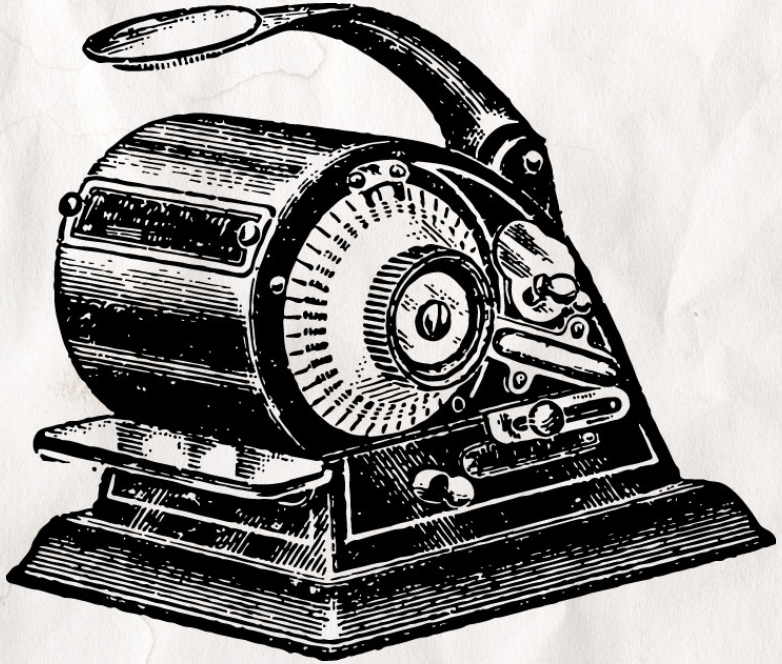
ENGAGING THE DATA MOMENT



Special issue
Volume 11• Number 1 • 2020

STS Encounters
Research papers from DASTS
SPECIAL ISSUE
Volume 11• Number 1• 2020

Data: a cosmopolitical approach

Peter Danholt
Dept. Digital Design & Information studies, Aarhus University, Denmark

Morten Bonde Klausen
VIVE, Denmark

Claus Bossen
Dept. Digital Design & Information studies, Aarhus University, Denmark

DASTS is the primary academic association for STS in Denmark. Its purpose is to develop the quality and breadth of STS research within Denmark, while generating and developing national and international collaboration.

Abstract

In this paper, we propose a cosmopolitical approach to, and understanding of, data, based on the work of Isabelle Stengers. This entails appreciating data as constituted through multiple actors and actions, and, accordingly, as something capable of producing unanticipated, surprising consequences. Cosmopolitics helps us think about data, and datafication, as actors in a more-than-human world in ways that transgress a common and widespread perception of data as either neutral, objective and representational or as socially constructed, perspectivist and endowed with human politics. The argument is thus that data and datafication change practices and can bring forth novel layers and qualities of those practices. We explore data through a cosmopolitical approach using two empirical examples generated during 2013-2017, where the authors carried out ethnographic fieldwork in a project on governing and managing healthcare data. We conclude by proposing the term cosmo-data-politics and discuss the implications of this neologism.

Keywords: cosmopolitics, data, healthcare, ethnography, actor-network theory.

Introduction

Data and datafication - practices in which processes, life, and phenomena are turned into data in order to create some sort of value – are associated with great potential and optimism (Mejias & Couldry, 2019). States and government bodies, tech. companies, consulting firms, media and many others all contribute to the prevailing data optimism. At all levels of society––from the individual user of health apps, smart semi-AI applications installed in phones, cars, and the home, to businesses and public organizations, municipalities, regions and nations––data is considered key (McKinsey Global Institute, 2011). It seems that our current moment is one in which everything exists as data *in potentia*; yet to be datafied matter. Also, for many of us not working with the actual construction of data and algoritms in computer science or in tech corporations, data and datafication are often invisible or ungraspable, and the concrete material practices and circumstances under which processes, objects and relations become datafied, are inherently complex, opaque or secret (Burrell, 2016; Edwards, 2013; O'Neil, 2016; Ruppert et al., 2017; Wang, 2016). Premised by these general assertions about data, we propose an alternative notion of data drawing on the work of Belgian philosopher and science studies scholar, Isabelle Stengers. Based on the field of science and technology studies (STS) and her decades long conversation with Bruno Latour's work, Stengers' cosmopolitics entails that the cosmos (nature) and politics (the social) are inextricably entwined (Stengers, 2010, 2011a). Cosmopolitics implies uncertainty as an ontological condition, which means that it is impossible to definitively settle on what exists and to what consequence. As such, it implies that we should think and act in the presence of this uncertainty and as Stengers suggests "care for the possible" (Stengers, 2011b). The world is an inherently dynamic and surprising place and this must not, and cannot, be ignored (James, 1996a; Whitehead et al., 1978). However, the problem is that it is often, in technoscience as well as in politics, more convenient and common to think of the world in ideal terms as a fully knowable, representable and stable place where science and politics are neatly separated and compartmentalised (Bruno Latour, 1992; Pickering, 1995). On this basis, the article address how cosmopolitics can flesh out moments of datafication and help appreciate these moments as processes of emergence and creation. The article thus proposes and evinces an understanding of data and datafication as something that *adds* to and *transforms* the world in unanticipated ways. An understanding which is in contrast to more dominant ideas about data as representationalist, instrumentalist and reductionist.

