Information in-formation: Algorithmic policing and the life of data

Abstract: Many processes in law enforcement increasingly rely on algorithmic processing of digital data. Whereas most recent critical scholarship focuses on the algorithm as the decisive factor in the production of knowledge and decisions, we foreground the data that feed algorithms. Based on insights gained from empirical studies on predictive policing software, we forward a theorization of "information in-formation". In bringing together new materialist thinking and approaches about the liveliness of data, we conceptualize data as matter that is in-formation due to both human and non-human capacities. We develop the analytic notion of the life cycle of data to better understand the liveliness and agency of data in any type of data-based environment, and illustrate the life cycle of data in the case of predictive policing. We show how data come into being, how they are selected and cleaned, how they multiply in distinct database systems and according to different logics of speed and urgency, and how they inform the workings of algorithms.

Key words: prediction, predictive policing, digital data, life cycle, life, software, algorithms, information, form

1. Introduction

Many processes in law enforcement increasingly rely on algorithmic processing of digital data: software packages are used to detect crime patterns (see PRECOBS, Predpol, ShotSpotter), court-room algorithms assist in the prediction of recidivism (see Equivant) and chatbot "Sweetie" exercises smart, covert surveillance to create digital evidence (TDH Netherlands 2014). Whereas most recent critical scholarship focuses on the algorithm as the decisive factor in the production of knowledge and decisions (Amoore 2009, Leese 2014, Saugmann 2017, Aradau and Blanke 2018, Kaufmann 2019, Krasmann 2020), we foreground the data that feed algorithms.

Even if the word 'data' tends to be translated as 'given' (see Hoad 1996), we argue that data are in a constant relationship of (ex)change with their environment. Hence, they exhibit a form of "liveliness" (Lupton 2016b), which also structures our knowledge about the world. Data are always an active part of and are shaped by production processes that select and prioritize (Gitelman 2013, Bechmann and Bowker 2019). For example, discussions about so-called 'data quality' are a signifier of such interaction processes within predictive policing due to variations in crime reporting, differing theories about crime predictors or varying classifications and categorizations of incidents. By following data through different stages of being in-formation, we offer a comprehensive angle of the relationship between data, algorithmic prediction and productive bias (Hildebrandt 2016), that is the idea that any interaction with data produces consequences in social life.

Based on insights gained from empirical studies on prediction software, we forward a theorization of "information in-formation" (cf. Dillon 2000). In bringing together new materialist thinking (Barad 2007, Mol 2008) and approaches about the liveliness of data (Savage 2013, Ruppert et al. 2013, Lupton 2016b), we conceptualize data as matter that is in-formation due to both human and non-human capacities. In other words, data have a life, which means that they take a role in the production of law enforcement. While this argument appears to indicate a trend towards post-human policing, we emphasize that humans are always part of the process – even if their role is changing. Datasets stand in an active relationship with the humans who generate information, who clean and prepare datasets, who write programs to analyse data and interpret results, and whose actions eventually become informed by these interpretations. During these many stages, however, data assume liveliness as their affordances can shape analytic projects and human life in unforeseen ways.

In what follows, we will first present the analytic notion of the life cycle of data (Kaufmann 2020) to better understand how humans imbue data with life, but also how data assumes liveliness and agency in any type of data-based environment. After that, we illustrate the life cycle of data in the case of predictive policing. Indeed, predictive policing and the life cycle of crime data varies across countries, which is why we need to take care to explain the specificities of predictive policing processes. The empirical data used for this paper come from research that was carried out between 2016 and 2019 with 11 police departments in Germany and Switzerland. The departments used different types of predictive policing software for the identification and subsequent intervention into domestic burglary series. Data were primarily conducted in the form of qualitative expert interviews and ethnographic observations of predictive policing practices. We use material from these empirical studies to illustrate how data come into being and how they are selected and cleaned. We include the role of storage in databases, the workings of algorithms on datasets and the reuse of data to show how final results come about and influence policing decisions.

2. Data and life – towards an analysis of data life cycles

Digital data are central to the way in which predictions take form. They inhabit this powerful position not only because data are the basis for any type of observation or analysis. We develop the argument that data are powerful because they are productive and performative. In fact, we argue against their very name – "a thing given or granted" (Hoad 1996: 224) as expressed in the singular Latin form *datum* – that data are dynamic. They are matter *in becoming* because they stand in constant communication with other data, with humans and infrastructures. If data are to exist in plural, this plurality of data – the fact that different pieces of information relate to each other – also means that they influence and work with each other and their environments.

