate’ (https://digi.me/sources/). DfG advances a similar claim on its website: ‘We offer better, faster and cheaper access to relevant data and insights for research organizations and private stakeholders without compromising the individual privacy’ (https://dataforgoodfoundation.com/en/what-why). With a vision to transcend established silos between sectors, organisations and nations, both startups strive to deliver combined access to data sources of established, emerging and experimental relevance for health and well-being (Vayena, 201x).[14] In my interview with Sandra, she describes how her ‘holistic’

---

[14] What is recorded about the lives of individuals, including health records, can however also be of value for purposes *beyond* digital health. This can be illustrated with the Digi.Me supported app UBDI, which allows users to monetise anonymised insights from this life record for market research. For market actors interested in health, insights can be derived from health records or inferred from anything from fitness habits to food purchases. The partnership with UBDI was by Digi.Me announced in early 2019 as a ‘data ethical monetization programme for consumers' (Firth, 2019). One dimension of this ethical claim rests on the infrastructural credentials of Digi.Me's ‘private sharing’ platform detailed in the previous section.

conception of health data can include informational sources on taxation to insurance claims 'because all forms of behaviour is health’. For Sandra, behavioural insights can ‘reveal a lot about[...] how good my life is and it is the quality of life that is ultimately decisive for my ability to contribute to the society and community’. This definition of health data is also shared by Timothy in relation to Digi.Me’s cross-sectoral platform and his personal belief: ‘We work across all data. And actually, from my personal point of view, I believe that all data is health data, because your Tesco supermarket data, for example, about what you've been buying, and the particular products tells a huge amount about your health. So, we are about this longitudinal record - Apple doesn't do the longitudinal record across all your life.’ The value proposition from both actors to third-parties is from a utilitarian perspective then not access to *health* records, but a self-curated record of *life* itself (Rose, 2009; Clarke 2010); all verifiably interlinked to a single individual.

Interim discussion

In the beginning of this section, I described the emergence and traction of a narrative concerning an “ethical paradigm shift” in tech that would reward those who embrace ethics. Just as the market makes demands on companies, I suggest that companies also use ethics to sell themselves back to the market. Based on the empirical analysis, ethics can more specifically be said to comprise part of the route by which the infrastructures are envisioned to deliver market value.

While MyData emerged from a grassroot movement in data activism, it is not always evident whose interests are ultimately reflected in the promotion of individual control (see Lehtiniemi and Ruckenstein, 2019). This ambivalence was for instance articulated at the 2018 MyData conference, which took place three months after the GDPR was adopted in May the same year. At the previous annual conference, the new right to data portability was the cause of tremendous excitement in anticipation of its introduction. While much of this energy was brought forward to the subsequent conference of 2018, the initial excitement entered a gradual decline as the event unfolded. This decline was particularly observable during a panel session on adoption of the new right as the “report from the field” (from various actors that include data intermediaries) generally delivered underwhelming results. On the screen with crowdsourced comments from the audience, there was one entry that seemed particularly incisive since it addressed the fundamental question why individuals should use data intermediaries: ‘Isn[‘]t the challenge of portability in practice more of a problem to organisations wanting to provide services to individuals than for individuals wanting to control their own data[?]’. What the author of this comment alluded to was a question of why individuals should be concerned with the portability of their personal data and who benefits?

The question of ‘*cui bono’* (to whom is it a benefit) has been an important heuristic in STS to elicit the politics of infrastructures (Star, 1990: 43). During my interview with Sandra in 2017, I asked my informant to assess the demand for individual control to which she immediately replied that it was “huge”. Interestingly, this demand was stated to be registered among private and public actors (those who look to benefit from enhanced data access) without mention of the citizens who are otherwise touted to become ‘empowered’. The previous year, during her presentation at the 2016 Data Ethics conference, Sandra conceded that she was known among her former colleagues in marketing as “Miss act on fact”. While she was unsure whether the nickname was given in good spirit, she explained that took pride in it since the name reaffirmed a core value in her worldview summarised as “maximising impact through intelligent insights”. This worldview is also reflected in the envisioned use scenarios for imagined users. What is now called the Personal DfG was by her own admission “actually just a practical tool for us [DfG] to create full privacy and data control, because what we really want is to exploit and utilise the potential that all this data has”.

In terms of data accumulation, the goal was “to create a 360 degree overview of the individual - of their needs, motives, barriers and so on”. With this 360 degree overview, individuals were envisioned to be “empowered” by a “personal GPS” that could deliver data intelligence to help “create the best possible life”. A similar logic was additionally replicated on a societal level to maximise the impact of public health interventions through prevention. In Sandra’s statements, ethical values of user protection and user agency are instrumentalised to exploit the potential of data reuse. The instrumentalisation of ethics is however in this example justified according to a perceived personal and public good, which in turn enkindle additional questions of ethics. As scholars on the ‘quality of life’ have long established, what constitutes “the best possible life” is drenched with ethics and politics (Wahlberg and Rose, 2015).

Previously, I introduced how Digi.Me had embedded an arm’s length principle (“we are never able to touch, hold or see user data”) in their infrastructural design. According to my interview with Sandra, third-party actors such as the pharmaceutical industry are looking for a somewhat similar solution, which is what DfG now aims to accommodate with the “secret sharing” technology provided by the infrastructural capabilities of their partner Secata. In her elaboration, Sandra explains that “they essentially do not want to know who you are[...], they would actually prefer the arm’s length principle, but they want the knowledge about groups”.

In this example, the arm's length principle can be translated to anonymised data access or insights. This argument was also presented during a health data session at the 2018 MyData conference, which was challenged by an audience member. In his comment, the audience member argued that pharma's desire for an arm's length principle was more to do with the *liability* of possessing personal data rather than the *possession* itself. Seen this way, the studied infrastructures are not only reconfiguring the responsibilities of their users, but also customers who seek third-party access. Leaning on the observations of others, I argue that it would be simplistic to assume that the legitimisation of such access is solely achieved at the individual level of consent mechanisms. Based on a study in the same region, Tupasela, Snell and Tarkkala for instance propose that the private-public collaboration arising from the platformisation of 'national data ecosystem' can itself be understood 'as a type of legitimation exercise for both public and private actors' (2020: 11). Revisited once more, infrastructures can in this light be viewed to reconfigure the responsibility of the state too that takes on the role of an *enabler* to partake in the data economy through a delegated mode of governance (Tupasela, Snell and Tarkkala, 2020: 11; see also Faulkner-Gurstein and Wyatt, 2021).