The article describes and exemplifies the implications of addressing data cosmopolitically. We do so by, first, presenting and conceptualizing what cosmopolitics entails. Second, we will present the project we

followed during 2013-17 about the quality of healthcare data in the Central Region of Denmark. This part consists of a presentation of the project, our research methods and fieldwork, and the analysis of two events in the project. We conclude by discussing the implications of cosmopolitics for understanding data and data politics.

A symmetrical approach to data

To begin with, we propose that, in general, debates regarding data––both in public and academic life today––often express the view that data is digital, big or comprehensive, as well as of a magnitude that at first makes it seem incomprehensible (Schutt & O'Neil, 2013). Data can thus come in many forms. The very act of identifying and circumscribing something is an act of datafication, or a 'captafication' as Rob Kitchin suggests (Kitchin, 2014). In the first instance, what data is and is capable of, is thus impossible to fully decide or define apriori. This means that for analytic purposes we should be agnostic about the qualities and consequences of data and approach them symmetrically (cf. Callon, 1986).

Popular accounts of data abound, promoting ideas about data as absolute, rational, objective, and accordingly, as key to developing better, more efficient, fairer, more objective etc. practices in business and society at large. These ideas are often promoted by those in the business of selling the idea of being "data-driven" (Chris Anderson, 2008; McKinsey Global Institute, 2011; Science Staff, 2011). These accounts include certain *ontological* assumptions, namely that data are seen as *instruments* for businesses, governance and management, and as *representing* reality, as well as a *means* for improvement and progress. In research and studies of data, these accounts are challenged and elaborated further. It is argued that data require work and sensemaking in order to actually *become* data in the sense promoted by the popular accounts mentioned above (Bossen et al., 2019; boyd & Crawford, 2012; Dourish & Gómez Cruz, 2018; Gitelman, 2013; Wang, 2016). Some of these studies highlight how data has come to play a significant role in all sorts of practices, businesses and governance procedures, in a manner where data has become detrimental to human lives. For instance, take the data practices of predictive policing or insurance services (Eubanks, 2019; O'Neil, 2016) as two examples. What these accounts show, is that profit or allocation of resources oftentimes trump questions of fairness, justice and equality. O'Neil and Eubanks make the important point that the problem in relation to for instance predictive policing or insurance cannot be reduced to a matter of insufficient, incorrect or wrong data. Rather it is a lack of concern and consideration with the consequences of data, and with the particular situations and lives which data influeses, that is problematic. In that respect, the problem, following O'Neil and Eubanks, is exactly what is often considered the quality of data, namely its decontextualized and decontextualizing nature. We consider the work of O'Neil and Eubanks important in understanding the role of data in contemporary society and also that this entails investigating data with *empirical specificity* (Zuiderent-Jerak & Bruun Jensen, 2007). Data considered in general terms leads to general and accordingly limited insights. Therefore, we suggest a focus on specific practices, situations, or moments of data and datafication. In what follows, we focus on how data in specific situations come to play a role that challenges ideas about data as either neutral representations or endowed with human politics based on Stengers' cosmopolitics. The article thereby contributes to further our understanding of data as an unruly actor in more-than-human ontologies. But before we turn to this, we wish to make a few further assertions about data based on an agnostic and symmetrical understanding grounded in actor-network theory.

As we pointed out above data do not transcend practice. Data are products of practice. They are used and made sense of, and made to work in practice, as Tricia Wang, Paul Dourish and Rob Kitchin among others have pointed out (Dourish & Gómez Cruz, 2018; Kitchin, 2014; Wang, 2016). Neither are data monolithic, neutral nor transcendent. Also we want to add an additional point, following from actor-network theory (ANT) and generalized symmetry (Callon, 1986; Bruno Latour, 1987), namely that this is equally so for other types of 'data,' be they

narratives or ethnographic accounts. Just as digital data cannot stand alone, but needs to be narrated–– as Dourish puts it––or need thick descriptions, as Wang referring to the work of anthropologist Clifford Geertz, states, it is equally the case with *allegedly* rich, thick and qualitative accounts. We stress this not to suggest that one type of data, say a number, is the same or equal to another type, say an ethnographic narrative. The point is to be agnostic with regards to *any* type of data. Specifically, we think it is crucial to resist this sort of thinking about data, where thick, qualitative narratives are per se considered more extensive than digital data and this is a way of thinking that one might fall prey to, when it is argued that data needs thick accounts. This understanding mirrors the understanding that digital data can indeed now provide the fuller picture, clearly illustrated, and strangely enough, by Bruno Latour and Tommasso Venturini when they argue for the relevance of digital methods in social science:

> Thanks to digital traceability, researchers no longer need to choose between precision and scope in their observations: it is now possible to follow a multitude of interactions and, simultaneously, to distinguish the specific contribution that each one makes to the construction of social phenomena. Born in an era of scarcity, the social sciences are entering an age of abundance. In the face of the richness of these new data, nothing justifies keeping old distinctions. Endowed with a quantity of data comparable to the natural sciences, the social sciences can finally correct their lazy eyes and simultaneously maintain the focus and scope of their observations. (Venturini & Latour, 2010).

In the article, Latour and Venturini thus argue that in "an age of abundance" of data the social sciences can indeed follow and trace the social from the micro to the macro and thus presumably––finally––get a *full picture*, as if this has been the ambition of (all) social sciences all along.

