Janet Chan

21 Datafication

Abstract: This chapter examines the origin and meaning of the term "datafication" and analyzes data practice in terms of the ideological and political nature of data, supportive and resistant narratives, and contexts and consequences of datafication.

Keywords: data, sociotechnical imaginaries, contextual factors, consequences

Datafication is a term coined by Mayer-Schönberger and Cukier in their book *Big Data* —"To datafy a phenomenon is to put it in a quantified format so that it can be tabulated and analyzed" (2013: 78). According to the authors, datafication should not be conflated with digitization (which merely converts analogue information into binary code so that it is in a computer-readable format). Jefferson (2020) suggests that police (at least in the United States) have started to represent events in the world in the form of data points (for example, in Geographical Information Systems) since the 1960s. Crime occurrence databases have existed even longer. In criminal justice, datafication has taken place not only in record keeping, but also in mapping, suspect identification, correlation, and prediction of crime occurrences (see Prediction by Ķīlis, Gundhus and Galis).

At the simplest semantic level, datafication is about turning something into "data." But what is "data"? Mayer-Schönberger and Cukier trace the word "data" to mean "given" in Latin, "in the sense of a 'fact'" but modern usage of the word "refers to a description of something that allows it to be recorded, analyzed, and reorganized" (2013: 78). In policing, Ratcliffe (2014: 99) has placed data along a DIKI (data, information, knowledge, intelligence) continuum, so that, for example, when the location of a burglary incident is recorded on a computer, it is regarded as *data*; when this incident is recognized (together with other data) by a crime analyst as an emerging pattern, it becomes *information*; when this information is shared with a detective and is used to build hypotheses about the "criminal environment," this new wisdom becomes *knowledge*; when this knowledge is used by police to mount a surveillance operation, this knowledge becomes *Intelligence* (see Intelligence by Gundhus and Lundgaard). Under this typology, datafication represents the first step in a cumulative process of transforming an incident (itself subject to processes of detection, definition, and decision to report) into a form that can be recorded, compared, and made sense of using human interpretation and/or analytical tools. Note that *Intelligence* is sometimes replaced by *Understanding* (and/or *Wisdom*) in different models (Quarmby and Young, 2010: 26–27).

With the advancement of analytic tools that can make sense of textual, sound, and visual signals, there is no reason to assume that datafication is only about *quantitative* data. For example, a frequency (count) of burglaries in a certain time period turns in-

 https://doi.org/10.1515/9783111062037-022

cidents into quantitative data, whereas photographs of stolen items—before they are labeled and counted—exist as qualitative data.

Datafication is more than a one-way process towards the accumulation of data. A productive way to conceptualize datafication is to regard it as a *practice* which involves two "interwoven processes: the use of *more and different data* … and the deployment of *more advanced methods* to analyze these data and feed it back into existing work processes" (Reutter, 2022: 905; original emphasis). To see datafication as practice has the advantage of appreciating it as both *dynamic* and *interactional:* datafication is constantly changing and its trajectory is dependent on both structural and symbolic conditions. Chan and Bennett Moses (2017), for example, make use of Pierre Bourdieu's framework to situate social practice as an interaction of the *field* (with its various types of *capital*) and the *habitus* (a generative mechanism based on shared assumptions about the nature and purpose of data) of security agents. For example, technological change can enhance the cultural capital of policing agents with technical expertise; it can also transform what agents regard as relevant information (Chan and Bennett Moses, 2017: 303). This framework implies that the practice of datafication in a particular field may be limited/facilitated by the presence/absence of structural constraints/resources as well as the nature of shared assumptions held by actors in the field.

Sheila Jasanoff's (2015) notion of sociotechnical imaginaries is useful for examining shared assumptions. Jasanoff posits that sociotechnical imaginaries are "collectively held, institutionally stabilized, and publicly performed visions of desirable futures" in relation to a specific technological change (2015: 4). This formulation acknowledges that "resistant imaginaries" can co-exist with "desirable" ones. By tracing how sociotechnical imaginaries are formed, resisted, negotiated, and implemented into "assemblages of materiality"—data assemblages that are not only manifested in values and thoughts, but also in "ownership structures, policy agendas, organizational practices, and legal frameworks"—in a particular administrative context (Reutter, 2022: 906), we are in a better position to identify the causes or mediators of harmful unintended consequences.

