Accessibility statement

This is an accessibility statement for the journal: Encounters.

Conformance status

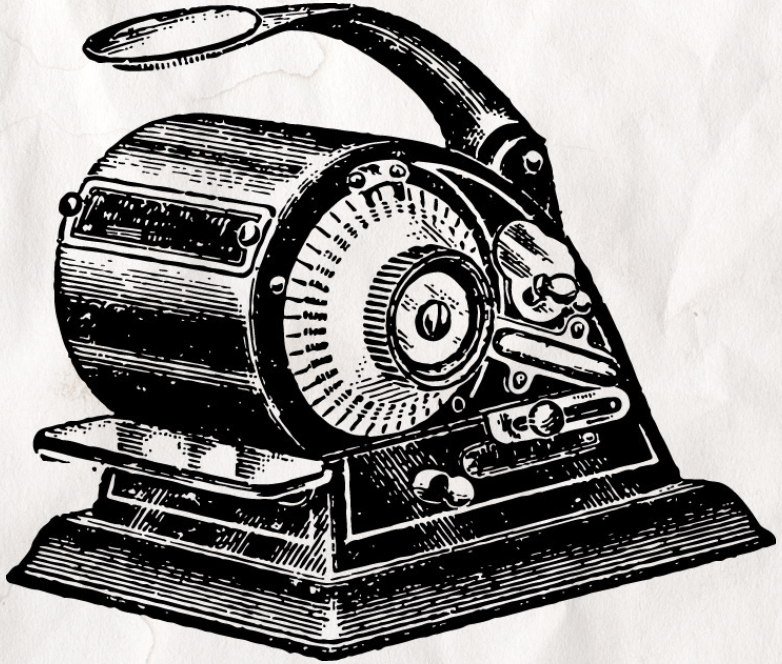
The Web Content Accessibility Guidelines (WCAG) defines requirements for designers and developers to improve accessibility for people with disabilities. It defines three levels of conformance: Level A, Level AA, and Level AAA. This statement is relevant for volume 10, number 5, 2018 through volume 12, number 1, 2021. Encounters is partially conformant with WCAG 2.1 level AA. Partially conformant means that some parts of the content do not fully conform to the accessibility standard. Despite our best efforts to ensure accessibility, footnotes and graphs may not be accessible for screen readers at this point in time.

Feedback

We welcome your feedback on the accessibility of the journal. Please let us know if you encounter accessibility barriers. You can reach us at:

E-mail: imvko@cc.au.dk
Address: STS Center, Helsingforsgade 14, 8200 Aarhus N

ENGAGING THE DATA MOMENT



Special issue
Volume 11• Number 1 • 2020

STS Encounters

SPECIAL ISSUE

Volume 11• Number 1• 2020

Data criticality

Mareile Kaufmann
Department of Criminology and Sociology of Law, University of Oslo, Norway

Nanna Bonde Thylstrup
Department of Management, Society and Communication, Copenhagen Business School, Denmark

J. Peter Burgess
École normale supérieure, Paris, Chaire Géopolitique du Risque, France

Ann Rudinow Sætnan
Department of Sociology and Political Science, Norges teknisk-naturvitenskapelige universitet, Trondheim, Norway

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

The data moment, we argue, is not a single event, but a multiplicity of encounters that reveal what we call 'data criticality'. Data criticality draws our attention to those moments of deciding whether and how data will exist, thus rendering data critically relevant to a societal context and imbuing data with 'liveliness' and agency. These encounters, we argue, also require our critical engagement. First, we develop and theorize our argument about data criticality. Second, by using predictive policing as an example, we present six moments of data criticality. A description of how data is imagined, generated, stored, selected, processed, and reused invites our reflections about data criticality within a broader range of data practices.

Keywords: Data, critique, criticality, predictive policing, digital

Introduction

Ever more powerfully and in an increasing number of ways, data have become critical in two senses of the word: firstly, by being decisively important for generating, structuring, and carrying knowledge and, thus, key to the generation and sustenance of life as we know it. By their ubiquity and agency in our lives, data have become our 'companion species', affecting us in ways that are in part beyond our control (Lupton 2016; Bellanova 2016; based on Haraway 2008). Secondly, data are critical in the sense of being a mirror to society whose essential knowledge they are intended to contain, but to which they are never entirely a simple servant. These two meanings of data converge and intertwine in events and encounters that reveal when digital data become critically relevant to a lived context. Analysing these moments helps us understand how data come into being, how they are worked with and put to work, and how they play their companion-species role(s) in our lives. Thus, these moments in which data become critical to societal life require our critical engagement. This article offers a conceptual and methodical analysis of data criticality, framing what 'critical data scholarship' (Lupton 2016) can mean in practice.

In the widest sense data are the foundation of any type of knowledge. This analysis, however, focuses on digital data practices. The fusion of knowledge practices with digital data––the discrete, discontinuous units of information that take the form of binary code––is a key event that gives rise to a broad variety of socio-digital practices that warrant our attention. Moreover, the digitization of data practices implies a crucial qualitative shift in the relationship between data and surveillance that came with the invention of the Internet in the 1960s. The Internet is in its essence a self-surveilling, digital network management machine born out of the organizational need for managing shared data and inseparable from ubiquitous surveillance (Chadwick 2006: 257-287; Zuboff 2019). More than an infrastructure of cables and servers, nodes and connections, the Internet is an ecosystem constituted and sustained through the circulation of digital data as synthesized units of information. Yet, while one constituting function of the Internet is to facilitate the flow of digital data, another is to catalogue, label, direct, and monitor digital data flows. Thus, in its most primordial form, the Internet is a surveillance system that contains and follows data. It is impossible to plug into the internet, let alone participate in the social intercourse of Internet 2.0, without also participating in dataveillance, be it as individual citizen, group, organization, or business.