Since data, infrastructures and humans stand in a mutual relationship of becoming with each other (Haraway 2015a; Latour 2013) we use the lens of *life* to better understand how information is in-formation and generates form. There is a growing interest in theorizing the relationship between digital data and life and we will draw attention to those authors who have already developed such theorizations.

Due to the steady rise of digital infrastructures for data collection, we need to think of digital methods anew. Karen Barad (2007) underlined the need to describe carefully how data come into being and how they are interacted with. Remarkably, Mike Savage (2013) argues that the proliferation of "lively data" engendered through information and communication technologies energizes the study of method in a way that makes him speak about the "social life of methods" (ibid: 3). Digital methods, he observes, are no longer subject to the "straitjacket imposed by positivist statistical procedures" (ibid: 6f). They can, thus, no longer "be treated as a 'given', but are themselves subject to multiple mobilization" (ibid: 10). Digital methods, too, begin to live new lives due to the liveliness of digital data and – we further argue – these mobile methods are part of shaping the liveliness of data in return. This lively interplay of humans, digital data and digital method will become apparent in our case studies of predictive policing. In order to further understand what lively data are and to substantiate the relationship between humans, digital data and life, we will first shed light on further key theoretical arguments.

The idea that life has a fundamental impact on ontology, observation, epistemology, and politics is collected in positions that favour becoming over being, flow and flux over structure, and emergent form over mechanisms (Lash 2006). Friedrich Nietzsche's materialism, Gilles Deleuze's focus on multiplicity (Deleuze and Guattari 1987) as well as Michel Foucault's (2008) biopolitics are considered fundamental to thinking information in relation to life. Later contributions, such as Nikolas Rose's "The Politics of Life itself" (2006) and Eugene Thacker's (2005) "The Global Genome" focused more concretely on the way in which digital information remakes society, social life and sociality. This train of thought is also promoted in Manuel Castells' seminal notion of "The Network Society" (1996).

Evelyn Ruppert, John Law and Mike Savage (2013), however, criticize the global and epochalist nature of such claims about information reconfiguring social life. While they acknowledge the relevance of the abovementioned works on digital information and life, they bring back Foucault's *devices* in order to underline the specificities of how data make new and different forms of life. They invite readers to pay attention to how digital devices are varied and composed within different environments. Ruppert et al. (2013) further ask us to pay attention to how digital devices are enrolled in knowledge creation by bringing data to life. In doing so, they contribute to a more thorough understanding of the relationship between digital devices and life by drawing our focus onto specificities. They do, however, tend to postulate the liveliness of data without further defining what that means. Much in line with Savage's "social life of methods"[1] the focus lies on the devices that shape data and make it lively. From this perspective, data "are the material of social lives" and "form part of many of the apparatuses for knowing those lives" (Ruppert et al. 2013: 24). Yet, Ruppert et al. do not focus on how

[1] The social life of methods is the theme of the special issue that both essays are part of.

exactly digital data form devices. While they offer a method- and device-oriented answer to our aim of theorizing the relationship between data and life, we can still ask: in what other ways are data lively?

Deborah Lupton takes us one step further with her study of social fitness (2015) and the quantified self (2016a). She observes how data are "lively", because they are generated from life by documenting human bodies. Digital data are continually generated and fundamentally about the life of humans. In return, data generate new insights into people's bodies, behaviours and sense of self (2016a: 5). Further, as participants in the digital economy, digital data have a vitality of their own. Data circulate, they are fluid and open to constant repurposing by a range of actors and agencies. Through that, data can become quite independent of those who originally generated them (ibid: Lupton 2015). This happens, for example,

> when users share their personal data with one another or attempt to synchronize the data across other devices or platforms, or when many users' data are aggregated and rendered into large data sets, which may in turn be deployed for a range of purposes by other human actors. (Lupton 2016a: 41)

As a commodity, data influence social relationships, life chances, livelihoods and everyday life. They are lively due to the advent of algorithmic authority, which once again underlines the importance of the devices that make data lively and that equally have effects on conducts of life and life opportunity (Lupton 2015: 563). While these writings mainly outline how data come of and influence life, Lupton (2016b) pushes that relationship even further when she draws on Donna Haraway's (2003) trope of "companion species" to suggest that data can be our digital companion species. To her, Haraway's (2015b: n.p.) comment on human-machine relationality is central:

> The kind of sociality that joins humans and machines is a sociality that constitutes both, so if there is some kind of liveliness going on here it is both human and non-human. Who humans are ontologically is constituted out of that relationality.

The liveliness of data and our own vitality, Lupton (2016b) argues, stand in a mutual dependency that needs to be recognized. The interdependency of digital data with humans is so profound that it becomes difficult to disentangle agencies. Lupton employs Annemarie Mol's (2008) essay "I eat an apple" that describes the entanglement of food and body via digestion, to argue that it is equally difficult to understand at what point digital data become part of us and we become part of digital data. Data can act in human life and human life can act in data assemblages.