The Community: Building the case for the monetisation and donation of personal data

In this final section of analysis, I explore how the two startups advance ethical arguments to support what they strive to make possible through their infrastructures: the ability for individuals to "donate" and "monetise" personal data. Based on my fieldwork, I highlight specific events where DfG and Digi.Me respectively present their case for why data donation and data monetisation are needed to peers and key stakeholders. Whether the startups can sell their ethical arguments has implications for the social, political and economic conditions of possibility for the development and success of their infrastructural enterprises. The analysis is in this respect not only an exploration of the ethical arguments, but the infrastructuring as a social practice. Common for how the case for data donation and data monetisation is presented is that they both tend to be advanced in extension of calls for data ownership. Since the invocation of data ownership has a tendency to remain 'ambiguous and even paradoxical' in such calls (Hummel, Braun and Dabrock, 2020: xx), I will strive to specify their deployments in the two empirical examples. Before engaging this endeavour, I take a brief moment to outline how data donation and data relationship has been conceptualised, along with their relation to ownership.

While some might be tempted to conflate data donation with data sharing, scholars contend that they can be seen as distinct kinds of transactions (e.g. Hummel et al., 2019: 25). As Barbara Prainsack notes (2019a: 12), legal definitions of donation all entail the following elements: 'The owner of a thing transfers it to another person or entity *without consideration of what she will receive in return'*. What Prainsack highlights from the emphasised part of the quote is the suggestion that data donation cannot be motivated by economic profit or understood to posit a *direct* reciprocity. In this respect, data *monetisation* can provisionally be viewed as an antonym to data donation, as the former posits that data are transferred or made accessible to another person or entity in *exchange* for monetary gain. Resuming the focus on donation, the absence of economic profit and direct reciprocity may seem trivial in the example clothing items. The importance of these characteristics however comes to fore if one considers the integrity of monetary donations for political campaigns or the ethics of organ donation in a global economy.

When it comes to the matter of ownership it is not inconsequential whether the item of donation is a material thing or personal data. This point is relevant for the donation and monetisation of data. If somebody buys a piece of clothing then that person can rightfully claim to have exclusive ownership rights over the material item as a personal property. Once an item obtains this status, then the owner is generally entitled to transfer the ownership rights onwards to another person or entity through an economic or altruistic transaction. Extending this logic to personal is however not straightforward. Whether a person can claim exclusive ownership over personal data is for instance subject of regional dispute (Purtova, 2015). The US discourse tends to promote the propertisation of personal data and in extension the transferability of its exclusive ownership. The European discourse, by contrast, tends to draw on its regional tradition of human rights to promote the anti-propertisation position that casts personal data as something inalienable. Another point that is relevant for the donation and monetisation of data lies in what Prainsack terms the 'multiplicity of digital data' (Prainsack, 2019a: 18). In short, this notion refers to digital affordances that enable data to reside in multiple places at once and be used by several actors simultaneously. This complicates the conventional understanding of donation, as the act of donating usually implies an exclusive transfer where the donor cannot simultaneously retain what is donated (e.g. a book). With this introduction, it is now time to resume the empirical analysis.

'The case for owning - and monetizing - your own data' (2500)

One place where calls for data ownership have been observably rife but contentious is in the MyData Community. This has for instance been observed in the extensive fieldwork of

Lehtiniemi and Ruckenstein (2019) that includes participatory observations at three annual conferences between 2016 and 2018. In lieu of the data ownership, the authors noticed how activists in the MyData community 'consciously employ the concepts of data management and control, focusing on individuals' practical capacity to make use of their data' (Lehtiniemi and Ruckenstein, 2019: 5). Despite this resistance, the language of data ownership continues to have a strong presence in the community, which I have observed at each annual conference between 2017 and 2019. Sometimes, the notion that individuals should be able to own personal data about them was articulated as an *ethical* right while at other times it was described as a *legal* right stipulated in the GDPR. The latter is based on a particular interpretation of the GDPR that most scholars consider to be unsubstantiated insofar that it is understood as exclusive property rights (REF).[15]

In the annual MyData conference of 2019, I attended a debate entitled *Data ownership - thanks, but no thanks.* The debate description in the conference programme elaborated the normative framing of the event with the following objective: 'that this round table will put the ownership debate to rest at least in the MyData Community so that we can focus on more constructive debates that emphasize human, civil, and political rights of individuals in respect of their personal data' (https://www.mydata2019.org/programme-page/data-ownership/). The debate was attended by three discussants who appear with the following pseudonyms: Lukas, Michelle and Christopher. Lukas was a Finish lawyer from Technology Industries of Finland who did policy work in the space of digital technology and data. Michelle was in the programme introduced as a 'legal advisor and digital rights advocate' who presented work she developed during her tenure at Privacy International. The final discussant was Christopher, a co-founder of acquired company Personal who introduced himself as 'the US CEO of Digi.Me'. Each discussant gave a ten minute presentation, which then informed a moderated debate.