One data practice that not only derives from, but also inspires new forms of dataveillance is predictive policing, which has gained considerable attention in recent years, although in-depth knowledge about its various data moments is still rare. Furthermore, the concept of criticality has a tendency to mark a political divide in the literature: the embrace of digital data and methods is either seen as critical in rendering police work more efficient and proactive (Ratcliffe 2004; Pearsall 2010), or such data practices are discussed from a critical perspective (Bennet Moses and Chan 2018; Degeling and Berendt 2018). In this paper, we draw attention to the encounters that underline how data become critical to a specific context, while also warranting

our critical attention.

Analysis of these moments is based on an interview study conducted by Mareile Kaufmann with experts, police officers, software designers, and ICT engineers on the specifications of seven predictive policing software models with origins in three different continents. The aim of the study was to understand in greater depth how digital data and technologies create new knowledge practices, rationalities, and concepts connected with crime control in a field that has a long-standing history of exercising surveillance, data analysis, and prediction. As in many other domains, digital data and analytic instruments used for predictive policing are embedded in many (non-linear) circuits and intersect with many lives. This encouraged an attempt to trace these circuits and identify moments in which data are rendered critical and require critical engagement. Quotes and insights from this study––marked with fictional first names––are selected to illustrate these moments. It should be noted that data imaginaries, generation, storage, selection, processing, and reuse empirically relate to different aspects of predictive policing, but they also serve as a more generalized catalogue for similar moments in other digital practices and fields. In order to create a framework for these empirical insights and their discussion, we first give more substance to our notion of data criticality.

Data criticality

Any engagement with data is a critical event. Data produce social and political meaning the instant they are set in a specific context and associated with other data. As noted above, this moment of engagement is 'critical' in two senses: first, in that it implies a moment of decision (ancient Greek: *krinein*), that is, the moment of their affiliation with other data and of a decision or determination of their form of existence. Decisions are made in the moment when data are ascertained in a given context, when they are imagined, generated, collected, stored, recycled, and chosen as a proxy or representation for a phenomenon. Part of an interpretative processing of the world, they are removed from their logical status as purely given and attached to the contextual elements through which they acquire and transmit meaning. Second, 'data criticality' has a normative meaning due to the political need that springs from the first sense, which is to remain vigilant to the political character of data. The first meaning is the fruit of critical observation, the second describes the sense of the political action.

There are myriad means by which humans and infrastructures coalesce data into meaning, each impacting on the destiny of data in its own way. Thus, these moments also warrant careful reflection about how data is constituted as relevant. If data have become crucial to society to the point of becoming a companion species, this companionship is multi-faceted and follows multiple trajectories, thus requiring a stepwise analytical approach to insights into data's roles in our lives. The concept of data criticality invites our engagement with "the possibilities for critical renewal that everyday companions might suggest" (Austin et al. 2019b: 5). Importantly, then, the purpose of data criticality is not to pass judgement on all data and data practices (cf. Felski 2012); rather, the concept can help us attune to the moments in which data attain meaning and what this means for their––and our own–– situation in the data ecosystem.

In pointing to data's dual criticality, we align ourselves within a history of theories on the relations between data and society: Merton's (1942) CUDOS concept, the empirical program of relativism (Collins 1981), actor network theory (e.g. Latour 1987) marxist and feminist standpoint theories (e.g. Marx, no date: 2nd edition, postface; Hartsock 1983, Harding 1991), agential realism (Barad 2007) have all sought to explain how data can at once be obviously social products yet also represent, be impacted by, and impact upon a world of realities seemingly beyond social determination.

Looking across all these theory categories, we see that data tend to be treated as stable products, an 'immutable mobile' in Latour's (1987) terms, that can carry information intact from one context to another. At the same time, data are also seen as animate in or animated by the precise moment a scientist interacts with them. In this article, we build

from the more or less common ground of the theoretical frameworks mentioned above to examine this "liveliness of data" and how it becomes critical in multiple senses (Ruppert et al. 2013: 29; see also Lupton 2015). Contrary to their reputation as technical, binary, and objective information we show that digital data cannot be divorced from the moments that we are describing in this article. How and according to what norms and grammar are digital data assembled? How are they made sense of? Who and what are part of making decisions and interpretations, and of translating data from one context into another? In describing these changes, we provide a catalogue of the different ways in which we can observe the criticality of data and think about data critically. In the following we use the emergent relationships in which data are situated as a starting point for describing how data criticality becomes a core property of the networks that suffuse and surround them.

Data and moments of meaning-making

While data are constantly dynamic, there are key moments that particularly reveal how they become critical: when they are imagined, generated, stored, selected, processed, discarded, and reused. These moments are at once temporal events and modularizing processes: in these moments, it becomes obvious how data become amenable to being associated, merged, or combined with other data. At each of these encounters, assemblages of designers, scientists, engineers, scholars, professionals, users, and target groups, as well as machines, routines, attitudes, concepts, and preconceptions collaborate in order to render data critical in a specific context or for a particular purpose. These collaborations can be observed over the course of many years and in different environments that are organized around the making and shaping of dataveillance. Using the case of predictive policing as an example, we illustrate data criticality with six moments that serve as inspiration to reflect about data criticality. They portray critical entry points for analysing how other digital practices are also co-constituted by many actors and involve––maybe similar, maybe different––moments that bring "liveliness" (Ruppert et al. 2013: 29) to data.