With our conceptualization of the data life cycle, then, we seek to answer some of Deborah Lupton's questions, namely: "How do we live with our digital data companion species? How do they live with us?" (2016b: 4). If we seek to understand this mutual relationship in more depth we do not only study how digital devices form data about life, and how digital data form apparatuses that know life in return (e.g. Savage 2013). We also focus on the specificities of how data come from life, how they become lively, how they exhibit life-like qualities and how they shape life (Haraway 2003; Ruppert et al. 2013; Lupton 2015, 2016a, 2016b). If we seek to grasp this mutual relationship between digital data and life in a consequent manner, we also need to include Donna Haraway's (2003) and Katherine Hayles' (1999) observations that basic principles of information do appear in life. Or as Scott Lash (2006) puts it: that the study of life is informational as can be seen, for example, in genetic instructions and coding, where information literally organizes life.

The above literature has given us good foundations to acknowledge that digital data shape life and life itself can be understood in terms of informational principles. We combine these strands and provide a detailed account of the liveliness of data – the discussion that Ruppert et al. (2013) and Lupton (2016b) already started. We thus follow an understanding of digital data that acknowledges their materiality and life-changing influence, but also their dynamic, life-like trajectories. The latter may be captured with Michael Dillon's (2000) ontology of information as being-in-formation[2]. We see data as matter that is in-formation, in relation (Haraway 2003), in a state of becoming. Because data stand in a relationship with human life - because they are imbued with life and have a life of their own - they do take an active role in socio-technical processes. In this vein, we find the analytical notion of the lifecycle helpful and forward the idea that digital data have a lifecycle. While Veerle van den Eynden (2014) has introduced the formal trajectory of the research data life cycle, she used the life cycle as a structuring element within the research process without tying it to the liveliness and agency of data as such.

When seeking to understand the mutual relationship between life and data, what more can the life cycle give us? The life cycle helps us bringing data and their dynamics into focus. It allows us to think of them as a digital companion species, as well as to always think their productive potential in and through relations. The lifecycle may appear linear, but we want to draw attention to its circularity. It emphasizes the ecological complexities of ingestion, digestion and emission (Mol 2008). The lifecycle captures the overlaps of continuous circulation, of coming to life, being-in-formation, of being deleted and re-used (cf. Kaufmann 2020). We use the lifecycle as that which can also illustrate the many overlaps with other lives and agencies – human and non-human. Lifecycles interact, intersect and overlap. They influence each other. From that perspective, the lifecycle can give us both *metaphoric*, but also *analytic* and *theoretical* entry points to think through the relationships between data and life, how data are brought to life, live their own life and influence human life. Most importantly, when we know these entry points, we know better how to try to shape the life with digital data that we want to achieve.

3. The rise of digital data in the context of policing

[2] This concept is forwarded in Dillon's account of poststructuralism and complexity. He sees their coincidence as a shift away from thinking with isolated components that stand in a relationship of exchange towards thinking through relationality.

As Barad (2007), but also Ruppert et al. (2013) remind us, we need to carefully analyse the specificities of the relationships between data and life. This is why we suggest using data life cycles to describe the agency of data in their respective contexts, which is in our case predictive policing. Predictive policing takes multiple empirical forms that speak to different types of crime, different models, different algorithms, and not least different data sources – all of which vary according to location and domain (Kaufmann et al. 2019). In our case, predictive policing software was used to identify ongoing domestic burglary series, marked by spatio-temporal patterns of professional offender behaviour (Egbert and Leese 2020). In principle, predictive policing aspires to make crime control more efficient. More specifically, predictive policing software promises to accelerate crime analysis through algorithmic processing of data to such an extent that insights about ongoing crime trends become available without delay and enable the police to intervene in criminal activities as they unfold (Perry et al. 2013). The police departments that we studied would, based on the knowledge produced by predictive policing software, dispatch patrols specifically in those areas with the highest likelihood for further burglary incidents. Predictive policing, then, must predominantly be understood as an attempt to enhance situational awareness and assist decision-making for operational measures such as patrolling or prevention measures. Computing recommendations of where and when to dispatch police forces are the central means to increase the likelihood of deterrence or arrests.