Thus, datafication practice can be understood in terms of (i) the ideological nature of data, (ii) the political nature of datafication, (iii) narratives that support datafication projects, (iv) narratives that resist datafication, (v) contextual factors that mediate or inhibit datafication, (vi) consequences of datafication, and (vii) data justice as a framework for action.

(i) *The ideological nature of data.* Data is more than the material presence of paper records or digital information stored on databases on physical or "cloud" storage. It is intrinsically ideological. Datafication relies on an ideology of *dataism* which assumes that data is objective and can be trusted for accurate tracking of human behaviors or social phenomena (van Dijck, 2014). In this sense, datafication is less about the production of data but more about the constitution of "symbolic and imaginative work" that makes it possible to "think of something as 'data' in the first place" (Dourish and Gómez Cruz, 2018: 2):

> Data do not speak for themselves. Data must be narrated—put to work in particular contexts, sunk into narratives that give them shape and meaning, and mobilized as part of broader processes of interpretation and meaning-making. (2018: 1)

Creating narratives is part of the process of *sensemaking* (Weick, 1995) that people engage in "to explicate the world and give it a sense of order" (Chan and Bennett Moses 2017: 302).

(ii) *The political nature of datafication.* Datafication is political in that it involves the privileging of certain ways of framing, selecting, and excluding information so that certain claims and discourses are legitimized or delegitimized (Jasanoff, 2017: 11–12; Chan et al., 2022). Dencik and Kaun (2020: 3) suggest that "Datafication … needs to be treated as a political development rather than as a technological one, and one that sets out methods of knowing and definitions of social life that have implications for what values, logics, and forms of responses are privileged over others."

(iii) *Narratives that support datafication.* Reutter (2022: 905) has found that datafication is central to visions about the future of the welfare state and public administration:

> Such sociotechnical imaginaries include notions of proactive, rather than reactive, modes of governance. Datafication provides the public sector with a sense of being able to do more, better, faster, and more cheaply and is therefore perceived as a solution to the growing complexity of society and administration and as a tool to reduce uncertainty …

Chan et al. (2022) have highlighted the symbolic attractions of data-driven policing approaches such as the "scientification" of police work, legitimizing discourse such as "smart" analytics and the likelihood of more "rational and objective" decision making.

(iv) *Narratives that resist datafication.* Critiques of datafication have arisen from many quarters. Some *police officers* see data-driven policing as "'deskilling' and 'devaluing' their experiential craft of policing … replace their discretion … [and] threaten the legitimacy of the case-based approach aimed a conviction" (Chan et al., 2022: 4). *Social scientists* have identified flaws such as "opaqueness, reinforcement of discrimination, and facilitation of surveillance" in datafication projects, while case studies have documented the unintended consequences of data-driven systems (Reutter, 2022: 904).

Chan (2021) points out that *data scientists* themselves can create narratives that resist datafication. In particular, D'Alessandro et al. (2017: 125–126) raise specific issues of discrimination that machine-learning systems can introduce, using predictive policing as a case study: sources of discrimination include data issues ("discrimination in, discrimination out"), systemic biases against certain social groups, sample bias (see Bias by Oswald and Paul), model misspecification, and process failure such as "inappropriate feedback loops" which could "lead to a self-perpetuating system that continuously targets poorer and more minority concentrated communities" (2017: 132).

Community groups who are affected by the adverse effects of datafication can also be authors of resistant imaginaries. Chan (2021: 55) has shown how community organizations, researchers, and other concerned parties in the US have expressed reservations about predictive policing, calling for a "more transparent, community-based and fair systems that are subject to informed public debate, independent evaluation and continuous monitoring as to their racial impact." In a highly controversial (and subsequently found illegal) application of data-driven fraud detection algorithm in Australia (the so-called Robotdebt Scheme which automated the income data verification and debt notification stages of welfare fraud detection), the resistant imaginary took the form of public criticisms and ultimately a class action lawsuit against the federal government. The class action resulted in a $112 million settlement in 2021 (see Commonwealth of Australia, 2023; Mann, 2020).