Lukas was the first discussant to take the stage where he used his time to discuss the fit of data ownership with Europe and legal theory. When it comes to matters of privacy and data protection Lukas established that he believed it was more appropriate to talk about 'rights' than 'ownership'. As most highlight in the scholarly discourse (Purtova, 2015; Stepanov, 2019), Lukas noted that his personal preference was also a European one where there has traditionally been strong belief in fundamental rights. This tradition should be seen, he argued, in context of the atrocities that took place during the second world war, which has in

---

[15] As suggested, Digi.Me and DfG are among the startups to deploy (and moreover champion) the language of data ownership in their promotion of individual control. This is for instance observable in the subtitle of Digi.Me's 2017 conference presentation Patient centricity: Empowering the individual to own and control their data, asserting a need to 'give data back to the individuals' (Fieldnotes).

part resulted in article 7 (*Respect for private and family life*) and especially 8 (*Protection of personal data*) in the European Charter of Fundamental Rights. For his second point, the lawyer articulated his preference for ‘control’ over ‘ownership’ since the latter notion is too ‘static’ and often implies ‘exclusive property rights’. Control was by contrast argued to be more ‘dynamic’ and compatible for thinking about rights for reasons relating to the aforementioned multiplicity of digital data. This observation has been identified as a common foothold in the argumentation of anti-propertisation scholars by Purtova (2015), which she contends for reasons I will address in due course. In terms of Lukas’s presentation, his main points were that although ownership could possibly make sense for certain types of data, this was not the case for personal data due a perceived incompatibility with European tradition and legal theory.

The subsequent discussant to take the stage was, who contrary to his fellow discussants took it upon himself to challenge the normative framing of the debate. This was evident from the off-set of his talk as the title of his presentation projected onto the screen: *The case for owning - and monetizing - your own data.* Through his talk, Christopher sought to affirm his opening question of whether it is ‘possible - or ethical - to own or monetise our own data?’. By the end, Christopher concluded that ‘Data ownership and monetization are essential keys to the future we [the MyData community] all want when used in the right context'. In order to clarify the ‘right context’, Christopher dismissed that an ownership model could work across the entire personal data ecosystem. The general idea that data should be ‘sold to the highest bidder’ was furthermore characterised as ‘crazy’, as Christopher insisted that selling and monetising are two different things. For data intermediaries like Digi.Me, however, data ownership was ‘not only a good idea; it was necessary’.

For Christopher, data ownership was necessary for two reasons. The first reason was described as an ‘ethical right’ for individuals to exercise privacy self-management in regard to personal data. With platforms such as Digi.Me, Christopher claimed that individuals could potentially combine ‘millions of data data points’ about themselves, which governments and companies could not match for legal and technical reasons. This personal ‘competitive advantage’ would necessarily need ‘legal protection’, which Christopher seems to propose is better accomplished through ownership rather than rights. The example that Christopher provides is a ‘data ownership agreement’ invented by *Personal* to assert individual ownership over what was stored in the personal data store and shared with third-parties. With a legal basis in contract law, Christopher’s proposal leans toward the idea of propertisation as a means to privacy that is more prevalent in US discourse (Putrova, 2009; Evans, 2011). On this latter point, Lawrence Lessig (2002: 255) posited the claim two decades ago that ‘If you could get people (in America, at this point in history) to see a certain resource as property, then you are 90 percent to your protective goal’.

In addition to privacy self-management, data ownership was secondly argued to give individuals an economic right to benefit from the personal data economy. This was founded on a strong belief that ‘individuals should fundamentally be able to participate in the economics of data without being exploited or subjected to risk or harm’. Introducing the Digi.Me powered app *Universal Basic Data Income* (UBDI), Christopher demonstrated that this was not an abstract idea. With UBDI, consumers could monetise anonymised insights for market research based on the self-disclosure of personal data. Christopher was a co-founder of this US based service where the conference provided the venue for its official launch in Europe where it is promoted as a ‘data ethical monetization programme for consumers' (Firth, 2019). For instance, Christopher explained, consumers ‘aged 20-30’ who ‘run more than ten miles per week’ could be compensated $25-50, which would only take a couple of minutes to analyse through edge processing (i.e. on the consumer’s device). ‘If you happen to also be an IT decision maker who spends more than $1000 a year on cloud services’, he continued, ‘then a single study could be worth $2-3-400’. In this way, UBDI was believed to disrupt the multi-billion dollar industry of ad tech by disproving what Christopher perceived as ‘one of the biggest lies coming from Silicon Valley’: that data is only worth something at scale, not at a personal level. In this instance, the necessity of ownership is articulated as a means for the marketability of personal data where the imagined user is moulded according to profit maximising and utilitarian traits of homo economicus:

> When [people] see $1000 a year within an hour or two, they drop what they are doing and are willing to plug in their accounts and figure out what their data can do for them. It’s just how we have been conditioned, it’s how we understand value - or one of the main ways we do. And so, I am here to tell you that is one path I think can and should also work.

As others have identified, this latter version of data ownership is also more prominent in the US discourse where the propertisation of personal data predominantly caters to utilitarian values (Prainsack, 2019a: 18).

Michelle’s stance in the debate was also clear from the beginning as she began her presentation entitled *Data ~~Ownership~~ Rights.* In line with Lukas, Michelle argued that a rights-based approach to control was the best option currently available to protect the privacy of citizens and that property rights were incompatible with the nature of data. In order to build a ‘healthy ecosystem’, Michelle added, control rights would need to be supplemented with governance principles regarding transparency, privacy-by-design, privacy-by-default and perhaps also non-market mechanisms. Commenting on the marketability of personal data, as exemplified with UBDI, Michelle was sympathetic towards those who felt a sense of entitlement to have a ‘cut’ of the economic gain made with their personal data. Her analysis was however less sympathetic toward ownership as a solution to the unethical practices in tech, unless it is merely a rhetorical framing of agency or autonomy. Unlike Christopher, propertisation of personal data could in the view of Michelle become a ‘dead end’ that could easily result in *less* control for individuals who would be left with limited rights if they for instance choose to sell their immaterial property. This idea of ownership was also dismissed on a structural level, as Michelle argued that it failed to address the fundamental dynamics that allowed data monopolisation, abusive practices or incentives for corporate targeting and profiling. On the contrary, it was argued to introduce an economic incentive for self-datafication and disclosure hereof, thus supporting the longevity of the ethical problems rife in the tech industry. Instead, MyData was implored to explore alternative models in their quest for a healthy ecosystem, such as national data funds.