While predictive policing asserts that police work can be rendered more efficient and effective through digital data and algorithmic processing, it also re-problematizes the data that are processed by the system and that structure police knowledge about the world. Important questions include what kind of data should go into the system and how adequate "data quality" could be ensured in a domain where knowledge is by definition limited, volatile, and uncertain (Maltz 1999; Cope 2008). Scholars and activists have pointed out how a questionable basis for knowledge production could backfire as police tactics become biased through data and extend such bias by over- or under-policing of populations, which again feed back into datasets for future policing (Bennett Moses and Chan 2018; Kaufmann et al. 2019). Such normative assessments of data and their repercussions are important and timely.

We are, however, less concerned with the problem of "data quality" as such, but with the ways in which data and concerns about their quality become problematized by the police and software engineers. Further, we are interested in how such concerns shape the ways in which data come into being and change form as they make their way from a crime scene to crime analysis. We will, throughout the following section, empirically trace how the police create data from crime scenes, how these data inform managerial processes as much as criminal investigations, and how they are subjected to amendments and consolidation before they serve as the basis for algorithmic crime analyses.

When tracing the life cycles of crime data in predictive policing, notions of speed and responsiveness become prevalent. Crime data, so the rationale goes, quickly lose their analytical value if the aim of the analysis is to produce insights into ongoing developments. The swift availability of data is thus considered paramount for predictive policing. As one respondent put it: “We have little time to react: if there’s a burglary today, then we’ll have to be in that area tomorrow. And not in a week, because [that burglary series] will be over by then.” (Interview 07) Although crime data are not produced exclusively for predictive policing, the implementation of crime analysis software puts considerable time pressure on the life cycle of crime data. Accordingly, police departments aim to speed up the generation of crime data from the crime scene, their transfer to the central database, their consolidation, and their readiness for analysis. Digital devices and data are here not only underpinned by notions of efficiency and acceleration in policing, but they also have a performative effect as they simultaneously foster that same rationality of speed and responsiveness in crime control on the street level.

4. Formations of information: Data life cycles in predictive policing

While data life cycles constantly intersect and are not linear, the life cycles of crime data we have studied generally begin with the generation of data from a crime scene. Even though some crimes are detected directly by the police and other officials and there are regular estimates of unreported cases, many crimes are detected and reported to the police by victims or witnesses. In practice, this means that citizens call the police, where their call automatically triggers the creation of a new file in a process management database. From this point onward this file serves as the primary means for the police to handle the issue internally. The process management database contains unique process identification numbers as well as meta-data: who called and when, the type of reported crime, and the address of the crime scene.

The call centre operator then directs a patrol car to the crime scene and officers proceed to generate data from the crime scene. These data usually include basic characteristics such as the type of crime, the date and time of occurrence, known victims or stolen/damaged goods, the method of the crime (e.g., the use of weapons or tools, the point and mode of entry), and the current disposition (i.e. was an arrest made or is the offender at large?) (Santos 2013: 69). Information about these issues is derived from forensics, as well as from the statements of victims and witnesses. It is important to highlight here that data about crime do not exist independent of human action. They are actively brought into being, and how exactly they come into being is shaped by the domain specifics of their later use: police officers will capture what they consider relevant for the task at hand (i.e. registering and solving crimes), and everything that does not pertain to this task will be discarded.

At the first stage of their life cycle, crime data are usually considered fragmented and little reliable. Only at a later point, after their consolidation, they are considered to provide a more comprehensive account of what happened. A comprehensive account is, however, not what is needed at this early stage of the data life cycle. One of our interviewees illustrated the idea to create initial data that provide a brief overview and serve as a starting point for further action:

> The idea of a process management database is to capture what you see: you go to the crime scene, you speak to the victim, and you get an idea of the crime. Was it burglary, was it

> larceny, or was it larceny in combination with trespassing? The location and time of the crime, three lines of text description. Whatever you know - or think you know - at this point. It needs to be quick, and the data are sent off immediately. (Interview 50)

"Sent off immediately", as the interviewee framed it, means that the data from the crime scene, usually captured in digital form already, are transmitted to the central database where they are combined with the meta-data that were already generated at the emergency call centre. It is, however, worth looking more closely at exactly how crimes are captured and turned into data. A key task for police officers, as Harper (1991: 294) frames it, is the translation of empirical phenomena into bureaucratically administrable, analysable, and comparable bits of information. Most police departments use standardized reporting forms that need to be filled out by officers in the field, and these forms employ more or less fine-grained classification systems that specify a list of pre-defined categories for each variable that is captured. Translating empirical phenomena into bureaucratic categories means that police officers have to try to fit whatever phenomena they encounter at the crime scene into this classification system. The more fine-grained classification systems are, the more difficult this operation can turn out to be. In order to ensure proper representation of empirical phenomena within the produced data, officers need to carefully draw boundaries between different phenomena and assign them to the 'right' categories. Just like the phenomena they are supposed to represent, data (and their classification systems) exhibit a liveliness that might even be characterized as recalcitrance. One interviewee illustrated this as follows:

> We have a very complicated and exhaustive system of categories that was originally supposed to facilitate the generation of crime statistics. Proper crime statistics should be detailed, and you can't do that if you only use 'residential burglary' as a category. So we have a category for 'residential burglary', one for 'armed residential burglary', one for 'organized residential burglary', one for 'grand larceny within a dwelling', and so on. In the end we have 50 different categories that have something to do with residential burglary, and it's quite the art to find the right one. A grand larceny from a dwelling is probably a residential burglary. A grand larceny that took place within a dwelling is probably also a residential burglary, but maybe the patrol officer used the wrong key, because he couldn't think of 436000 and he didn't have the time to do his research. So he used 400000, which is just 'grand larceny'. But the report also says that the site of the crime was an apartment. Now I'll have to look into the free text description: maybe it was the brother who smashed the piggy bank and took the money. Technically that's grand larceny. The money was secured, and the incident took place within the apartment. Hence, it's grand larceny within a dwelling. Maybe the door was kicked in, then all of a sudden it's residential burglary. (Interview 50)

This quote demonstrates not only the challenges of translating crime into data, but it also hints at the repercussions that epistemic classification practices (cf. Bechmann and Bowker 2019) have for predictive policing. While deciding whether a crime is classified as burglary or larceny may appear trivial at first, the decision is not trivial for generating what is considered the "right" crime prediction. Since the software we studied was geared towards the identification of patterns in domestic burglary activities, a crime classified as larceny would not end up in the data analysed by the system. This would in turn influence the produced knowledge. Another interviewee further illustrated the problem of uncertain classification: "How did someone get access to an apartment? Mostly it's levering a window. And even if it's not levering, the category for levering is the one that our colleagues know by heart. So to be honest, it's often standard categories that end up in the report." (Interview 80) There might be various reasons for this, from time-pressure and late-night shifts to simple forgetfulness (Interviews 07, 51, 77).

We can see that the generation of crime data – the bringing of crime data into life - is dependent on many different entities, such as police officers, victims or witnesses (and their willingness to report), as well as police forms and systems, as well as those who translate reported cases into administrable data. There is variation in the ways in which data are brought to life and made to live through different categorizations. This variety in classification, however, implies that crime data must not only be considered uncertain in ontological terms, i.e. concerning the phenomenon itself, but also in epistemic terms. Data come into being through human actions, but once "alive" they do not remain static, but rather enter ongoing, transformative relations with other data, infrastructures, and humans.

As we have argued above, predictive policing imposes particular time pressure on the production and analytical processing of data. The main data input for the software that we studied consisted of data about burglary cases and their characteristics. From the temporal and spatial distribution of these cases, combined with insights about modus operandi and haul, the software would compute short-term (3-7 days) risk areas that the police would then patrol more intensely in order to prevent follow-up crime. Hence, the crime data that would be considered relevant to predictive policing would need to contain information about all these aspects. We have seen, however, that at the point in time when crime data are first generated, knowledge about the criminal incident tends to be preliminary and fragmented. The time of the committed crime might, for example, not be known because the victim of a burglary only got back from a vacation. From the mess that the offender left in the apartment, it might not be immediately clear what exactly was stolen. In addition, the modus operandi of a crime might not easily be defined without results from the laboratory.

Moving on in their life cycle, crime data are thus consolidated before they can be used as input for the production of crime predictions. For such data 'cleaning' processes, police departments usually have several layers of 'quality control' in place. Initially, supervisors and central information management units double-check for inconsistencies, such as missing values or other obvious mistakes (Interviews 26, 50, 51). These types of quality control mainly address formal issues. Crime data can, however, only be solidified more substantially via further investigations: information that might not have been available in the beginning might become amended by additional witness statements, as victims sort out the damage that was caused, or as forensic evidence comes back from the laboratory (Interviews 07, 18, 77). Although data have been brought into existence already, their epistemic properties are not yet fixed. Further interaction with human capacities is considered necessary to ensure their validity. The information provided by the data quite literally remains in-formation, as it changes form and content through interactions between police officers and databases.

At this point, the intersection of different data life cycles becomes prevalent. For investigations that can "solidify" crime data, police departments usually draw upon a dedicated *case file management database*. In opposition to *process management databases* that are mainly geared towards administrative purposes, case file management databases are geared towards in-depth knowledge production about cases. The data that they contain have been cross-checked via different investigative processes and are generally considered less volatile. These data are, however, not immediately available. As criminal investigations might take some time to produce tangible results, a case file management system is likely to lag behind a process management system in terms of quickness and readiness for data processing. The temporal difference between the databases effectively splits the representation of the empirical phenomenon of crime across disparate structures that each come with their own characteristics, advantages, and drawbacks.