Our point is not to suggest that the above is representative of Latours work, which has, in our opinion been about demonstrating the opposite, namely to problematize ideas about overarching essentialist structures or pre-existing transcendent orders, and accordingly, the ability to be able to produce a full picture (Callon & Latour, 1981; Bruno Latour, 1998, 2005). Our point is instead, that evidently even Latour may slip into a way of thinking about digital data that resembles popular understandings of digital data as potentially providing a full or fuller picture of reality. It is this sort of imagining of a full or fuller picture through data, that we find important to resist because it harbours and promotes a representationalist understanding of data and information. It relates also to the point made by Donna Haraway, Susan Leigh Star and Lucy Suchman, namely the partiality of every perspective. Every narrative is circumscribed, contingent and partial. No narrative, no matter how thick, long or rich, is a full account (Haraway, 1990; Bruno Latour, 1988; Star et al., 1994; L. Suchman, 2002; Lucy Suchman, 2007). We argue, that the above sketched representationalist understandings does not help us in appreciating what Andy Pickering terms a performative understanding of data as something that creates novelty and adds to the world (Pickering, 1995, 2011). Also, and related to refusing ideas about a full or fuller perspective helps remind us that the problem of any data or account is a matter of relation. Our experience of its richness, its adequacy or self-explanatory qualities depends on our specific relation to the data in question (Loukissas, 2019).

The symmetrical approach means that data are different products of different practices with different modes and capacities. If this is the case, as we claim it to be, it also means that we must be able to consider their ontological status as variable and ambiguous and in this regard Isabelle Stengers cosmopolitics is a helpful companion to think with.

Cosmopolitics

Isabelle Stengers' concept of cosmopolitics entails that we exist in a world in which the cosmos and human life, and how we arrange them (politics), are inherently intertwined, and accordingly we, as human

beings, must think, live and act with this as our condition (Stengers, 2000b). One implication of this is that Stengers is critical of both a social constructivist and realist assertion of science. Science is a practice in which scientists are hard at work at creating a situation––an event––in which an entity is made to exist in such a manner that it can be said to exist *autonomously* from the scientist. As an example, Stengers speaks of the neutrino's paradoxical mode of existence:

> [...] the neutrino is as old as the period in which its existence was first demonstrated, that is, *produced in our laboratories*, and [that] it *dates back to the origins of the universe*. It was both constructed and defined as an ingredient in all weak nuclear interactions and, as such, is an integral part of our cosmological models." (Stengers, 2010: 20-21 our italics).

First, it is important to note that this way of thinking about the neutrino seems paradoxical. One might immediately object to the idea that things can be both produced in laboratories and be a cosmological building block of the universe. You cannot have it both ways! But you can and we do, Stengers argues. Her point, borrowing from Latours concept of the factish, is that the world changes dramatically at the moment when the neutrino is produced in a laboratory and also in that respect becomes part of our cosmology. This event becomes consequential for how the universe is theorized and studied from that moment onwards, not to mention how it affects the invention of new technologies inside and outside of the lab (Stengers, 2010). On that basis it makes good sense to acknowledge the moment of production as indeed also a legitimate and relevant part of reality, instead of diminishing or deleting it from our understanding of the world. So the point is that the neutrino is real *and* as old as the universe and the moment in which it was realized through a very concrete, challenging, technological and constructed work process in a laboratory, is equally real. To choose between one or the other version implies a bifurcation of nature, which leaves us

with a poorer understanding of reality, not a more objective or correct one (Whitehead, 1920). Cosmopolitics holds that the production and construction of scientific facts make those facts more––not less––real (Jensen, 2004; Latour, Bruno, 2000; Bruno Latour & Stark, 1999). Cosmopolitics thus offers an irreductive way of thinking about science and reality. Science not only discovers and represents what the world consists of, it adds to the world and changes it. But cosmopolitics accordingly also means that how scientists conduct science can and must be scrutinized, which is indeed what Stengers does. When science adds to and not just depicts reality, then what it produces and how, becomes a crucial matter of concern. In that respect, Stengers is full of admiration of science practices that evoke novel qualities of reality based on a passionate interest in what it studies. But for the same reason, Stengers is highly skeptical of scientific practices that, under the banner of science, reduces or molests its objects or in an authoritarian manner claims to hold the only and objective truth about a given subject (Stengers, 2000b, 2000a). Stengers is critical of scientific practices that do not acknowledge that science is indeed a matter of knowledge production and as such always at risk of being wrong or of not having been able to create a situation in which the object of study can articulate itself in a manner that is not prefigured by the researcher (see also Despret, 2004; Despret et al., 2016; B. Latour, 2004).

Cosmopolitics implies that things and objects may be partially existing and that what exist in the world is a continuum of more or less existing objects rather than a matter of binary either/or (Latour, Bruno, 2000). Accordingly, we propose to think of data in a similar manner, namely, as a continuum between being human constructs and detached representations of reality. It seems trivial to point out, since evidently what data *are* at a given moment and place and with which consequences indeed varies dependent upon the circumstances.

We consider cosmopolitics to be a productive concept by which to study data for several reasons. First, it entails seeing data as both a constructed object that requires great effort and work in order to become data, while also acknowledging that data are not simply or only

a human construct. This position simultaneously resists the idea that data are 'objective facts' detached from human interests and that data are merely human constructs endowed with "human politics". Second, cosmopolitics suggests that what data produces or may produce, cannot be fully known, but must be curiously and closely investigated. And last, that datafication and data must be made "in the presence" of this uncertainty. What cosmopolitics implies is that data while may be under our control, they do not feel obligated by our human politics, and our datafication projects need to take that uncertainty into account.