(v) *Contextual factors that mediate or inhibit datafication.* As Chan et al. (2022: 2) point out, the pace of datafication has been driven by "Society's demand for risk information …, technology's promise of organisational efficiency …, and the attractions of the rhetoric of 'intelligence-led policing' …". However, Reutter's (2022: 918) research shows "how policy, organizational structures, legal frameworks, subject matter experts, and existing data infrastructures are able to mediate datafication in significant ways. These constraints act as counterforces against dominant sociotechnical imaginaries strongly dominated by the private sector …." By the same token, the above-mentioned ill-conceived Robodebt Scheme in Australia was facilitated by an organizational culture that discourages the delivery of "bad news" to superiors, disregards questions about the scheme's legality, and presents obstacles to the giving of independent advice (Commonwealth of Australia, 2023: 124; see Privatization by Lomell). Research on the impact of datafication on the production of police intelligence has identified challenges such as the symbolic mismatch between police experiential knowledge and data-driven intelligence and deficits in material conditions such as technical knowledge and poor system integration (Chan et al., 2022: 11).

(vi) *Consequences of datafication.* Consequences of datafication may not become obvious until it has been implemented for some time, either through lack of independent monitoring/evaluation or difficulty of accessing reliable information. For example, the biases of the risk assessment tool COMPAS were not identified until Angwin et al. (2016), at least six years after it was rolled out by probation departments in New York State. Even though there was a "comprehensive statistical evaluation of the tool" published in 2012, racial differences were not evaluated. Similarly, the Robodebt Scheme in Australia started operating in 2015 and did not end until 2019. A robust assessment of its consequences was only published in 2023. A scheme that was designed to save $1.7 billion in its first five years had the opposite effect: some 380,000 individuals were affected and the government's subsequent reimbursement and writing off of debts amounted to $1.751 billion

(Commonwealth of Australia, 2023: 471; xxix). By the time the harms caused by ill-conceived schemes become known, it is usually too late to "fix" the problems.

(vii) *Data justice as a framework for action.* Data justice has emerged as an overarching framework for engaging with the consequences of datafication, even though the concept is itself "unsettled" (see Dencik and Sanchez-Monedero, 2022 for how different disciplines approach the analysis of and responses to data injustice). This framework recognizes that "the burdens of datafication overwhelmingly fall on resource-poor and marginalised groups in society" and invites us to "focus on what function datafication—as a discourse and practice—serves in different contexts, the social and political organisation that enables it, and who benefits" (2022: 9–10). Even though political actions may range from refusal, divestment of resources, strategic litigation, to a movement towards data sovereignty, the common ground is "a need to tackle the actual conditions that lead to experiences of injustice as they exist on the ground rather than necessarily pouring efforts into appealing to ideal formations of data and technology in contemporary society" (2022: 11). Such actions are to be "nurtured through solidarity" and social relations (2022: 11).

Conclusion

Viewing datafication as a form of *practice* that involves the translation of sociotechnical imaginaries into data assemblages provides opportunities for citizens, users, and researchers to "alter these imaginaries prior to or even during their translation" (Reutter, 2022: 918). Participating in the alteration or improvement of these imaginaries may take a number of paths. The search for data justice may be approached in different ways, depending on the circumstances and conditions that result in injustice.

- The contestation of dominant sociotechnical imaginaries may be an important step. It provides opportunities to find out what the obstacles are and how imaginaries can be changed.
- It is important to observe how sociotechnical imaginaries are negotiated by institutions, workers, and users and who are the winners and losers (Reutter, 2022).
- Participants must avoid "data-driven determinism"—instead, they should look at how "policy, organisational structures, legal frameworks, subject matter experts, and existing data infrastructures can mediate datafication in significant ways" (Reutter, 2022: 918).

More generally, researchers could investigate datafication practices in different settings using ethnographic methods such as "situational analysis" which identifies "all the discourses, objects and actors involved … as well as their relations and negotiations" and uncovers "how the structural, cultural and operational contexts … led to differing definitions and perceptions of technology" by different actors (see Sanders and Chan, 2023: 778–779).