Imagining data

As Evelyn Ruppert (2018) notes, some of the most forceful 'socio-technical imaginaries' (Jasanoff and Kim 2009) we face are those involving digital technologies and data gained via dataveillance. These imaginaries drive and frame many of the critical data infrastructures with which we surround ourselves and on which we base our lives, politics, and decision making. Data feed an imaginary of form, especially within the various fields of prediction; their digital format not only fits but invites continued pattern recognition. The imaginary of digital data as liquid and malleable––we can drown in the 'data deluge' (Bevan 2015), be overtaken by a 'data tsunami' (Rubinstein 2013), fix leaks through 'data plumbing' (Davenport 2014), and even 'sweat data' (Gregg 2015)––proposes their endless re-evaluation for forms and patterns. Digital data encourage the identification of correlative shapes, not necessarily explanations (cf. Striphas 2015; Kaufmann et al. 2019a). This imaginary of data as susceptible to form integrates well with predictive policing, since both mainly work with plausible suggestions about patterns and not why phenomena come into being. Frank, who works on software for predictive policing, confirms this:

> That was what the basic research was all about: to figure out the mathematical structure or phenomena of crime patterns. And when you understand the general structure of that, then you can use that as a basis for a general learning process. … And just to be clear: we're only focused on predicting where and when crime is most likely to occur. We don't predict why or how or who. Those are things that our particular process doesn't focus on.

Digital data do not stand for the idea that all identifiable forms are

meaningful. However, in the context of policing, digital data fuel the quest for the pattern that 'works'. Data imaginaries of liquidity and formability are coupled with ideas about which of these malleable datasets works best to identify meaningful forms. Here, the data imaginaries become more refined. While still understood as yielding practicable and actionable patterns, data imaginaries tie in with methodologies about the right choice of dataset and correlative methods. These include data-opportunistic approaches, such as relating police datasets to any available data, as well as approaches that work with more selective datasets. Chris, for example, who works on prediction models, explains his data imaginary. He correlates police data "with various other statistics, like weather being one, traffic data … basically you use whatever data you have available. It's very opportunistic. … the number of people buying headache medicine … The more you know, the better system you can make."

Other imaginaries and approaches include pre-processing to further define patterns of interest. Amanda works on a project developing prediction software for policing purposes. In contrast to the imaginary of 'big data', she describes a rather focused and selective data imaginary. Amanda explains how she and her team discussed and tested which data they considered relevant for meaningful predictions, thus also formulating how the teams' specific imaginary of 'select data' unfolded:

> We created an index including socioeconomic status, because there is research that suggests that economically disadvantaged areas are more likely to experience crime than prosperous or affluent areas. … We looked at residential stability and how long people have been living in those neighborhoods, because there is research to suggest that the longer people have lived in an area, the more they are invested in an area, the more attachment they have to that place, and they may be more willing to step in or prevent crime or they have more social capacities to prevent crime from happening in the first place. … We looked at linguistic isolation, especially indo-European linguistic isolation. I am not as familiar with that body of research, but I know that immigrant areas––I don't know about the international scale––but at least in the Unites States, but there is actually less crime in places of immigrant concentration. So that is another variable that we put in. And we also included a race variable, because there is a lot of research specifically in the US, again I'm not sure about the international, about how race is related to crime. There is a whole bunch of research about racial oppression that is driving this relationship. It's not that the minorities are more criminal than the rest of the population, but there are a lot of structural and macro-level policies that unfortunately even still today are driving crime in minority areas. So we compared all these structural variables with our crime variables and we only selected the variables that had a consistent relationship with all type of crimes. … And we did not include the linguistic isolation and the residential stability, because they were going in the wrong directions sometimes for certain types of crime.

With an eye to more selective datasets, developer Georgios discusses, for example, whether it makes sense to include social media data or not:

> I know that some crime forecasting systems use social media as indicators; we have not used social media in any way and we don't plan to use it for crime forecasting. I think it's most valuable to use it for situational awareness––say a bomb goes off––to know what has happened, to get pictures; then it's super-useful. But I think it's less useful for prediction. It suffers from some problems, meaning that any time you want to analyse social media data you need a language-processing component … . I

> just don't think it makes a lot of sense to use it when we have already a lot of other data that are … less private.

Johannes heads a team that develops a software for predictive policing. Of all the interviewees, he was the most outspoken about the fact that any correlation, any pattern recognition, also needs to include theories about causation, pointing out, "A correlation is not a causality! You can always find a correlation, but when you take a close look, it is not a sensible one … I am not a friend of including just any type of data in software. … Good software builds on knowledge bases. It is based on content, not only pure statistics, mathematics, and algorithms." The type of research which is then quoted as claiming causality in datasets ties in with more complex combinations of theories and dataset imaginaries. Yet an overriding imaginary seems to persist, namely, that digital data are susceptible to mathematical form and that there is such a thing as unbiased data that can reveal meaningful patterns. As IT professional Bertrand states, "If you have … high quality unbiased data for machine learning, I wouldn't rule out that you can have a prediction algorithm that can actually outperform a skilled police officer."

Even at the stage of conceptualizing data we can already observe how they are considered crucial to processes of prediction. This overview of data imaginaries in predictive analytics thus highlights which critique becomes pertinent. While purporting to offer efficiency––a politics of form that can exclude and include notions of causality––the imaginary of malleable, unbiased datasets underlines the necessity of critically describing the theories, correlations, and causalities that are expected to sit in these datasets and that render them critically relevant to the process of prediction.