In the following, we offer two cosmopolitical accounts of data that came out of an ethnographic fieldwork in the Danish healthcare sector. We consider the accounts to be exemplary of practices in which datafication plays a central role. Both accounts exemplify cosmopolitics, since they are simultaneously about constructing and retrieving data and they show how datafication can have emergent and novel consequences. In this regard we also claim to do empirical philosophy (Gad & Bruun Jensen, 2009; Mol, 2002)

Field and methods

The authors were invited together, along with other researchers from Aarhus University, to follow and study a project initiated by the Central Region of Denmark. The Region is the governing body for healthcare in the central part of Jutland, Denmark. Denmark is divided into five regions and the central Region is the second largest with approximately 1.3. million people. In 2013, the region proposed a pilot project in which nine different hospital wards were to be exempted from productivity measurement via the established DRG-system (Diagnosis Related Groups). In brief, the DRG system is the one through which the hospitals are reimbursed for the treatment procedures they carry out (Reinhard Busse, 2011) (R. Busse et al., 2013) (Bonde et al., 2018; Bossen et al., 2016). The Region initiated the project "New governance from the patient's perspective". The idea was to measure quality of treatment instead of productivity (number of treatments) and the nine wards were given full liberty to develop their own criteria and indicators for quality with which they would attempt to govern their wards. Examples of indicators were number of re-admissions of patients (fewer re-admissions indicates good quality of treatment), mortality rates (the lower the better), time from referral from general practitioner to diagnosis and treatment (the shorter the better), patients' satisfaction with treatment, and so on. The overall ambition of the project was to give healthcare professionals the autonomy to decide on the best treatment for their patients, dissociated from economic concerns. The project began in January 2014 and lasted three years. We were asked to follow the project and were offered full freedom to do so in accordance with the methods and theories we preferred (Bonde et al., 2018, 2019).

Our research project was an ethnographic qualitative study. We conducted qualitative interviews and observations aimed at following and understanding the development of indicators and infrastructures, and the concrete changes at the departments as a result of the re-direction of performance measurements towards quality and health benefits for patients. We conducted semi-structured interviews in 2015 and 2016 and did participant observations of meetings and workshops with heads of departments and region officials. Interviews lasted between 60 to 90 minutes. 25 interviews we conducted with head doctors and head nurses from the nine departments; two interviews with the management of a center, to which five departments belonged; and two interviews with staff from the business intelligence (BI) unit responsible for operationalizing indicators. Inspired by grounded theory (Glaser & Strauss, 1967), the interviews were transcribed and coded by means of qualitative software by all three authors. The accounts below are comprised of events that occurred across several departments.

Partially existing data

In this first example, we detail how the wards, in the beginning of the project, decided upon a range of indicators in an effort to measure and govern quality. However, it quickly became apparent that choosing

such indicators was a complicated and demanding process. The idea, central to the overall project, of building a data-driven governance infrastructure based on indicators was thus much more difficult to realize than first assumed. Data existed, but in ambiguous ways in different places and formats. The following account unfolds this and offers a cosmopolitical response.

Denmark is at the forefront of IT infrastructures for healthcare with all citizens having personal id numbers used, amongst other things, for tax, work and health purposes. All five Danish Regions have implemented electronic health record systems (EHR's) that allows for the collection and processing of patient data. In addition, Denmark also has a substantial number of national clinical quality databases, each of which collects and process data about each patient's disease history. This means that there is already many indicators and data on quality of treatment. Additionally, the right to define and select their own quality criteria and indicators only added to this already abundant availability. In total, the nine wards came up with over 100 different indicators. Each ward handpicked those that fit their medical specialization. The idea of having a handful of cross ward general indicators seemed, from early on, unrealistic.

Gathering data on the 100 plus indicators––for instance mortality rate or time from referral to treatment––turned out to require extra work and collaboration amongst clinicians and IT-technicians. Even though the departments had experience with documenting and registering indicators, acquiring new, or re-purposing existing, data proved extensive and challenging. In some instances, existing data from the EHRs could be repurposed and used for the project. This was the case for 57 indicators. However, for the remaining 43 indicators either a lot of work and expertise were required, or data retrieval turned out to be impossible.

Repurposing data from EHRs to support indicators required extensive collaboration between clinicians and the data workers at the business intelligence (BI) unit (See figure 1 below for a simple graphical representation of the central bodies and their relations in the project).

For instance, clinicians at one department had chosen 'non-attending patients' as an indicator and aimed for a 20% reduction of this group in order to increase efficiency. But this required negotiation and discussion between the clinicians and the BI Unit staff. What was needed in this specific case was to decide upon a baseline, and whether the 20% was a decline in absolute numbers (e.g. from 100 to 80 patients) or a decline in percentage points (e.g. a reduction from 10% to 8%). The data-worker at the BI Unit required more information in order to be able to "…tell the data how to behave…". The data-worker had to develop the scripts and algorithms required to process the––in principle––already available data (Interview with data-worker 1, BI Unit). Working out indicators, even with existing data, was dependent upon a dialogue between clinicians and data-workers, since the former were experts on clinical practices, but not on data retrieval, accumulation and analysis.