Throughout the debate, Christopher seemed particularly invested in communicating his position when others had the word in cases where he subscribed to a similar view or value. I understood his repeating verbal and gesticulatory affirmation (e.g. saying “definitely” or nodding) as a conscious attempt to manage Digi.Me’s reputation and stakes in MyData, as a member and sponsor, wary that ownership and monetisation were polemic notions in the MyData community. As the debate approached an end, the host invited a closing question from the audience. Christopher became active in his body language once more, but this time he seemed to communicate with a particular person in the audience instead of the general assembly. The person had taken a seat a couple of chairs from mine only moments earlier, who I recognised as the other UBDI co-founder from my informal interview with her at the company’s promotional booth at the conference. Before the host finished his invitation to the audience, Christopher oriented the host’s attention to his fellow co-founder whose affiliation was never addressed. The co-founder apologised for a possible repetition - because she had missed most of the debate - before posing her question to Christopher’s fellow discussants. Paraphrased, she asked whether their ‘stance’ against ownership rights and monetisation might be hurting what they care about the most given that the latter could also incentivise privacy?

Provided that her reasoning is consistent with our foregoing interview, the logic behind this question can be understood accordingly: By putting a price on personal data, monetisation is seen to make its value visible to consumers who should therefore be incentivised to protect their immaterial property. A similar reasoning is provided by Lessig, who also articulates ownership as a way to mitigate the 'privacy paradox' (Kokolakis, 2017) in contemporary digital life: 'If people see a resource as property, it will take a great deal of converting to convince them that companies like Amazon should be free to take it. Likewise, it will be hard for companies like Amazon to escape the label of thief' (Lessig 2002b: 255). With a few adjustments, Christopher directed the question specifically to Michelle, asking whether she considered the outcome of ownership and monetisation so 'worrisome' that it should be entirely dismissed - or if they could co-exist with alternative approaches. In response, Michelle reiterated her stance by dismissing both the notion of ownership and monetisation - cognisant of the fact that the latter need not rely on the former - but would welcome counter arguments. The lawyer reiterated his previous arguments, adding that contractual agreements to control and monetise personal data was a challenge in terms of scalability since they would need to be re-written along with other demanding processes. Christopher acknowledged that ownership, control and agency might be different from a legal perspective. From a 'non-lawyer' perspective, however, the difference was a matter of semantics in the belief that they were 'effectively[...] trying to say the same thing'.

Data donation

Part of what drew my attention to DfG during fieldwork was the explicit use of the rhetoric of 'good', which for me beckoned further empirical research to pin down its practical translation. During the *Data for Good Summit* in 2020 co-hosted by DfG, I overheard a conversation between two representatives who occupied neighbouring booths to promote their tech companies. Whereas several conference attendees crowded the DfG booth, the ones occupied by two representatives were devoid of activity. Looking down the line of booths with envy, one of the representatives told the other that they were victims of 'unfair competition' since nobody could be opposed with the bannered slogan of 'data for good'. Although it was delivered as a witty remark, the comment highlights a relevant observation regarding the rhetorical enrolment of ethics in the promotion of tech. A closer look at the data sharing discourse reveals that the rhetoric 'good' is not unique. In fact, the precise slogan of *Data for Good* was the name of a public campaign in 2014 launched by another digital health platform called PatientsLikeMe. As STS scholar Btihaj Ajana details, the campaign promoted the importance of 'donating health data' and ameliorating a cultural 'distrust' that had previously tainted public attitudes towards the pharmaceutical industry (Ajana, 2017: 9).

Coincidentally, DfG also appends the language of donation to personal data in relation to the pharmaceutical companies that as an industry also invokes scepticism in Europe (Skovgaard, Wadmann and Hoeyer, 2019). As I have outlined in the introduction of this section, the idea of ownership is closely tied to the concept of donation, but also how it is not unproblematic when applied to data in certain contexts. One of the contexts where this was observed in my fieldwork was at the first event I attended that was hosted by DfG.

“Who owns the citizens’ data?”

In November 2017, DfG hosted an event on the premises of the Danish Parliament entitled *Who owns the citizens' data?* As noted in the invitation, the event was the public disclosure of the DfG initiatives that was described as a ‘Danish proposal for a forward-looking big data community for data-driven prevention and health’. As suggested by the event title, the question of ‘who owns the citizens’ data?’ was a central element in the pitch. The proposed answer from DfG is indicated by the apostrophe placement in the title, as the ownership of citizen data was assigned to the citizen in terms that were suggestive of an exclusive entitlement. Legal support for this suggestion was found in the imminent introduction of the GDPR. Although the formulation is now retracted, a similar legal interpretation was until recently also available on DfG’s website: *‘GDPR*[...] supports that we as citizens in the future have the *rights to own our personal data’*.[16] While DfG’s contribution to the ownership question may have been seen as a stronghold going into the event, it did not seem to elicit the expected reception among the politicians and other key stakeholders in the audience. Even when the event approached its conclusion, a key political stakeholder explicitly conceded how it remained a challenge to comprehend what DfG could bring to the table. At the reception, several of those affiliated with the startups were visibly disappointed, which was also vocalised in closed circles. While this disappointment cannot be reduced to the reception of DfG’s call for data ownership, it was a key framing device for the event itself. The struggle to convey this call with an adequate effect of persuasion is however less surprising if one considers the implications of exclusive ownership in the context of Danish healthcare.