Police departments are therefore faced with a practical trade-off when it comes to predictive policing: should they use promptly available yet more uncertain crime data as input for crime analysis, or should they wait and work with the more reliable, yet potentially outdated data version of the crime? This question illustrates well how different datasets about the same phenomenon come to live and have the chance to actively influence policing practices. For our respondents, the answer to this question was a matter of tinkering with what could be accommodated in practice rather than it was a deliberate epistemic choice (Interviews 09, 26, 45, 51, 78). As one analyst framed it:

> That was the main challenge: to really find out if it works with weak data, with [process management] data, which are the data that the emergency call centre records, and with data that our patrol units collect at the crime scene with their tablets. (Interview 07)

At this point, we have moved to yet a different phase of crime data life cycles: data are now analysed by the predictive policing algorithm in order to inform policing practices. As the above quote illustrates, the police differentiate here between "weak" data and "strong" data, depending on the level of consolidation and reliability that they ascribe to different representations of the same empirical phenomenon. Against the backdrop of the need for speed in predictive policing, they use this distinction to ensure that they do not end up with 'weak' analytical results that would undercut the possibilities of intervention into ongoing crime series. Data that have passed through different – fast and slow – life cycles are thus seen as more or less fit for the production of knowledge. This illustrates the argument that data are lively and malleable: they undergo multiple transformations and repurposing processes throughout their respective life cycles and have different capacities to "act back". Due to the specific, rapid temporality of predictive policing, data that were never meant to be analysed in the first place become mobilized for analytical purposes nonetheless. With the rise of algorithmic policing tools new and different data are filled with life and given an additional purpose that fundamentally changes their character and importance. One interviewee described the character of this operation:

> [The process management database] is always a bit blurry. […] It is supposed to administer internal information: who is responsible for a case, when did they distribute what kind of information to whom? The fact that we use that for analytical purposes is more like a by-product. (Interview 46)

This statement illustrates two aspects of data life cycles very well: first, it underlines the liveliness of data, because data that were originally considered uncertain and incomplete made their way into data analyses anyway. Second, this statement describes an instance of data reuse – a form of data repurposing, which triggers yet another data life cycle. In this case, data reuse creates a new role for human analysts. While originally, one of the rationales of predictive policing software was to relieve the human analyst of doing the 'heavy-lifting' via the automation of tasks, the inclusion of 'weak' data again creates a new need for the close supervision of algorithmic analysis. Here, human analysts are responsible for a retro-active curation of the data that have already been analysed in order to prevent the transformation of weak data into "irrelevant" knowledge (Interviews 01, 03, 77, 76, 79).

Crime data are at this stage of their life cycle productive of a reformulation of the relationship between human and machine, as the human task is now to investigate analytical results through a re-problematization of the data input. Predictive policing software, as one respondent put it, "would process the data right away. And that's why you need an operator who is able to dig a little deeper if necessary, and is able to evaluate a crime on the basis of his criminological knowledge." (Interview 77) The assumption is here that crime data could have been updated or complemented with additional or contextual knowledge that is, however, not represented in the rudimentary process management database, or the "fast" version of the data life cycle. Human analysts enact thus what could be called a 'consolidation after the fact', adding more moments of curation into data life cycles or triggering new, slightly altered variations data life cycles. Another interviewee framed the ongoing interaction between data and humans as follows:

> It is a matter of probabilities. And the software uses a certain probability threshold. I have to be aware of that. And if the software has produced an alert, I can go ahead and ask myself whether I have additional information. Can I substantiate the alert? What does that additional information tell me? And there's always the possibility of rejecting the alert. (Interview 80)

Analyses of predictive policing tend to foreground how data 'move' humans, as the algorithmic production of crime risk indicates the locations where patrol forces should be dispatched in order to enact maximum crime control. They are not wrong. The transformation of data into spatial instructions is indeed a key rationale that informs predictive policing. And yet, they neglect the multiple and intricate relations between humans and data throughout different life cycles of crime analysis and crime prevention. As we have shown through our empirical data, the life and liveliness of data are coined by a number of interaction effects. Data move people as much as people move and manipulate data. Data start out as fragmented, uncertain, and volatile. They are consolidated in order to fully unfold their analytical surplus value. At their peak, they have the power to indicate which neighborhoods should be patrolled and which ones could be neglected.