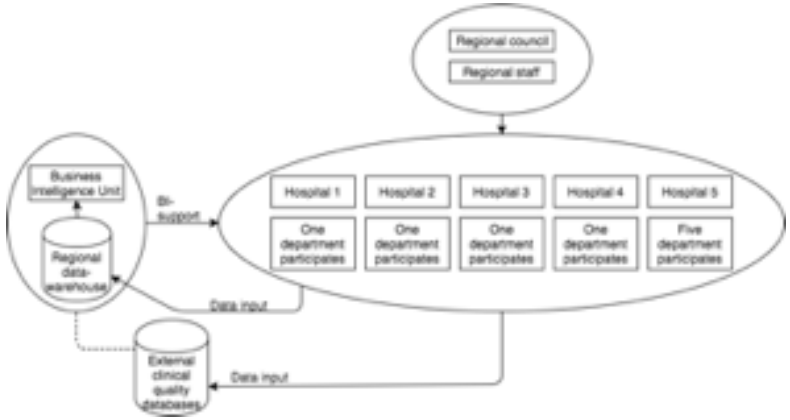


Figure 1. This diagram shows the organizational set-up of the project and the relations to the external bodies relevant to the building of indicators

Acquiring data from national quality databases also required collaboration and clarification between clinicians and data-workers, and in some cases had the additional challenges of limited access to data and incompatibility with the Region's own systems. These national research databases are administered by medical interest groups and act as

quality and research repositories for different specialties such as back surgery, head and neck cancer and others.[1] However, the departments that wanted to utilize this data learned that they could only retrieve data on a yearly or half-yearly basis and not continuously as they had imagined, and which was important in order to establish a near to real-time assessment of quality. Second, they found out that data was not easily retrieved, because the data formats of the databases were incompatible with the EHR and the BI Unit's IT systems. Hence, some of these indicators had to be discarded or needed to be established in other ways.

The necessity of interdisciplinary collaboration and limitations arising from existing IT infrastructures also became apparent for the departments that strived to generate data on the 'patient's perspective'. This turned out to involve a lengthy, and to some frustrating, dialogue between the departments and the Region concerning the development of a questionnaire. Agreeing on what the 'patient perspective' entailed and on which questions to ask across diverse patient groups proved challenging. As the head of one department stated:

"We said 'we're in, but you [the Region] have to help us', because you know about this [the patient's perspective]. …And we have had numerous discussions about who is to measure the patient's perspective. We can't! It is naive to ask a small department to develop such a product, when even the quality unit of the Region cannot accomplish it. We have had six or seven meetings with the quality unit by now…"

In the end though, a questionnaire was developed. But implementing the questionnaire at the departments required the development of techniques for gathering, accumulating and making data from patients accessible. Questionnaire data was gathered either by nurses or Red Cross volunteers at discharge, both of whom required renegotiations of work agreements. Paper questionnaires meant that the response was transcribed and added to a common sheet (analogue or digital). Using tablet PC's alleviated this work, but made data generation vulnerable

[1] The databases in question are Danespine, Dahanca, Thykir, Rhino.

to infrastructural contingencies: network connections were unstable, or login requirements posed difficulties to patients and volunteers. Furthermore, these data were not compatible with the existing standards of the Region's data warehouse and thus required substantial efforts in developing an IT-interface. Thus accumulation, presentation, and distribution of these data could not be automated in ways similar to the other indicators.

In these instances, we see healthcare professionals expressing various degrees of frustrations and surprise with regard to the challenges they encountered with data. We see them work to transfer or produce the data they assumed to be readily at hand. Generating data required extensive work as well as collaboration between clinicians and data workers at the BI-unit, all of which was contingent upon existing data infrastructures. In different ways the people involved in the project, were challenged by the difficulties of realizing data that were presumed to be already available. But the point, from a cosmopolitical perspective, is that the practitioners were not wrong to assume data availability. The problem was that although data preexisted in some form, it still required work, effort and configuration to the particular practices at hand. From a more-than-human ontology that cosmopolitics implies, the example shows that data evades representational and instrumental understandings. It preexists 'out there' and is in some form already available. Cosmopolitics dissolves this perplexity, because indeed data can and do exist *and* require work and configuration. Cosmopolitics thus interferes with predominant ways of thinking about data as either available or not. Consequently, it offers alternative strategies and dispositions towards data projects by equally mitigating a naïve data as "plug'n'play" understanding and the disappointment and frustration that may follow what is become evident that it is not.

In the next account, we show how datafication is productive in surfacing complexity and as such exemplifies another cosmopolitical point, namely datafication as processes of emergence and creation:, data as event.

Datafication as emergence and creation

One department was concerned with high numbers of surgery cancellations, seeing these as detrimental for the quality of patient experiences. Therefore, they decided to count the number of cancellations in order to decrease them. However, when reviewing cancellations, it turned out that cancellations were not one specific thing. It was necessary to distinguish between four different types, some of which were detrimental to the organization (staff and equipment was idle; other patients waiting were not treated etc.), while others were not (overscheduling), and yet others actually beneficial (a patient was treated earlier than scheduled). This, in turn, led the department to develop procedures for measuring the detrimental cancellations in order to specifically reduce them, and furthermore enabled the department to consider the reasons for cancellations. They wondered why patients scheduled for surgery cancelled or simply stayed away on the day of surgery. This led the department to investigate the problem. They interviewed some patients and discovered that patients that initially had decided to have surgery, sometimes changed their minds, when they had had the time to reflect on and discuss the procedure with their relatives. The department concluded that, ironically the problem was that despite the best of intentions, they provided too swift and efficient a service by immediately (after diagnosis) giving the patients the opportunity to sign up for surgery. Hence, the patients were not given the time to consider the pros and cons of the surgical procedure and then decide whether or not they actually wanted it. The example illustrates how the attempt to manage a specific problem––a high level of cancellations––led first to the attempt to measure the problem and turn it into data as a simple number of daily cancellations; then to a further development and specification of the problem; and finally, to the concern as to whether patients were actually provided the conditions that enabled them to be sufficiently involved in decision-making. This process is evidently a process of emergence. It exemplifies how the attempt to turn a problem into data is interrupted and becomes complicated, leading to both a more specific and targeted data collection, but more importantly to crucial insights about organizational matters––such as patients concerns and reasoning. Ultimately, it could potentially lead to improved quality of treatment and efficiency and resource management.