In one of his best selling books, the American cardiologist and author Eric Topol (2015: 281) for instance writes: ‘The ownership of property is central to emancipation. It’s unquestionably appropriate, a self-evident truth, that each individual is entitled to own all of his or her medical data’. While this statement might at first glance make sense, it can rapidly become complicated insofar as ownership is understood as an exclusive entitlement. An initial

---

[16] https://dataforgoodfoundation.com/en/what-why, author’s emphasis (accessed 23/09-2020)

complication can be invoked by considering the rights of those who actually enter data in the medical record, such as general practitioners.[17] A second complication regards the fact that personal data can concern more than one data subject. The classic example is genome data, since it can potentially disclose a risk of hereditary illness among biological family members. The social dimension of personal data is however visible in numerous other examples, such as the day-to-day coding practice of general practitioners that can for instance implicate the illness of a spouse. A third complication to consider concerns the hierarchies of law that in various examples elevates the collective interest over the individual interest (e.g. Vayena and Blassime, 2017). This practice is particularly extensive in Denmark and its neighbouring welfare states where the individual right to erasure or opt-out is superseded by a legal obligation of retention in the interest of public health (e.g. Hoeyer, 2016, Nordfalk and Hoeyer, 2020), but also other interests such as accountability (e.g. Winthereik, Van Der Ploeg and Berg, 2007). This latter complication illustrates why it is misleading to conflate the right to data portability with property rights, since the former lacks the defining element of the latter even when combined with the right to erasure: the right to exclude (Greaf, Husovec and Purtova, 2018: 1368).[18]

Data donation as a personal, corporate and public good

While the public reveal of DfG struggled to meet expectations, the startup would soon encounter better fortune. As mentioned previously, a series of successful funding applications would in 2019 and 2020 secure the financial means to develop and pilot the DfG platform envisioned to enable data donations for a good cause. As illustrated in the pamphlet for the HeDaX project (see figure 6), patients are invited to participate with the following question: “Do you want to participate in the securement of a better and more precise cancer treatment?”. Participation in this instance includes the donation of personal information on exercise and dietary habits from fitbit, retail records and bank record as a supplement to registry data. Contrary to the conventional understanding of donation, the data would in this instance remain in the personal data store while it is made accessible to third-parties anonymously. Returning to the pamphlet quote, the health benefit for the citizens engenders

---

[17] Although the Danish healthcare system is predominantly financed with public funding, general practitioners are self-employed enterprises contracted by the state.

[18] The fact individual empowerment will often be limited to aggregate a copy of their data to a personal data store is sometimes recognised in Digi.Me’s online promotion. Yet, Digi.Me maintains that individuals can still have exclusive ownership over a digital photo even if it is copied in the ownership space of social media because the individual will have the ‘original’ (Firth, 2016a). Critics could however argue that the notion of an ‘original’ has a more limited value in the digital space where computers can process each zero and one as ‘a mimetic machine par excellence’ (Boon, 2010: 167). This has been experienced first hand by the entertainment industry in regard to piracy communities, which moreover demonstrates the limitations of exclusive ownership as a means to control or protect immaterial property through copyright.

a question about whether the reciprocity in the data donation should be characterised as direct or indirect. A more elaborate description of the envisioned benefits of HeDaX can be obtained from the project homepage. By participating in the project, citizens are informed that they "contribute to new research and new treatments that can secure them better and more effective treatment with fewer side effects ('Gevinsten ved projektet', n.d.). This will additionally provide citizens the opportunity to gain insights of individual relevance from research more quickly". In this example, donation is articulated with a personal good that most patients with a life-threatening illness will probably find more compelling than a monetary transaction. The promise of a better treatment is notably worded with conviction in the first sentence ("can secure"), which is less present in the concluding sentence as access to individual health insights is noted as an "opportunity".

Figure 6. Patient pamphlet for HedaX the project

Shifting attention towards the other beneficiaries highlights how data donation is also articulated to support a public good. Among the beneficiaries are public and private researchers who are articulated to benefit from “easier access to more data with potential significance for health”. Whereas the contribution of citizens was previously described as something that “can secure” better treatment, the use of “potential” in this latter example recognises the experimental nature of the project. Other benefits include a cost-effective production of research and an efficient dissemination of results that can also include the data donor when relevant. Another beneficiary is the health system where the gains are articulated to derive from the “possibility to develop and optimise new treatments”. The envisioned benefits are improved effectiveness of care trajectories and the advancement of prevention and precision medicine. The “society” is also noted as an explicit beneficiary in itself. The health benefits are articulated as improved conditions for self-care and prevention, as well as a more effective healthcare system. The societal benefit is however also articulated in economic terms: “New insights on the basis of Danish health data can attract companies and clinical trials to Denmark. This will strengthen the Danish ecosystem within life science even more and create more jobs and economic growth”. In short, data donations are considered a win for all parties involved.

Whereas the reuse of health data for public health is widely accepted as a common good, the same level of acceptance cannot necessarily be assumed when the secondary purpose is economic gain through commercial collaboration. This is for instance indicated in a review of public attitude towards health data reuse where commercial reuse was registered as a common point of scepticism in studies of people living in the EU (Skovgaard, Wadmann and Hoeyer, 2019). This scepticism was however in some cases contingent on the industry in question and other factors. The attitude towards health data access for pharmaceutical companies was for instance contingent on whom the respondent believed would benefit. Despite these concerns, it remains a fact that the (re)use of health data for jobs and economic growth has been a national strategy in Denmark for several years (e.g. Tupsula, 2021). To this end, the framework of registry stipulates that health data access is conditioned upon a collaboration with Danish research environments. The national strategy was explicitly formulated in the *Growth plan for health and welfare solutions* from 2013 in which the policy paper stated that “The Danish health data constitute an important competitive parameter for the attraction and quality of research, development and documentation of new products” (Regeringen, 2013). The economic exploitation of health data as a policy objective was however the following year articulated as public concern following its enrollment in a national health data controversy (Wadmann and Hoeyer, 2018).

Fast-forward six years, the public scepticism towards the commercial and economic exploitation of health data reappeared during a panel debate at Data for Good Summit in

2020. The conversation that caught my attention was between two discussants, who I will call Søren and Mathias. Søren was the founder and CEO of a Danish entrepreneurial health tech hub while Mathias was a spokesperson for national authority on statistics, Denmark Statistics. The public scepticism was initially articulated by Søren who characterised it as an impediment holding back the economic opportunities of an “asset” worth “billions and billions of money”. Asserting a call for action, Søren issued a caution that the worth of Danish health data will only diminish with time as competitors will not be as hesitant. This problem was illustrated with the current framework in which the use of health data in the university sector cannot legitimately be justified on the basis of its monetary return alone. “it has to be because it helps sick people”, he concluded, if scepticism is to be avoided. With a swift response, Mathias affirmed that the observations of his fellow discussant were “spot on”. Mathias proceeded to explain how it was only recently that Denmark Statistic had an internal discussion of whether there should be a differentiation between public and private actors. Both actors were however recognised to be of societal importance, but Søren underscored that it was also paramount to preserve the public trust to state institutions as data holders. One way forward, Søren concluded, was therefore to “work with the narrative that private value is public value”.