Generating data

Data do not exist per se. Rather, someone or something, with or without specific intentions, always generates data. Imagining data and generating them are intertwined processes, as data are often (but not always) produced for a specific purpose. Purpose-driven data generation is informed by ideas about what kinds of data will match a purpose best. Incidental data generation reflects imaginaries in other ways but may introduce purposefulness along with further imaginaries at later steps. Both purposeful and incidental data generation incorporate imaginaries regarding what is true, knowable, acceptable, and complete. The recent activist and scholarly trend of distinguishing between 'good data' (e.g. Mann et al. forthcoming) and 'bad data' (Galdon Clavell 2018) is indicative of reflection about the way in which data imaginaries and data generation speak to each other. In these articles, data are understood to either embrace or disrespect fundamental rights, which implies that datasets can reproduce social in/equalities from the moment of their creation. While discussions about data as 'good' or 'bad' follow specific ethical imaginaries, this article emphasizes more generally that, taken together, moments of imagining and generating data channel the further direction data analyses take, since data are considered critical to explaining or mastering a specific phenomenon.

For example, in the context of predicting crime patterns, the actual generation of data is of high relevance. Any software model that seeks to predict such patterns relies, amongst other things, on data from police reports: incidents that are recorded by the police in a specific geographic area over time. Data from police reports are highly dependent on organizational factors, such as who registers crimes, what forms are used, and where exactly the incidents occur. While variation in self-reporting by victims is already a factor known to influence available data for analyses (not least when it comes to gender difference and intimate partner violence [cf. Chan 2011]), there are many other elements that shape the actual production of data. The interview with police officer Dihyah disclosed that, in his area of responsibility, "approximately 20% of the police population are registering 80% of the information in the database. It's lots of data, but very few register very much." Thus, officers' recording activity also influences the data available for crime prediction. In addition, each officer has a different threshold for deeming an incident worthy of report and, depending on

the reporting system, there is also leeway for an officer's interpretation of the reported case, often an implicit aspect of crime data generation. Other administrative elements that affect the generation of data are the level of detail that crime-reporting forms provide and whether they are in digital format or have to be digitized. Software designer Amanda notes that in her own country, "The police department is still using paper forms", which are then digitized. The very translation from analogue into digital forms also influences the kind of data available for analysis. Thus, not only humans, but also their situations, as well as forms and programs are part of producing and shaping the data available for analysis.

When discussing specific information-organizing software for intelligence purposes, Dihyah explains that officers and software designers are well aware of the differences in data generation and related moments of observation. Almost in the spirit of Karen Barad's call for thorough description of the data-producing and recording apparatus (2007), the officers and designers decided to add a field into the software's interface in which the 'story'––that is, the circumstances of data production––is described by the recording officer. Police officer Dihyah explains, "What story are you trying to tell me? You are delivering a lot of data, but where is the story? So they were obliged to fill in a short story. … You have to put it in words. Because we can't really tell that from the data you provided." This context information would then be used to achieve a higher standard of reproducibility and reflexivity in the software.

These examples illustrate how much variation can be found in the preconceptions, routines, and standards for data generation just within the field of policing. Both, human and non-human, intentional and unintentional, reflexive and un-reflexive processes shape the datasets available for analyses of crime patterns and make data act back.

Storing data

There are no data without a database, without storage or retention (or, at least, only very short-lived data). We have seen that data cannot take shape or meaning without imaginaries of form. Neither can data gather meaning without containment, that is, without limits or borders. Such containers build upon rules of what is contained and what is not. While technological specifications and frameworks for storage come to mind, the digital ecosystem also includes the norms, values, and rules that generate decisions on criteria for inclusion and exclusion in any given database. This is the value-based and regulatory framework of data, which not only orders and structures the borders of the stored data but also the manner of storage (or storage infrastructure), its internal hierarchies, and relations between elements or points. The rules that order databases alter, as a matter of course, the relations of their data to data subjects and much more. While the logic of data storage may not fully determine data and data subjects, we can say that it co-determines the existence of data, the data subjects, processes of handling data, and even the fields that are eventually affected by predictions. The moment of creating and maintaining data storage is key to rendering data critical, especially within predictive policing: the worth of data is established by keeping them and making them available to analysis. As discussed below, the multiplicity and complexity of that moment also needs critical observation.

Within predictive policing we can see that the ways in which data infrastructures and databases are built already have a forceful impact on the data as well as the forms and patterns that data eventually reveal. Software developer Amanda indicates a crucial moment of database-generation that we tend to forget when thinking about digital data analyses, "Officers handwrite when an incident happens, they fill out the paperwork, they submit the paperwork and then it is recorded into database." Building data storage, then, is not only a part-analogue process, but also includes a critical moment of reformatting and translating information. Once digitized, Amanda says, "… you can query the data, you can select a crime incident that you like––which you wouldn't be able to do if you just had stacks of paper forms sitting on your desk. It makes analysing the data much easier and more time-efficient." Even

if data storage were at some point to omit analogue infrastructure and procedures of reformatting data, different databases would still be dependent on those who register data into a database, and the rules of storage. To add to this complexity, police officer Dihyah points out that existing rules about databases do not necessarily aid in the process of building a knowledge base. When building a knowledge base, technical and legal rules are interdependent with unstructured decision-making processes. As Dihyah explains,

> We have the law on how to store and how to delete data. We have all this data, all this information, but we don't have procedures, we don't have any systems that further help us in deciding which data to keep, which to delete. This data management is manual. Every time something is registered in the database, someone has to sit and read text. … Every bit of information has to be read and assessed. While quality indicators should be objective, they end up in fact being subjective assessments: How necessary is this? How well can you connect this data with other data, about which criminals, victims? All these assessments about how and why to keep this information are made by people.