Another department wanted to reduce the number of re-admissions, which is often taken as a (negative) quality indicator, since re-admissions are often, and for obvious reasons, considered indicative of poor quality treatment. In addition, re-admissions are burdensome for patients and the healthcare system in general. However, the problem in relation to data and performance indicators is how to differentiate between preventable and non-preventable re-admissions? Just counting re-admissions and deciding upon an acceptable rate is not sufficient, because some re-admissions––for example, those that are not due to maltreatment of some sort but to a worsening of the patient's condition for other reasons––are good and should therefore not be counted. So instead, an analysis and evaluation of each re-admission was required. Thus, the ambition to reduce re-admissions began as a matter of just counting them, which was then quickly realized as insufficient and meaningless, because the actual matter of concern was to discern between 'good' and 'bad' re-admissions and this required a much more in-depth analysis.

One last example concerns a department's ambition to have several diagnostic tests planned and performed on the same day, instead of patients having to come to the hospital multiple times. This required defining an indicator for the number of diagnostic tests a patient should receive during a hospital visit. But as a physician noted:

> So, what is the right number of tests per day? Is it ten? Just to suggest a random number. But what if by the eighth test the diagnosis is established? Then of course you should not do the last two tests, just because there is an indicator saying ten. And what if the diagnosis is established after just two tests?

As with re-admission, failing to meet the standard set by the indicator might both indicate negligence and excellence. Deciding upon a standard in order to measure performance, also in this case required further investigation.

In summary, it may be difficult to turn phenomena and concerns into data due to their complicated nature. Although this may be seen as a challenge to ambitions of data-driven healthcare, we wish to stress its value. Our examples show that although data production may be hampered, knowledge production is not. The attempt to produce data may, as we have illustrated here, lead to a more profound understanding of a specific problem and provide an insight into organizational and clinical concerns. Although one might be disappointed that re-admissions, cancellations, and same-day treatment, turn out to be complicated problems to 'datafy', professionals, nonetheless, can gain crucial insights. Datafication can, as these cases illustrate, thus be understood not as detrimental and reductive of real-life matters, but as processes by which these matters emerge, become articulated and ultimately taken care of.

From a cosmopolitical perspective, we would like to point out how the attempt to datafy produces a novel situation in which what is presumed to be relatively simple––countable events, such as cancellations––turns out to be more complicated and in need of reconceptualizations, new taxonomies and accordingly, different actions. We consider this to be a matter of cosmopolitics, not only because it stimulates reflections on the limitations of the very thing that initiated the process, namely datafication, but more importantly because it produces a situation in which something new is learned. Datafication comes to a stop. It is not an all sweeping territorialising event, but becomes concretised and constrained in, and with, the particular practices in which it is intended to be productive.

Cosmo-data-politics

In this article, we have proposed studying data by way of cosmopolitics.

We have argued that cosmopolitics enable us to study and appreciate data and datafication as ambiguous and as both already existing and something to be constructed. Thereby we resist notions of data as either given or as something to be constructed: data are indeed both. Furthermore, cosmopolitics entails the ability to observe and appreciate datafication as processes of emergence and creation, which, in the end, may moderate data ambitions. To help us think about this we propose the term cosmo-data-politics.

Cosmo-data-politics implies that we cannot know what data are capable of, and that they must be studied and analysed with empirical specificity. We have attempted this by providing examples of datafication processes in a hospital setting, and by demonstrating how they can be conceptualized as cosmopolitical. Cosmo-data-politics is about resisting simple assumptions about data, such as that data per se leads to improvement or violence and that it, by definition, is a human construction and instrument. As many other things in a more-than-human world, data and datafication escapes human mastery in various ways and cosmo-data-politics implies that our data projects and our data ambitions must take this into account. We suggest, again referring to Stengers, that we should think of data as a pharmakon (Stengers, 2010). A pharmakon is an agent that in certain doses are poisonous, whereas in others are nurturing and invigorating. Whether it is the one or the other is dependent upon the subject to which it is applied. Following this train of thought, a central cosmo-data-political concern is that the qualities and uses of data must be analysed and evaluated in the presence of those to whom it matters. This may seem a trivial point perhaps, but it nonetheless goes against many of the prevalent ideas about data as detached and decontextualized. In fact, it is exactly detachment, which is often considered a main strength of data. But thinking with the term cosmo-data-politics, we argue that data can be thought of as a type of cosmos. What this entails is that data may, and oftentimes will, come to matter and have consequences beyond the mastery of its human initiators. As such, similar to the term cosmos, they may be indifferent to human politics and intentions. And it is this uncertainty that we as

human actors need to take that into account: we do not control the consequences of "our" datafied actions. Consequently, the more data are promoted as detached or universal, the more concerned we should be and the more we should work to bring them into the presence of those they affect. Take, as one example, those who are 'managed' via data, such as the less privileged described by O'Neil and Eubanks (Eubanks, 2019; O'Neil, 2016). But cosmo-data-politics for the same reason also implies a consistent curousity about what data do, a curioustity about what it may do, what kind of surprises and unexpected consequences they may produce. So cosmo-data-politics resists idealist notions of data and is instead about exploring data usage and datafication with a passionate interest in what data do and how and whose existence they potentially transform.