Data donation as a rhetorical device? (Redefining public value?)

If we take a moment to reflect upon the presented empirical material, then it becomes visible how the DfG are actively enrolled in the exercise of narrating private value as a legitimate public value. With events such as the Data for Good summit, DfG takes proactive steps to provide a stage for the narrative to be vocalised. As mentioned in the second section of analysis, a key function of the foundation structure is in Sandra’s own words to secure “legitimate access to data, insight or knowledge - for commercial actors, public players and everyone”. On the HeDaX project homepage, the benefit of “easier access to more data” is explicitly stated to include private research, as is the collaboration with private companies in the pharmaceutical and tech industry. As such, it is fairly transparent that the project aims to deliver a win in the form of personal, corporate and public value where the former two are argued to support the latter.[19]

Recalling the witty but incisive remark from the booth, what is framed as a ‘good’ is not inconsequential. Based on research in self-tracking communities, scholars for instance observed how data sharing is framed as a ‘public good’ to the extent that ‘solidarity becomes almost synonymous with data sharing and information giving’ (Ajana, 2017: 9). This framing

---

[19] While the name of private actors features among the collaborators, it is not explicitly stated that the data are used for private research or economic growth in the pamphlet.

has also been encountered in my fieldwork by practitioners in fields such as artificial intelligence. Given the technoscientific opportunities, it was argued that the burden of proof in terms of ethical justification should not befall the reuse of health data, but the *absence* hereof. This inverted argumentation is forcefully narrated in the title of the Danish public-private partnership *Data saves lives*, which was established in the backdrop of a series of public health data scandals to advance the agenda of "better use of Danish health data" (https://www.cphhealthtech.dk/data-redder-liv). A similar inversion of ethical justification is also argued by scholars in support of the donation of medical records for public good, but with a proviso on the need to acknowledge and respect the conflicting beliefs and interests involved (Jones, 2019). The term "donation" and umbrella of "data for good" can in a similar fashion be of rhetorical utility when deployed to describe private data access and gain if approached in the right manner.

Since the ethnographic material was gathered when the HeDaX project was in its infancy, it is limited what can be said about the practical implication of data donation. Critical scholars do however caution that although calls for patient control and self-access over health records might appear in the rhetorical guise of empowerment the result can be disempowerment. This caution has been voiced in relation to the regulation of health data access in US healthcare (e.g. Ebeling, 2019), but also regarding platform initiatives that operate with data donations. An infamous example of the latter is 23andMe (Van Dijck and Poell, 2016; see also Sharon, 2016), which originated in 2006 as a direct-to-consumer service for genetic testing. In exchange for submitting a saliva sample through a special kit and filling out a "fun" online questionnaire, the consumer would gain a personalised overview of their genetic profile with a risk assessment for certain disease (Van Dijck and Poell, 2016: 3).[20] Consumers could also donate the submitted data for the alleged purpose of advancing genetic research in the interest of the public. In 2012, 23andMe was involved in two events that exposed consumer control as a vulnerable notion. The company was granted a patent related to Parkison's disease, prompting conflicting interests among consumers who donated their data to support public interest and not private gain. The acquisition of CureTogether in the same year revealed another vulnerability of consumer control over personal data. CureTogether was a patient experience platform akin PatientsLikeMe where the acquisition resulted in an corporate absorbent of data on 600 medical conditions. In 2015, 23andMe launched a new business model focused on drug development, illustrating

---

[20] As the authors note, this service was in 2013 replaced with ancestry identification following a ban from the American Food and Drugs Administration due to the algorithmic outputs of misleading information.

again the immediacy by which donated or acquired data can be repurposed in ways that depart from personal preferences.

Interim discussion

Although I have used Digi.Me and DfG to respectively analyse data monetisation and data donation, the distinctive use of each transaction is in reality not that clear clut. In response to the Covid-19 pandemic, Digi.Me launched a “Data Donor programme” to accelerate academic and private research access to “real world evidence” from individuals for the purposes of treatment development. Based on Digi.Me’s “privacy-centric technology”, users were afforded the opportunity to donate medical records, self-assessment reports and other data. In my interview with Sandra from 2017, the prospect of facilitating users a monetary compensation for data access was described as a possible use case scenario for DfG in relation to the pharmaceutical industry. While both claim to be neutral intermediaries, looking at what they aim to make possible - which they to an extent already facilitate - through their infrastructures reveals the political nature of their doings. This is made explicit by paying attention to the ethical arguments advanced or supported in relation to the monetisation and donation of data alike.

For Digi.Me and UBDI, the ethical justification of data monetisation was in part articulated in relation to an economic right for individuals to participate in the data economy with personal data. A similar idea can be traced in the concept of a ‘universal micropayment system’ proposed by Jaron Lanier, who aside from being a computer scientist is often characterised as a ‘visionary’ figure from Silicon Valley (e.g. Kahn, 2011) and credited for pioneering virtual reality:[21]

> A new kind of middle class, and a more genuine, growing information economy, could come about if we could break out of the ‘free information’ idea and into a universal micropayment system. We might even be able to strengthen individual liberty and self-determination even when the machines get very good. (Lanier 2014: 9)

As Metcalf, Moss and Boyd (2019) observe, Silicon Valley has a tendency to revert to technological solutionism when confronted with ethical problems. While the foresight of Lanier highlights the universal micropayment system as a catalyst for a ‘new kind of middle class’, critical scholars have cautioned that this liberatarian vision ‘may inadvertently lead to

---

[21] Lanier was also referenced explicitly in Christopher’s presentation to assert that the annual economic gain for consumers was most likely closer to the visionary’s estimation of $20,000.

the “proletarization” of users and the transformation of privacy into a luxury of those who can afford not to sell their data (Casilli, 2019)’ (Sharon and Lucivero, 2019; see also Lehtiniemi and Ruckenstein, 2019). Responding to power imbalance and inequality *caused* by the tech industry, Lanier and UBDI can in this light be viewed to propose a solution without addressing the underlying logics that gave rise to the problem it seeks to address.