While imaginaries of systemic objectivity are still prevalent in the idea of building data storage infrastructure, police officer Dihyah also underlines the need for horizon scans: overviews performed by professionals who then understand how they would like to develop the database further. He also acknowledges how challenging this exercise is when connecting data from different databases for that purpose. Yet, despite acknowledging these difficulties and seeing the complexity of socio-technical collaboration, the imaginary of a unified, objective knowledge base persists. Dihyah says, "We need to know what we know. We need to connect all databases so that we get one answer: this is what we know! Then we can ask [about] what we don't know. … [and] what we need to get … from those who collect information." Equally, in software developer Christian's narrative, the idea of a complete database mingles with the acknowledgement of imperfection, and it is interestingly the human data cleaning process that brings databases closer to perfection. "What's in the dataset? Is it complete? Have they given us everything? We need to first understand whether data needs to be cleaned, we need to understand quality of the data. … There are errors in all databases, you will never find the perfect database."

These examples underline how the making of containers for data storage – technically, via legal rules, and crafted by hand – is shaped by professionalized decisions and visions. These moments of data storage co-determine how and which data are rendered critical and which material data point will eventually be made into a marker of meaningful human experience or behavior.

Data thus pass through a process of imagination, generation, and storage in which each of the socio-technical moments involved co-shapes the criticality of data. Here, the acknowledgement of incompleteness, imperfection, and context meets the ideal of unbiased, complete datasets in curious ways. Initially, data are highly dependent on those who imagine them, those who create and collect them, and the infrastructure they have at hand. Yet the moment of entering them into containers–– storage platforms that follow their own rules and logics––disconnects data to a certain extent from their creators, owners, and collectors. While this disconnect will never be achieved in full, it creates new options for rendering data more malleable, supple, and impressionable.

Selecting data

As part of most scientific and engineering procedures, data selection takes place before they are subjected to further analyses. What happens here has some similarities with the moment of data generation, but at the stage of data selection differences in modelling the representative quality of data are even more pronounced. After data are generated––for

example, by officers filling forms, software capturing data traffic, or sensors receiving impulses––datasets still require engagement and are sometimes even changed as they are selected for analysis. They are 'cleaned' or translated into specific analytic categories. The assessment of data quality and the selection of data for further analyses are tied to specific understandings of the world, of the procedure's purpose, or phenomenon to be analyzed. Sometimes, these cleaning and selection processes can almost become the core of the analytic project. It may take enormous resources to develop a common standard for data selection, to define different data categories, to assign existing data to them, and discard other data. As Sabina Leonelli observes, technology-centric science projects in particular tend to argue over the correct procedures for "data selection, formatting, standardization, and classification, as well as the development of methods for retrieval, analysis, visualization, and quality control" (2016: 16). Some scholars have written manifestos advocating the importance of digital data handling in research projects (Geoff et al. 2011), since not all projects dedicate specific resources to this particular moment.

As the history of the relationship between science and data has illustrated, positions on the selection of data for analysis can vary drastically, something also found in the context of predictive policing. Some designers of predictive policing software, like Georgios, choose to run their analyses on any available data, including public databases on weather, societal events, or phases of the moon. As he observes, "Some cases seemed unusual at first … For example, the phases of the moon. Some of these variables are used for similar kinds of crime. There is no literature about why that is that case, but with full moon you may be seeing more outside." Others, like software designer Johannes and his team, include only highly select data in their analyses, which have been thoroughly examined and curated by policing experts. Unsurprisingly, each approach to data selection ties in with different ideas of data processing, as well as variation in pursued results. Georgios' approach is based on the assumption that data quantity can reveal unexpected patterns, even though explanations for such patterns may not (yet) exist, as long as large, little-curated datasets still provide the user with a 'correct' result (e.g. a crime in a specific area). Johannes' approach, on the other hand, is informed by specific criminological theories and explanatory models of crime. These include, for example, Routine Activity Approaches or Near Repeat-Modelling (based on Cohen and Felson 1979), whereby the same offender is believed to follow specific routines or geographic patterns, or theories about Situational Crime Prevention (originally Clarke 1997) that suggest crime occurs when targets are inadequately protected. These theories determine the selection of data for analysis. Furthermore, while humans curate most data selection processes, the increasing automation of data selection adds new layers to the process.

Differences in data selection approaches and the––sometimes arduous––procedures of cleaning and organizing data characterize this moment as a central part of data's becoming critical. For example, assigning data to new categories may require their reinterpretation or reorganization, which may question their status as immutable (as suggested by Latour 1987) or always intact. Data can never be scrubbed clean and often they are also difficult to assign to categories––whether because no compromise can be found amongst those who organize and engage with data, or because ambiguous data resist interpretive consensus. When data are cause for debate, it may be argued not only that humans render data critically relevant, but that data also introduce controversy or debate.

Processing data

Data processing may be the moment that is hardest to comprehend in full since its procedures are increasingly automated. The most common types of data-processing software follow specific analytic parameters and are then trained on datasets to identify patterns of interest. Within these training datasets the 'correct' patterns are known to the engineer so that algorithms and their parameters can be adjusted until the algorithm identifies all the relevant patterns. Once it passes the test of

finding the 'correct' patterns, the software is put to use on new datasets, where the correct matches are not yet known. These are so-called discriminating algorithms (cf. Smith and Buechler 1975), although not because they can impact on the right to non-discrimination by being trained on discriminatory datasets, which is also an important debate (see e.g. Benjamin 2019). Technically, discriminating refers to the algorithms' mode of operation, which is based on making distinctions. Other forms of automation are Generative Adversarial Networks (GANs, originally designed by Goodfellow et al. 2014), which create at the same time as they discriminate. GANs still identify patterns in the datasets that they are processing, but they are not trained or given information about what a 'correct' pattern would be. Rather, the algorithm identifies, interprets, expresses, and re-creates what it identifies as 'the essence' of the processed data––without the engineer intervening, determining or even knowing what this essence may be.

