

Nanna Bonde Thylstrup and Kristin Veel

## 7 Archives

**Abstract:** This chapter addresses questions about the ramifications of digital technologies on criminological research through the lens of archival theory, highlighting cases of how criminal records are used in the training of machine learning and the digital archives that arise from the digitization of domestic abuse.

**Keywords:** archival theory, criminal records, ethics, domestic abuse, machine learning

Criminology relies on different sources of data to study crime, criminal justice, and crime control. While rooted in longer trajectories, such data is increasingly collected, stored, shared, and analyzed via new mechanisms of data extraction, novel algorithmic arrangements, and emergent infrastructures of cloud computing. These mechanisms and infrastructures are embedded within broader societal forms of anticipatory governance (see *Infrastructures* by Grisot and Parmiggiani). In the field of criminology data-based anticipatory governance takes shape, for instance, in predictive policing (Kaufmann et al., 2020). Here, potential future crimes and criminals are targeted on the basis of historical crime data, records of arrests, and the known social networks of criminals (Kaufmann et al., 2019).

Large and complex datasets are, for example, generated by digital devices, platforms and networks, such as social media, mobile phones, sensors, and camera which are used to provide new insights into crime patterns, offender behavior, and victimization (Smith et al., 2017). This strand also includes the use of aggregated data collection, such as location data collected by GPS-enabled devices, for example smartphones, cars, and watches, which are then used to map spatial and temporal aspects of crime such as crime hotspots, offender mobility, and routine activities. Related here-to is also the use of digital platforms (Egbert, 2019) and machine learning (e.g., natural language processing and computer vision) to automate tasks and explore patterns in crime data (Hayward and Maas, 2020).

Digital criminology offers a range of novel perspectives on how we understand and govern crime; at the same time, digital data collection, distribution, and interpretation also present a set of more enduring ethical and political questions in criminology of how we ‘make up’ criminals (Hacking, 1986), including gendered and racialized biases and the tension between surveillance and protection. These emergent large-scale data collections and practices mean that digital criminology relates as much to the study of crime control as crime itself.

This chapter argues that the theoretical framework of archives offers a particularly well-suited point of departure for exploring how technologies present both emergent and residual issues to research. The rise and consolidation of government archives, for instance, was what enabled the state to gather information about citizens, surveil them, and on the basis of this develop new scientific approaches to criminal justice policy,

notably criminal anthropology and statistics. These statistical forms of classification enabled states to ‘make up’ (Hacking, 1986) new categories of citizens, including new categories of criminal behavior (see *Categorization and Sorting* by Franko). Archives therefore offer a productive vantage point from which to approach both residual and emergent dynamics in criminology, helping us sharpen our attention to how digitization alters the answers to questions such as: who governs the archives of crime data? Under which conditions are they gathered and made available? And, finally, how do they reshape ontologies and epistemologies of crime and crime control? In this chapter we will employ *archives* as a lens through which to address these questions about the ramifications of digital technologies on fundamental inquiries within criminological research.

## Working with archives

In broad terms, we find two overarching—and increasingly intertwined—approaches to theorizing archives. One is shaped by archival studies, which is situated within the disciplines of history, library science, and more recently, information studies. This approach is dedicated to understanding the nature, management, and uses of records and is often also tied to the profession of archivization and the discipline of history. The second approach finds its origin in poststructuralism and deconstructivism, characterized by a theoretical reframing of the archive from an objective resource into a subject of research. Anthropologist Ann Stoler called this move from archives-as-sources to archives-as-subject *the archival turn* (Stoler, 2002). The archival turn expanded the space of archival research from records management into critical analysis of archival epistemologies and ontologies. Finally, a strong body of scholarship has emerged that brings these two archival perspectives into conversation (Caswell, 2016; Thylstrup et al., 2021). Taking its point of departure in the latter strand, this section provides an overview of how archives have been theorized in the humanities and social sciences; how archival studies and cultural theories of the archive have mutually enriched each other; and, finally, how digitization has reinvigorated, and in some cases presented ruptures to, the social (after)lives of archives and archival subjects.