Cosmo-data-politics sees data and datafication as processes that add to the world and *potentially* create learning and novelty, while at the same time resisting data as detached, simple instruments. In other words, cosmo-data-politics entails that data projects are looked upon for their evocative potentials *and* are conducted 'in the presence' of those to whom they come to matter.

Acknowledgements

The authors would like to thank all clinicians and the Business Intelligence staff participating in interviews for their time and detailed responses. Furthermore, the authors thank the wider research group around the project and DEFACTUM for organizing the research project. Also, thanks to the reviewers and editors for carefully reading and commenting on the article.

References

Bonde, M., Bossen, C., & Danholt, P. (2018). Translating value-based health care: An experiment into healthcare governance and dialogical accountability. Sociology of Health & Illness. https://doi.org/10.1111/1467-9566.12745

Bonde, M., Bossen, C., & Danholt, P. (2019). Data-work and friction: Investigating the practices of repurposing healthcare data. Health Informatics Journal, 25(3), 558–566. https://doi.org/10.1177/1460458219856462

Bossen, C., Danholt, P., & Klausen, M. B. (2016). Diagnoser som styringshybrider Diagnoserelaterede grupper i sundhedsv\a esenet. Tidsskrift for Forskning i Sygdom Og Samfund, 13(25).

Bossen, C., Pine, K. H., Cabitza, F., Ellingsen, G., & Piras, E. M. (2019). Data work in healthcare: An Introduction. Health Informatics Journal, 25(3), 465–474. https://doi.org/10.1177/1460458219864730

boyd, danah, & Crawford, K. (2012). CRITICAL QUESTIONS FOR BIG DATA: Provocations for a cultural, technological, and scholarly phenomenon. Information, Communication & Society, 15(5), 662–679. https://doi.org/10.1080/1369118X.2012.678878

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

Busse, R., Geissler, A., Aaviksoo, A., Cots, F., Hakkinen, U., Kobel, C., Mateus, C., Or, Z., O'Reilly, J., Serden, L., Street, A., Tan, S. S., & Quentin, W. (2013). Diagnosis related groups in Europe: Moving towards transparency, efficiency, and quality in hospitals? BMJ, 346(jun07 3), f3197–f3197. https://doi.org/10.1136/bmj.f3197

Busse, Reinhard (Ed.). (2011). Diagnosis-related groups in Europe: Moving towards transparency, efficiency and quality in hospitals. Open Univ. Press.

Callon, M. (1986). Some Elements of a Sociology of Translation: Domestication of the Scallops and the Fishermen of Saint Brieuc Bay. In J. Law (Ed.), Power, action, and belief: A new sociology of knowledge? Routledge & Kegan Paul.

Callon, M., & Latour, B. (1981). Unscrewing the big Leviathan: How actors macro-structure reality and how sociologists help them to do so. In K. Knorr-Cetina & A. V. Cicourel (Eds.), Advances in social theory and methodology: Toward an integration of micro- and macro-sociologies. Routledge & Kegan Paul.

Chris Anderson, A. (2008). The End of Theory: The Data Deluge Makes the Scientific Method Obsolete. WIRED. https://www.wired.com/2008/06/pb-theory/

Despret, V. (2004). The Body We Care for: Figures of Anthropo-zoo-genesis. Body & Society, 10, 111–134. https://doi.org/10.1177/1357034X04042938

Despret, Vinciane, Buchanan, B., & Latour, B. (2016). What would animals say if we asked the right questions? University of Minnesota Press.

Dourish, P., & Gómez Cruz, E. (2018). Datafication and data fiction: Narrating data and narrating with data. Big Data & Society, 5(2), https://doi.org/10.1177/2053951718784083

Edwards, P. N. (2013). A vast machine: Computer models, climate data, and the politics of global warming (First paperback edition). The MIT Press.

Eubanks, V. (2019). Automating inequality: How high-tech tools profile, police, and punish the poor.

Gad, C., & Bruun Jensen, C. (2009). On the Consequences of Post-ANT. Science, Technology & Human Values, 35(1), 55–80. https://doi.org/10.1177/0162243908329567

Gitelman, L. (Ed.). (2013). "Raw data" is an oxymoron. The MIT Press.

Glaser, B., & Strauss, A. (1967). The Discovery of Grounded Theory: Strategies for Qualitative Research. Aldine Transaction.

Haraway, D. J. (1990). Simians, Cyborgs, and Women: The Reinvention of Nature (1st ed.). Routledge.

James, W. (1996a). Essays in radical empiricism. University of Nebraska Press.

Jensen, C. B. (2004). Researching partially existing objects: What is an electronic patient record? Where do you find it? How do you study it? Science Technology Society.

Kitchin, R. (2014). The data revolution: Big data, open data, data infrastructures & their consequences. SAGE Publications.

Latour, B. (2004). How to Talk About the Body? The Normative Dimension of Science Studies. Body & Society, 10(2–3), 205–229. https://doi.org/10.1177/1357034X04042943

Latour, Bruno. (1987). Science in action: How to follow scientists and engineers through society. Harvard University Press.