Despite the fact that software becomes a prominent actor in the processing moment, data still play a crucial role here. Data are part of determining what, exactly, algorithms are able to identify. Even GANs, which are often presented as independent, creative agents, cannot escape or bypass those moments in which data are imagined, generated, stored, accessed, and selected before being processed. However, during this moment of processing, data and algorithms collaborate in ways that humans cannot necessarily know. This collaborative moment of data processing is also difficult to reconstruct due to the computing powers and processing speed that machines exhibit. In the context of predictive policing software, for example, two interviewees explicate that engineers may define the parameters that they use to program the algorithm, but they cannot know exactly how algorithms combine these parameters when processing data to produce results. Police officer Hans reflected about the effect this has, observing, "I guess it's harder for people, then, to question those patterns if these parameters are not visible or accessible. You just accept the parameters." Thus, data become critical and begin to act not just when humans engage with them, but also when processed by an automated agent.

Reusing data

At its core, datafication is a problem of recycling (Thylstrup 2019): data is broken down and re-emerges as new data in new contexts. Drawing on related work on recycling, therefore, we finally draw attention to the moment of data reuse and repurposing. Once extracted and selected as suitable for processing, data are repurposed for new and different kinds of uses. Hence, waste metaphors such as 'data exhaust' and 'data traces' have played a significant role in the rise of data practices, with tech companies redefining data flows and digital traces as waste material (Mayer-Schönberger and Cukier 2013). Data analytics companies structure and reuse digital traces to turn them into valuable resources. Such data management, data integration, and data structuring can be understood as the development of data value chains; and it is not only data that are reused. Algorithms also undergo cycles of use and reuse in systems such as facial recognition, biometrics for service provision, and welfare 'decision support' tools. Neither data nor algorithms thus die in digital data ecologies; rather they are recycled: broken down to re-emerge as new matter that enfolds people, times, and places in entirely new contexts. Again, predictive policing tools are a case in point. Despite the practice that each prediction tool is trained on local and very recently produced datasets, the recycling of data is also observable in the original sense of the word: different data points are extracted and 'put together', collected from several databases. Interviewee Christian was an outspoken supporter of combining data from as many different sources as possible. Yet even those who are more selective about their data sources recycle and compose information from different databases. Police officer Dihyah explains that he sees the added value of combining police data with financial information and data from other public databases, not necessarily for predictive policing in the narrow sense, but to assess a person's risk factor:

> [The system] connects all these types of information––financial information and all the other information that we

> have in all the other databases¬¬––and then it gives each object a relevance factor based on the rules that impact each object. So, after this automatic process, person A can have a factor of 700 and B can have a factor of 400, telling us that person A could be a bigger risk factor than person B.

This example exhibits a typical effect of recycling. Not only are data originally produced for different purposes and contexts (financial administration, public administration, and police administration), they are reassembled, reused, and repurposed in order to produce new insights. The logic of risk and prevention, originally emerging from the financial and insurance sector, also begins to co-determine policing practices. However, more problematically, since the moments of imagining, generating, storing, selecting, and processing data differ in each dataset, recycling becomes a complex process, in which tracing the histories of datasets becomes a practical and an ethical challenge. The training data used by the National Institute of Standards and Technology (NIST) to develop intelligent facial recognition solutions (NIST 2019) exemplify this. Nikki Stevens, Os Keyes, and Jacqueline Wernimont (2019: online) recently found that the NIST database and training system relied heavily on images of people in vulnerable situations, such as "images of children who have been exploited for child pornography; U.S. visa applicants, especially those from Mexico; and people who have been arrested and are now deceased", as well as images "drawn from the Department of Homeland Security documentation of travelers boarding aircraft in the U.S. and individuals booked on suspicion of criminal activity" (ibid.).

As the problem of discriminatory datasets is well-known in predictive policing (Browne 2015), recycling data to solve crime problems needs critical attention. This insight is also formulated by programmer and expert Bertrand who says, "History is biased! … They arrest Blacks and all the historical data say, 'Well, we have all these wonderful arrests of Blacks possessing dope' … And the algorithm basically says, 'Sure, it's ok, it's not racist, you can go on [ironically] because algorithms are absolutely apolitical and you can just go on harassing Blacks.'" In his statement, Bertrand denounces procedures of correlating any available data, particularly with police data, that is, software models that heavily cultivate data reuse.