In the examples given, both startups struggled to advance their calls for data ownership. While the relation between ownership and the two transactions remains muddle on a conceptual level, this has neither prevented practical development of data monetisation and data donation nor their explicit narration as ethical transaction. In a Danish context at least, it seems that the latter type transaction is an easier sell. This is in part suggested by the discontinuation of the data intermediary startup company Cima Technologies that entered the Danish market around the same time as DfG. Whereas the promotion of DfG highlighted the ‘common good’ of individual control, Cima Technologies (2017) took another approach by promoting the prospect where citizens could gain an annual revenue €1000 by “selling” personal data to third-parties.

Conclusion

In the present study, I have explored the following question: What role does infrastructure play as Digi.Me and DfG promote themselves as ‘the ethical choice’? Drawing on an infrastructural approach, I have answered this question by attending to three different audiences: the user, the market and the communities Digi.Me and DfG engage in.

Beginning with the user, I highlighted technical architecture (privacy-by-design principles), governance model (quasi-market regulation), and consent solutions (e.g. digestible terms for informed and explicit consent) as infrastructural components used in the promotion of user protection and/or user agency as ethical values. Each component was argued to cast the imagined user as a data subject empowered to regain control over personal data. A critical inspection of the promoted ideas of “private sharing” foreground the practical limitations of control and its political ramifications in terms of responsibilisation.
Advancing to the market, the second part of the analysis situated the trend of ethics in Denmark by highlighting the promotion and embrace of a particular narrative in the national tech scene: a market framing of ethics as a competitive advantage. Locating a niche market between state and established data market actors, the startups draw on their infrastructures to offer ethics ‘as a services’ for third-party actors to gain legitimate access to self-curated

life records without the burden of GDPR compliance. Just as the market makes demands on companies, I suggest that companies also use ethics to sell themselves back to the market forming new collaborations, roles and responsibilities between state, market and citizen. Shifting attention to the community, the final part of the analysis explored how the two startups advanced ethical arguments to rally support and legitimacy for what they strive to achieve through their infrastructures: Data monetisation and data donation. Infrastructuring new data flows is shown to draw on narrational practices. Key to the promotion of these transactions was the promotion of exclusive data ownership that is shown to cause confusion and tension in a Nordic context for cultural and practical reasons. Whereas the ethical argument for data monetisation is still looking for suitors, data monetisation has gained traction in Denmark with the support of state funding. I suggest that the latter might be considered as a means to rewrite established narratives of public value and in this way legitimise new formations hereof. Both transactions however offer potential incentives for third-party access where data monetisation is arguably more ideologically vulnerable in comparison to data donation.

The main argument I put forward is that more ethnographic research is needed to illuminate the infrastructural practices that underpin the ethical claims put forward by tech actors in the data economy. As summarised above, the article provides a descriptive demonstration of how infrastructures are also enrolled in the promotion of ethical tech that differentiates between user, market and state in the Nordic Region. These descriptions shed light on how ethics is operationalized not simply in the tech industry in a generic sense, but in the specific context of highly regulated countries in Europe through data intermediaries. The article also provides insights specific to data intermediaries examined through two competing models. In the wake of an ethical crisis, DfG and Digi.Me represent a burgeoning solution to restore trust for a continued intensification of data reuse. The new possibilities for data-intensification are not merely built through technical solutions, but also social arrangements that take on the form of governance models, collaboration as well as narration. These possibilities promoted in the name of ethics however introduce new questions that are similarly to those they claim to address also of ethical, societal and political importance. A central theme in these questions pertains to roles and responsibility that call into negotiation the relation between concern state, market and citizen that will require continued scrutiny from scholarly and regulatory actors.

References

Ammanath B (2021) Thinking Through the Ethics of New Tech…Before There's a Problem. *Harvard Business Review.*

Anand N, Gupta A and Appel H (2018) *The promise of infrastructure.* Duke University Press.

Bowker GC (1995) Second Nature Once Removed: Time, Space and Representations. *Time and Society*, 4(1): 47–66.

Clarke AE, Mamo L, Fosket JR, et al. (2010) *Biomedicalization*. Duke University Press.

'Declaration' (n.d.) MyData. Available at: https://mydata.org/declaration/

'How do we secure trust in handling personal data?' (n.d.) DfG. Available at: https://dataforgoodfoundation.com/en/what-why (Accessed 23 Sep 2020)

Douglas-Jones R, Burnett JM, Cohn, M et al. (2018) A Bestiary of Digital Monsters. In *Living with Monsters?: Social Implications of Algorithmic Phenomena, Hybrid Agency, and the Performativity of Technology.* Springer, IFIP AICT - Advances in Information and Communication technology, vol. 543, pp. 177-190, Living with Monsters, San Francisco, United States.

Edwards PN (2003) Infrastructure and modernity: Force, time, and social organization in the history of sociotechnical systems. *Modernity and technology,* 1: 185-226.

European Commission (2016) General Data Protection Regulation.

European Commission (2020a) A European strategy for data.

European Commission (2020b) Proposal for a Regulation of the European Parliament and of the Council on European data governance (Data Governance Act)

European Commission (2021) Citizen-centred EU-EHR exchange for personalised health. Available at: https://cordis.europa.eu/project/id/826117 (Accessed Mar 2021).

'Event: Data intermediaries launch' (2022). *MyData Global Blog.* Available at: https://mydata.org/2021/12/20/event-data-intermediaries-launchpad/ (Accessed 31 January 2020).