Latour, Bruno. (1988). The Politics of Explanation: An Alternative. In S. Woolgar (Ed.), Knowledge and Reflexivity, New Frontiers in the Sociology of Knowledge (pp. 155–177). Sage.

Latour, Bruno. (1992). We have never been modern ([3. print.]). Harvard Univ. Press.

Latour, Bruno. (1998). Paris ville invisible (Les empecheurs de penser en rond). Institut Synthelabo pour le progres de la connaissance.

Latour, Bruno. (2000). On the Partial Existence of Existing and Nonexisting Objects. In L. Daston (Ed.), Biographies of Scientific Objects (pp. 247–269). University of Chicago Press.

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

Latour, Bruno. (2005). Reassembling the social: An introduction to actor-network-theory. Oxford University Press.

Latour, Bruno, & Stark, M. G. (1999). Factures/Fractures: From the Concept of Network to the Concept of Attachment. RES: Anthropology and Aesthetics, 36, 20–31.

Loukissas, Y. A. (2019). All data are local: Thinking critically in a data-driven society. The MIT Press.

McKinsey Global Institute. (2011). Big data: The next frontier for innovation, competition, and productivity. McKinsey Global Institute. https://www.mckinsey.com/business-functions/mckinsey-digital/our-insights/big-data-the-next-frontier-for-innovation#

Mejias, U. A., & Couldry, N. (2019). Datafication. Internet Policy Review, 8(4). https://doi.org/10.14763/2019.4.1428

Mol, A. (2002). The body multiple: Ontology in medical practice. Duke University Press.

O'Neil, C. (2016). Weapons of math destruction: How big data increases inequality and threatens democracy (First edition). Crown.

Pickering, A. (1995). The mangle of practice: Time, agency, and science. University of Chicago Press.

Pickering, A. (2011). The Cybernetic Brain: Sketches of Another Future. University Of Chicago Press.

Ruppert, E., Isin, E., & Bigo, D. (2017). Data politics. Big Data & Society, 4(2), https://doi.org/10.1177/2053951717717749

Schutt, R., & O'Neil, C. (2013). Doing data science (First edition). O'Reilly Media.

Science Staff. (2011). Challenges and Opportunities. Science, 331(6018), 692–693. https://doi.org/10.1126/science.331.6018.692

Star, S. L., Odense Universitet, Department of Feminist Studies, & Feminist Research Network: Gender-Nature-Culture. (1994). Misplaced concretism and concrete situations: Feminism, method and information technology. Feminist Research Network, Gender-Nature-Culture.

Stengers, I. (2000a). For en demokratisering av vitenskapene. Spartacus.

Stengers, I. (2000b). The invention of modern science. University of Minnesota Press.

Stengers, I. (2010). Cosmopolitics I (R. Bononno, Trans.). Univ Of Minnesota Press.

Stengers, I. (2011a). Cosmopolitics II. University of Minnesota Press. University Presses Marketing [distributor].

Stengers, I. (2011b). The Care of the Possible: Isabelle Stengers interviewed by Eric Bordeleau (E. Bordeleau, Interviewer) [Interview]. http://www.scapegoatjournal.org/docs/01/01_Stengers_Bordeleau_CareOfThePossible.pdf

Suchman, L. (2002). Located accountabilities in technology production. Scandinavian Journal of Information Systems, 14(2), 7.

Suchman, Lucy. (2007). Agencies in technology design: Feminist reconfigurations. Unpublished Manuscript. http://www.lancs.ac.uk/fass/sociology/research/publications/papers/suchman-agencies-technodesign.pdf

Venturini, T., & Latour, B. (2010). The Social Fabric: Digital Traces and Quali-quantitative Methods. Proceedings of Future En Seine, 30–15.

Wang, T. (2016, January 20). Why Big Data Needs Thick Data. Ethnography Matters. https://medium.com/ethnography-matters/why-big-data-needs-thick-data-b4b3e75e3d7

Whitehead, A. N. (1920). The Concept of Nature: Tarner Lectures. Cambridge University Press.

Whitehead, A. N., Griffin, D. R., & Sherburne, D. W. (1978). Process and reality: An essay in cosmology. Free Press.

Zuiderent-Jerak, T., & Bruun Jensen, C. (2007). Editorial Introduction: Unpacking 'Intervention' in Science and Technology Studies. Science as Culture, 16(3), 227–235. https://doi.org/10.1080/09505430701568552

Author Bios

Peter Danholt is associate professor at Information Studies, Aarhus University. His research is on the role of IT and data in healthcare, organizations, design and daily life. His research draws on STS, feminism, post-structuralism and actor-network theory. He is vice president of the Danish Association of STS (DASTS), chief-editor of STS Encounters.

Morten Bonde Klausen earned his PhD degree in 2015 from Aarhus University at Information Studies. After his PhD, Morten worked as post.doc doing qualitative research in value-based healthcare and health care financing. Currently Morten works as a senior research analyst at the Danish Center for Social Science Research (VIVE), still engaged with projects in the health care domain.

Claus Bossen is associate professor, Aarhus University, has a PhD in Anthropology and his research focuses on qualitative studies of the intertwining relations between IT and human practices. His research interest include healthcare, medical informatics, participatory design, and Science-Technology-Society. His most recent research includes value-based healthcare, as well as emergent occupations and data work in healthcare.