Suggested reading

Chan, J., & Saunders, P. (eds.) (2021). *Big Data for Australian Social Policy: Developments, Benefits and Risks*, Academy of Social Sciences in Australia, Canberra, https://socialsciences.org.au/publications/big-data-for-australian-social-policy/

Egbert, S., & Leese, M. (2021). *Criminal Futures: Predictive Policing and Everyday Police Work.* London and New York: Routledge.

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

McDaniel, J. L. M., & Pease, K. G. (eds.) (2021). *Predictive Policing and Artificial Intelligence.* London: Routledge.

References

Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016). *Machine bias.* Pro-Publica, 23 May 2016 https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

Chan, J. (2021). The future of AI in policing: Exploring the sociotechnical imaginaries. In J. L. M. McDaniel and K. G. Pease (eds.) *Predictive Policing and Artificial Intelligence* (pp. 43–59). London: Routledge.

Chan, J., & Bennett Moses, L. (2017). Making sense of big data for security. *British Journal of Criminology*, 57, 299–319.

Chan, J., Sanders, C., Bennett Moses, L., & Blackmore, H. (2022). Datafication and the practice of intelligence production. *Big Data & Society*, 9(1), 1–13. DOI: 10.1177/20539517221089310

Commonwealth of Australia. (2023). *Report.* Royal Commission into the Robodebt Scheme.

d'Alessandro, B., O'Neil, C., & LaGatta, T. (2017). Conscientious classification: A data scientist's guide to discrimination aware classification. *Big Data*, 5(2), 120–134. DOI: 10.1089/ big.2016.0048.

Dencik, L., & Kaun, A. (2020). Datafication and the welfare state. *Global Perspectives*, 1(1), 12912. https://doi.org/10.1525/gp.2020.12912.

Dencik, L., & Sanchez-Monedero, J. (2022). Data justice. *Internet Policy Review*, 11(1). https://doi.org/10.14763/2022.1.1615

Dourish, P., & Gómez Cruz, E. (2018). Datafication and data fiction: Narrating data and narrating with data. *Big Data & Society*, 5(2), 1–10. DOI: 10.1177/2053951718784083

Jasanoff, S. (2015). Future imperfect: Science, technology, and the imaginations of modernity. In S. Jasanoff & S.-H. Kim (eds.), *Dreamscapes of Modernity: Sociotechnical Imaginaries and the Fabrication of Power* (pp. 1–33). Chicago: University of Chicago Press.

Jasanoff, S. (2017). Virtual, visible, and actionable: Data assemblages and the sightline of justice. *Big Data & Society*, 4(2), 1–15.

Jefferson, B. (2020). *Digitize and Punish: Racial Criminalization in the Digital Age.* Minneapolis: University of Minnesota Press

Mann, M. (2020). Technological politics of automated welfare surveillance: Social (and data) justice through critical qualitative inquiry. *Global Perspectives*, 1(1), 12991. https://doi.org/10.1525/gp.2020.12991.

Mayer-Schönberger, V. & Cukier, K. (2013). *Big Data.* London: John Murray.

Quarmby, N., & Young, L. J. (2010). *Managing Intelligence: The Art of Influence.* Sydney: Federation Press.

Ratcliffe, J. H. (2014). *Intelligence-led Policing.* London: Routledge.

Reutter, L. (2022). Constraining context: Situating datafication in public administration. *New Media and Society*, 24(4), 903–921.

Sanders, C. B., & Chan, J. (2023). Methodological reflections on researching the sociotechnical imaginaries of AI in policing. In S. Lindgren (ed.) *Handbook of Critical Studies of Artificial Intelligence* (pp. 773 – 782). Cheltenham UK and Northampton, MA, USA: Edward Elgar Publishing.

van Dijck, J. (2014). Datafication, dataism and dataveillance: Big Data between scientific paradigm and ideology. *Surveillance & Society*, 12(2), 197 – 208.

Weick, K. E. (1995). *Sensemaking in Organizations.* Thousand Oaks: Sage Publications.