But already shortly after, data need to be either updated or discarded, as they represent only a snapshot of a dynamic world that keeps revolving around them. Data, in other words, quickly begin to lose their predictive force after they had been subjected to algorithmic analysis and 'grow old'. Once they have yielded their analytical value, their relevance quickly diminishes as they become replaced by new crime, new crime data, and new insights about criminal activities and intervention potentials. Thus, for prediction purposes these data now decrease in liveliness. Generally speaking, however, these data are far from dead. They still hold productive force for ongoing criminal investigations, for insurance claims, for court cases, and not least for the production of crime statistics and reports, where they may be re-used. Moreover, they are mobilized for analytics that seek to identify long-term trends or model the occurrence of crime in retrospect.

5. Conclusion

In exploring what the formation of information means in the context of policing, we narrated the way in which the interaction between data and humans produces knowledge and action. We have shown the many points at which human bodies take part in the generation, selection, processing and re-analysis of data and data infrastructures. In return, data and data infrastructures give form to predictive politics that have a direct impact on human bodies and behaviours. This mutual relationship becomes tangible in the domain of algorithmic policing and law enforcement. Indeed, the relationship between data lives and human lives includes the use of *force*, which underlines how important it is to carefully analyse data life cycles and their impacts on human lives. The data life cycle is here more than a mere metaphor or a notion to structure research processes. It is an analytical and theoretical concept that reveals the multiple agencies and recursive trajectories of data in professional data environments, but also in society at large. In the case of predictive policing, the entanglement of data with the production of evidence and law enforcement politics creates implications that strike at the core of the relations between the state and the citizen. That is to say, data have become so lively that they are part of reproducing order and power.

As Lupton (2016b) described, it is difficult to understand at what point data become lively, when and how they are entangled with our lives. Thus, we argue that reflections about data life cycles makes a difference in the implementation of algorithmic law enforcement. Understanding data life cycles brings the multiplicities of data trajectories, their liveliness and their power to the fore. Indeed, if a growing amount of processes in law enforcement rely on digital data - if data have become integral to productivity and human lives - it is crucial to investigate how data also saturate life with an economy of accumulation. Thus, there is a necessity to continue investigating the specificities of data life cycles. Only if we know these, we can start shaping the lives with data we want to achieve.

References

Amoore, L. (2009). Algorithmic war: everyday geographies of the war on terror. *Antipode: A Radical Journal of Geography*, 41, 49-69.

Aradau, C., & Blanke, T. (2018). Governing others: Anomaly and the algorithmic subject of security. *European Journal of International Security*, 3(1), 1-21.

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

Bechmann, A., & Bowker, G.C. (2019). Unsupervised by any other name: Hidden layers of knowledge production in artificial intelligence on social media. *Big Data & Society*, 6(1), 10.1177/2053951718819569

Bennett Moses, L. & Chan, J. (2018). Algorithmic Prediction in Policing: Assumptions, Evaluation, and Accountability. *Policing and Society*, 28(7), 806-822.

Castells, M. (1996). *The Rise of the Network Society: The Information Age – Economy, Society and Culture*, Vol. 1. Oxford: Blackwell.

Cope, N. (2008). 'Interpretation for Action?': Definitions and Potential of Crime Analysis for Policing. In T. Newburn (Ed.), *Handbook of Policing.* Cullompton/Portland: Willan Publishing, 404-429.

Deleuze, G. & Guattari, F. (1987). *A Thousand Plateaus: Capitalism and Schizophrenia.* Minneapolis: University of Minnesota Press.

Dillon, M. (2000). Poststructuralism, Complexity and Poetics. *Theory, Culture & Society*, 17(5), 1-26.

Egbert, S. & Leese, M. (2020). *Criminal Futures: Predictive Policing and Everyday Police Work.* London/New York: Routledge.

Foucault, M. (2008). *The Birth of Biopolitics. Lectures at the Collège de France 1978-79.* New York: Palgrave Macmillan.

Gitelman, L. (2013). *"Raw Data" is an Oxymoron*. Cambridge: MIT Press.

Haraway, D. (2003). *The Companion Species Manifesto: Dogs, People, and Significant Otherness.* Chicago: Prickly Paradigm Press.

Haraway, D. (2015a). Anthropocene, Capitalocene, Plantationocene, Chthulucene: Making kin. *Environmental Humanities*, 6, 159-165.

Haraway, D. (2015b). *Birth of the kennel: A lecture by Donna Haraway*, August 2000. The European Graduate School. http://www.egs.edu/faculty/donna-haraway/articles/birth-of-the-kennel/. Accessed 26 May 2015.

Harper, R.R. (1991). The Computer Game: Detectives, Suspects, and Technology. *British Journal of Criminology*, 31(3), 292-307.

Hayles, K. (1999). *How we became posthuman*. Chicago: The University of Chicago Press.