Yet the prediction procedures based on curated datasets also feed the precarious practices of recycling. The more opaque the relations between data subjects, owners, and creators––be it through data storage design, processes of cleaning, or trading datasets––the more difficult it becomes to 'follow the data' along its value chains. A classic claim made by those choosing to reuse data is that their datasets are merely "operational" (Grother et al 2019: 18). However, we wish to foreground the point that the data wrought by these datasets remain "sticky" (Ahmed 2004: 90): they cannot be wrested from their agency, sanitized, and presented as new data with no social stains or remains. Rather, they inevitably display the effect of their histories of contact between bodies, objects, and signs. They leave residues, carrying and spreading material, social, and ethical entanglements with critical infrastructures. At worst, such recycling processes can result in the creation of prediction technologies that distribute vulnerability unevenly through sticky associations while simultaneously invisibilizing these ties. Indeed, contemporary efforts to problematize data trajectories also show how data transactions develop haunted data (Blackman 2019). In these cases, data often end up reproducing violence, whether racist, misogynist, or classist. Acknowledging the critical moment of data reuse raises significant questions, then, about the ways in which data are extracted by "documenting humans' bodies and selves", while also making them "open to constant repurposing by a range of actors and agencies, often in ways in which the original generators of these data have little or no knowledge" (Lupton 2015: 563). This entanglement affects not only the opportunities of those whose lives remain as residue in data piles, but also everyone else whose data becomes enfolded into these moments. It matters what data are added to a dataset, under what conditions and according to which parameters. The critical moment of

data recycling thus warrants pervasive scholarly engagement with the reality and ethics of reuse that counters the imaginary of 'raw' data, and instead examines the sticky trajectories of dataset ecologies (Keyes, Stevens, and Wernimont 2019; Benjamin 2019; Kaufmann et al. 2019a).

Conclusion

The 'data moment' is not a single moment in time, nor is it a notion descriptive of a 'digital era'. Instead, we have described a recursive, not necessarily linear set of encounters that help us in navigating criticality within today's data ecosystems. Every time data are extracted, selected, stored, processed, and/or recycled a new series of relations and realities is established. This reveals the criticality of data and the need to study data critically. Data criticality draws our attention to the moments when humans and machines choose when, where, and how data will exist and what their agencies will be. The concept responds to Barad's call for describing the circumstances under which data is produced (2007) at the same time as it builds on the observation that data have become our companion species, one that exhibits "liveliness" (Ruppert et al. 2013: 29).

As we have shown in relation to predictive policing, recognizing data as critical to a specific context allows us to see the socio-technical processes of data ecologies. A complex assemblage of agencies, software, forms, regulations, and norms comes together in constantly shifting ways to create data and breathe new life into old data. This generative, creative process can take on animate characteristics. Data is neither sentient nor will-based but, nevertheless, it has agency, conditioning, structuring, and applying pressure on a range of analytic processes. In other words, data criticality reveals that data cannot be rendered exclusively as data. Rather, data are characterized by a radical relationality (Fraser et al 2005: 3), ceaselessly circulating in processes of emerging, breaking down, and reconfiguring. Data are, thus, neither immutable (Latour 1987) nor inanimate. Rather, they are constantly changing and always contingent on the system as a whole. There is agency in our companion species when it interacts with humans and non-humans, when it engages, and is engaged with, in different moments of meaning making. This interaction invites careful, critical observation. Only through critique can we be part of shaping the way our companion species becomes critically relevant in today's society.

References

Ahmed S. (2004) The Cultural Politics of Emotion. Edinburgh: Edinburgh University Press.

Austin J, Bellanova R and Kaufmann M (2019) Doing and Mediating Critique: An invitation to practice companionship. Security Dialogue 50(1): 3-19.

Barad K (2007) Meeting the Universe Halfway. Quantum Physics and the Entanglement of Matter and Meaning. Durham/London: Duke University Press.

Bellanova R (2016) Digital, politics, and algorithms. Governing digital data through the lens of data protection. European Journal of Social Theory 20(3): 329-347.

Benjamin R (2019) Race After Technology. Cambridge: Polity.

Bennett Moses L and Chan J (2018) Algorithmic prediction in policing: assumptions, evaluation, and accountability, Policing and Society 28(7): 806-822.

Bevan A (2015) The data deluge. Antiquity 89 (348): 1473-1484.

Blackman L (2019) Haunted Data: Affect, Transmedia, Weird Science. Bloomsbury Academic.

Browne S (2015) Dark Matters: On the Surveillance of Blackness. Durham: Duke University Press.

Chadwick A (2006) Internet Politics: States, Citizens, and New Communication Technologies. New York: Oxford University Press.

Chan KL (2011) Gender differences in self-reports of intimate partner violence: A review. Aggression and Violent Behavior 16(2): 167-175.

Clarke RV (1997) Situational Crime Prevention. 2nd edition. New York: Harrow and Heston.

Cohen LE and Felson M (1979) Social Change and Crime Rate Trends: A Routine Activity Approach. American Sociological Review 44 (4): 588-608.

Collins HM (1981) Stages in the empirical programme of relativism. Social Studies of Science 11(1): 3–10.

Davenport TH (2014) Taming the 'Data Plumbing' Problem. Wall Street Journal, May 21. https://blogs.wsj.com/cio/2014/05/21/taming-the-data-plumbing-problem/. (Accessed 29 April 2019).

Degeling M and Berendt B (2018) What is wrong about Robocops as consultants? A technology-centric critique of predictive policing. AI & Society 33: 347–356

Felski R (2012) Critique and the hermeneutics of suspicion. M/C Journal 15(1). Available at: http://journal.media-culture.org.au/index.php/mcjournal/article/viewArticle/431 (Accessed 25 March 2019).

Fraser, M, Kember S and Lury C (2005) Inventive Life: Approaches to the New Vitalism. Theory, Culture & Society 22(1): 1–14.

Galdon Clavell G (2018) Bad Data Challenge. Dataethics. Available at: https://dataethics.eu/bad-data-challenge/ (Accessed 26 March 2019)

Geoff SA, Vaughn M, McKay S, Lyons E, Stapleton AE, Gessler D and Matasci N et al. (2011) The iPlant Collaborative: Cyberinfrastructure for plant Biology. Frontiers in Plant Science 2: 34.