Firth E (2017) Digi.Me allowing Icelandic citizens to download their own health data in world first. Digi.Me Blog. Available at: https://blog.digi.me/2017/05/31/digi-me-allowing-icelandic-citizens-to-download-their-own-health-data-in-world-first/ (Accessed 8 November, 2017).

Hartman T, Kennedy H, Steedman R et al. (2020) Public perceptions of good data management: Findings from a UK-based survey. *Big Data & Society, 7*(1): 1-6.

Harvey P and Knox H (2015) *Roads: An Anthropology of Infrastructure and Expertise.* Cornell University Press.

Hogle L (2016) The ethics and politics of infrastructures: Creating the conditions of possibility for big data in medicine. In: *The Ethics of Biomedical Big Data*, Springer: Cham, pp. 397-427.

Hughes T (1983) *Networks of Power: Electrification in Western Society, 1880-1930.* The John Hopkins University Press.

Hummel P, Braun M, Tretter M et al. (2021). Data sovereignty: A review. *Big Data & Society*, 8(1): 1-17.

Jasanoff S (2015) Imagined and Invented Worlds. In: Jasanoff S and Sang-Hyun K (eds) *Dreamscapes of Modernity*. University of Chicago Press, pp. 321-341.

Kokolakis S (2017) Privacy attitudes and privacy behaviour: A review of current research on the privacy paradox phenomenon. *Computers & security* 64: 122-134.

Larkin B (2013) The politics and poetics of infrastructure. *Annual review of anthropology*, 42: 327-343.

Lehtiniemi T and Kortesniemi Y (2017) Can the obstacles to privacy self-management be overcome? Exploring the consent intermediary approach. *Big Data & Society,* 4(2): 1-11.

Lehtiniemi T and Ruckenstein M (2018). The social imaginaries of data activism. *Big Data & Society*, 6(1): 1-12.

Petersen A (2015) *Hope in health: The socio-politics of optimism.* Springer.

Phan T, Goldenfein J, Mann M et al. (2021) Economies of Virtue: The Circulation of 'Ethics' in Big Tech. *Science as Culture.*

Plough T and Holm S (2016) Meta Consent – A Flexible Solution to the Problem of Secondary Use of Health Data. *Bioethics*. 2016 Nov, 30(9): 721–732.

Poikola A, Kuikkaniemi K and Honko H (2015) MyData – A Nordic Model for human-centered personal data management and processing. Available at: https://julkaisut.valtioneuvosto.fi/bitstream/handle/10024/78439/MyData-nordic-model.pdf.

Prainsack B (2019a) Data donation: How to resist the iLeviathan. In: Jenny Krutzinna and Luciano Floridi (eds). *The Ethics of Medical Data Donation*. Dordrecht: Springer. pp. 9–22.

Prainsack B (2019b) Logged out: Ownership, exclusion and public value in the digital data and information commons. *Big Data & Society*: 1-15.

Metcalf, J., & Moss, E. (2019). Owning ethics: Corporate logics, silicon valley, and the institutionalization of ethics. Social Research: *An International Quarterly*, 86(2): 449-476.

Riso B, Tupasela A, Vears DF et al. (2017) Ethical sharing of health data in online platforms – which values should be considered? *Life Sciences, Society and Policy*, 13(12): 1-27.

Rose N (2009) The Politics of life itself: Biomedicine, power, and subjectivity in the twenty-first century. Princeton University Press.

Sanyal N (2021) How to Manipulate Customers … Ethically. *Harvard Business Review.*

Sharon, T. (2018). When digital health meets digital capitalism, how many common goods are at stake? *Big Data & Society*, 5(2): 1-11.

Sharon T and Lucivero F (2019) Introduction to the Special Theme: The expansion of the health data ecosystem – Rethinking data ethics and governance. *Big Data & Society*, 6(2): 1-5.

Smith IH and Kouchaki M (2021) Building an Ethical Company: Create an organization that helps employees behave more honorably. *Harvard Business Review.*

Solove DJ (2013) Privacy Self-Management and the Consent Dilemma. *Harvard Law Review*, 126: 1880-1903.

Star SL (1990) Power, technology and the phenomenology of conventions: on being allergic to onions. *The Sociological Review, (38)*1: 26-56.

Star SL (1999) The ethnography of infrastructure. *American Behavioral Scientist,* 43(3): 377–91.

Star SL (2002) Infrastructure and ethnographic practice: Working on the fringes. *Scandinavian Journal of Information System,* (14)2: 107-122.

Rasmussen JG (2010) Læger manipulerer med journaler for at beskytte patienter. *Dagens Medicin*. Available at: https://dagensmedicin.dk/lager-manipulerer-med-journaler-for-at-beskytte-patienter/ (accessed 7 May, 2021).

Thomsen S (2006) Industrial Foundations: Foundation Ownership of Business Companies. In Prewitt et al. (eds) *Legitimacy of Philanthropic Foundations: United States and European Perspectives.* Russell Sage Foundation: 236-251.

Van Dijck J, Poell T, and De Waal M (2018). *The platform society: Public values in a connective world.* Oxford University Press.

'Video' (n.d.) Digi.Me. Available at: https://digi.me/video/ (Accessed Mar 11 2020)

Wadmann and Hoeyer (2018) Dangers of the digital fit: Rethinking seamlessness and social sustainability in data-intensive healthcare. *Big Data & Society, 5*(1): 1–13.

Wahlberg A & Rose N (2015) The governmentalization of living: Calculating global health. *Economy and Society*, *44*(1): 60-90.

What is digi.me? (n.d.) *Digi.Me.* Available at: https://digi.me/what-is-digime/.

1,000s of sources of data. 1 SDK. (n.d.) *Digi.Me.* Available at: https://digi.me/sources/.

Winner L (1980) Do artifacts have politics? *Daedalus,* 109(1): 121-136.

Zuboff S (2019) *The age of surveillance capitalism: The fight for a human future at the new frontier of power.* New York: PublicAffairs.