Hildebrandt, M. (2016). New Animism in Policing: Re-animating the Rule of Law? In B. Bradford, B. Jauregui, I. Loader & J. Steinberg (Eds). *The SAGE Handbook of Global Policing*, 406–28. London: Sage.

Hoad, T.F. (1996). Datum. *The Concise Oxford Dictionary of English Etymology.* 224. Oxford: Oxford University Press.

Kaufmann, M. (2019). Who connects the dots? Agents and agency in predictive policing. In M. Hoijtink & M. Leese (Eds). *Technology and Agency in International Relations*. London: Routledge. Chapter 7. 141 – 163.

Kaufmann, M. (2020). Vocations, visions and vitalities of data analysis. An introduction. *Information, Communication & Society*. 10.1080/1369118X.2020.1777320 (online first).

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

Krasmann, S. (2020). The Logic of the Surface: On the epistemology of algorithms in times of big data. *Information, Communication and Society*. 10.1080/1369118X.2020.1726986 (online first).

Lash, S. (2006). Life (Vitalism). *Theory, Culture & Society*, 23(2-3), 323-349.

Latour, B. (2013). *Facing Gaïa: Six Lectures on the Political Theology of Nature*. Gifford Lectures 18-28 February 2013. https://www.giffordlectures.org/lectures/facing-gaia-new-enquiry-natural-religion. Accessed 12 Aug 2020.

Leese, M. (2014). The New Profiling: Algorithms, Black Boxes, and the Failure of Anti-discriminatory Safeguards in the European Union. *Security Dialogue* 45(5), 494-511.

Lupton, D. (2015). Lively Data, Social Fitness and Biovalue: The Intersections of Health Self-Tracking and Social Media. *SSRN Electronic Journal*. http://dx.doi.org/10.2139/ssrn.2666324. Accessed 12 Aug 2020.

Lupton, D. (2016a). *The Quantified Self.* Cambridge: Polity Press.

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

Maltz, M.D. (1999). *Bridging Gaps in Police Crime Data: A Discussion Paper from the BJS Fellows Program.* U.S. Department of Justice. https://www.bjs.gov/content/pub/pdf/bgpcd.pdf. Accessed 12 Aug 2020.

Mol, A. (2008). I eat an apple. On theorizing subjectivities. *Subjectivity*, 22, 28–37.

Perry, W.L., McInnis, B., Price, C.C., Smith, S.C., Hollywood, J.S. (2013). *Predictive Policing: The Role of Crime Forecasting in Law Enforcement Operations*. Santa Monica: RAND Corporation.

Rose, N .(2006). *The Politics of Life Itself*. Princeton, NJ: Princeton University Press.

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

Saugmann, R. (2017). Video, algorithms and security: How digital video platforms produce post-sovereign security articulations. *Security Dialogue*, 48(4), 354–372.

Santos, R.B. (2013). *Crime Analysis With Crime Mapping*. London: Sage.

Savage, M. (2013). The 'Social Life of Methods': A Critical Introduction. *Theory, Culture & Society*, 30(4), 3-21.

TDH Netherlands (2014). *Sweetie: First Conviction in Australia.* http://www.terredeshommes.org/sweetie-first-conviction/. Accessed 20 Feb 2020.

Thacker, E. (2005) *The Global Genome: Biotechnology, Politics and Culture.* Cambridge, MA: MIT Press.

Van den Eynden, V. (2014). The Research Data Life Cycle. In L. Corti, V. Van den Eynden, L. Bishop, M. Woollard (Eds.). *Managing and Sharing Research Data*. London: Sage. 17-23.

Websites

Equivant https://www.equivant.com (Accessed 20 February 2020)

Precobs https://www.ifmpt.de (Accessed 20 February 2020)

PredPol https://www.predpol.com (Accessed 20 February 2020)

ShotSpotter https://www.shotspotter.com (Accessed 20 February 2020)

Biographies

Mareile Kaufmann investigates changing understandings of crime and practices of veillance, as well as the role of metrics and data politics in digitized societies. Her background is in Criminology, Sociology and Cultural Studies. She researches and teaches at the Department of Criminology and Sociology of Law, University of Oslo, and holds a minor position at the Peace Research Institute Oslo. Mareile works with qualitative research designs that combine theory with innovative angles and strong empirical components.

Matthias Leese is a Senior Researcher at the Center for Security Studies (CSS), ETH Zurich. His research is primarily interested in the social effects produced at the intersections of security and technology. It pays specific attention to the normative and governmental repercussions of new security technologies across society, both in intended and unintended forms. His work covers various application contexts of security technologies, including airports, borders, policing, and R&D activities.