Goodfellow IJ, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A and Bengio Y (2014) Generative Adversarial Nets. Proceedings of the 27th International Conference on Neural Information Processing Systems, Montreal, Canada, 2: 2672-2680.

Gregg M (2015) Inside the Data Spectacle. Television and New Media. https://nls.ldls.org.uk/welcome.html?ark:/81055/vdc_100058579298.0x000012. (Accessed April 26 2019).

Grother P, Ngan M, Hanaoka K, Information Access Division and

Information Technology Laboratory (2019) Ongoing Face Recognition Vendor Test (FRVT): Part 1: Verification, April 4. Available at: https://www.nist.gov/sites/default/files/documents/2019/04/15/frvt_report_2019_04_12.pdf (Accessed April 26 2019)

Hamilton JM and Neimanis A (2018) Composting Feminisms and Environmental Humanities. Environmental Humanities 10(2): 501-527.

Haraway DJ (2008) When Species Meet. Minneapolis: University of Minnesota Press.

Haraway DJ (2016). Staying with the trouble: making kin in the Chthulucene. Durham: Duke University Press.

Harding S (1991) Whose Science? Whose Knowledge? Thinking from Women's Lives. Ithaca, NY: Cornell University Press.

Hartsock N (1983) "The Feminist Standpoint: Developing the Ground for a Specifically Feminist Historical Materialism." In: Harding S and Hintikka MBP (eds) Discovering Reality. Dordrecht: Reidel. 283-310.

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.

Kaufmann M, Egbert S and Leese M (2019) Predictive Policing and the Politics of Patterns. British Journal of Criminology 59(3): 674–692.

Keyes O, Steven N and Wernimont J (2019) The Government Is Using the Most Vulnerable People to Test Facial Recognition Software. Slate Magazine. https://slate.com/technology/2019/03/facial-recogni-tion-nist-verification-testing-data-sets-children-immigrants-consent. html (Accessed 29 April 2019).

Latour B (1987) Science in Action: How to Follow Scientists and Engineers Through Society. Cambridge: Harvard University Press.

Leonelli S (2016) Data-Centric Biology. A Philosophical Study. Chicago: The University of Chicago Press.

Lupton D (2015) Lively Data, Social Fitness and Biovalue: The Intersections of Health Self-Tracking and Social Media. SSRN Electronic Journal. Available at: http://dx.doi.org/10.2139/ssrn.2666324

Lupton D (2016) Digital companion species and eating data: implications for theorising digital data-human assemblages. Big Data and Society 3(1): 1–5.

Mann M, Devitt SK and Daly A (forthcoming) What Is (in) Good Data?. Good Data. Amsterdam: Institute of Network Cultures Theory on Demand Series. Available at: https://ssrn.com/abstract=3297103

Marx K (no date, oldest published edition 1906) Capital: A Critique of Political Economy. In: Engels, F (ed) New York: Modern Library.

Merton RK (1942) Science and technology in a democratic order. Journal of Legal and Political Sociology I: 115-26.

NIST (2019) FRVT 1:1 Verification, March 20. Available at: https://www.nist.gov/programs-projects/frvt-11-verification (Accessed April 29 2019)

Pearsall B (2010) Predictive policing: The future of law enforcement. National Institute of Justice Journal 266(1): 16–19.

Ratcliffe JH (2004) The hotspot matrix: A framework for the spatio-temporal targeting of crime reduction. Police Practice and Research 5(1): 5–23.

Rubinstein IS (2013) Big Data: The End of Privacy or a New Beginning? International Data Privacy Law 3(2): 74-87.

Ruppert E (2018) Sociotechnical imaginaries of different data futures: an experiment in citizen data. Rotterdam: Erasmus University Rotterdam.

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

Smith P and Buechler G (1975) A branching algorithm for discriminating and tracking multiple objects. IEEE Transactions on Automatic Control 20(1): 101-104.

Striphas T (2015) Algorithmic Culture. European Journal of Cultural Studies 18(4-5): 395-412.

Thylstrup NB (2019) Data Out of Place: Toxic Traces and the Politics of Recycling. Big Data & Society 6, no. 2: 205395171987547.

Zuboff S (2019) The Age of Surveillance Capitalism. The Fight for a Human Future at the New Frontier of Power. London: Profile Books Ltd.

Author Bios

Mareile Kaufmann works at the Department of Criminology and Sociology of Law, University of Oslo and the Peace Research Institute Oslo. Her research interests lie in the sociology of technology, security research and digital criminology. A large part of her work focuses on surveillance practices, but also on how people engage with these from within surveillance systems.

Nanna Bonde Thylstrup is associate professor at the Department of Management, Society and Communication at Copenhagen Business School. Her research interests broadly revolve around the politics, epistemologies and sustainabilities of digital media. She has published extensively in the areas of media theory, cultural memory, infrastructure and datafication including MIT Press, New Media & Society, Surveillance & Society, and Big Data & Society. At the moment, Nanna is working on issues related to data reuse and data sustainablility.

J. Peter Burgess is professor of Philosophy and directs the Chair in Geopolitics of Risk at Ecole normale supérieure, Paris. He is also Professor in the Research Group on Law, Science, Technology and Society at the Vrije Universiteit Brussel. His research concerns the meeting place between culture, politics and technology, with emphasis on questions of risk and uncertainty

Ann Rudinow Sætnan is professor emerita in Sociology at the Norwegian University of Science and Technology. Her research has ranged from working conditions in health care organizations via gender, science, and technology studies, to surveillance studies. Currently, with Rocco Bellanova, she is exploring surveillance issues through comparisons of forensic surveillance with bird-watching.