A classic point of departure for the ‘archival turn’ in the humanities and social sciences is French historian Arlette Farge and French philosopher Michel Foucault’s book *Disorderly Families: Infamous Letters from the Bastille Archives* (1982). The book consists of an edited collection of 94 18th-century letters from the prison archives in Bastille. The subjects in the letters include subjects of policing, which was then—as today—very broad. The archival records revolve both around then-marginalized behaviors by vagabonds and libertines, but also involve marital spouses seeking divorce and parents appealing to the authority of administrative power in conflicts with their offspring. Through their analyses, Farge and Foucault show how archives and archival collections should be understood less as static deposits of past data and more as providing “a history of the present” (Foucault, 1972) of how apparatuses of power and knowledge

shape criminal categories and subjectivities. Moreover, Farge and Foucault use the archive to offer a view of social transgression 'from below' by charting the spaces of so-called disorder through sustained attention to the lives of everyday people (as opposed to elites). In her later work, *The Allure of the Archive* (2013) Farge further sensitized readers to the significance of archives, as sites not only of information but also affect and affection. She thus reflects on the ethics of archival work, documenting both the emotional highs and lows of researchers exploring new dimensions of the past in the 'inventory rooms' and the emotional afterlives of subjects emanating from the fragile documents in the archive into the present. Scholars working on crime-related datasets used, for example, for facial recognition technologies today echo Farge's reflections on the affects of the archival encounter, emphasizing the value of thinking about critical data studies within the context of archival theories and frameworks: "These booking photos were taken from law enforcement records; the people in them were apprehended near the US-Mexico border and were most likely deported shortly after their photos were taken. It is difficult to look at the images and not feel as if they are, in some way, haunted" (Toft Djanegara, 2023: 49).

Farge's transtemporal perspective is also core to another classic archival text, *Archive Fever* by the French philosopher Jacques Derrida. Derrida famously foregrounded the important interplay between authority and interpretation in his etymological reminder that the concept of the *archive* derives from the Greek *arkhé*, which names at once a commencement and a "commandment" (Derrida, 1998: 1). He mobilized this etymological tracing to argue that archives combine a historiographical concern with origins and the concerns of law about where authority and social order are exercised. Moreover, since the etymological meaning of *archive* also stems from the Greek *arkheion*, the archive also raises questions about the 'archons,' namely the superior magistrates who command and guard the law, and who possess the 'hermeneutic right' and competence to interpret it. While Derrida, Farge, and Foucault share an interest in the relation between archive and power, Derrida shifts his perspective from the archives as sites of institutional power to also include the Archive as a symptom of humans' feverish obsession with origins and order. He therefore lodges his analysis in Freudian frameworks of memory, forgetting, and desire, rather than theories of power and sovereignty. Particularly powerful is Derrida's concept of what he calls "archive fever," an unsettling engagement with memory traces which leads to "a compulsive, repetitive, and nostalgic desire for the archive" (Derrida, 1998: 91). What Derrida calls "archive fever" is thus the "anarchival" urges that stir and destabilize orders of knowledge, making them living and dynamic sites of recollection, constantly moving and expanding, but also always at risk of (self-)destruction (Derrida, 1998: 10). Derrida's work has since been mobilized by scholars working on, for example, surveillance archives to explore both the performative power of archives and their the slippery status in their attempt to keep the lives of its subjects in an impossible orderly totality (Ring, 2014).

Echoing archives as sites of performative power that are also haunted by anarchival spectres of disorder, Achille Mbembe (2002) later ascribed the power of archives to

their role as “relic,” that is, as sites that tame the past through its inscription of the dead into a “sepulchre.” Articulating the power of (post)colonial archival regimes, Mbembe for instance points out how states have used archival institutions to “ensure that the dead do not stir up disorder” by burying not only the dead, but also “their ‘remains’” within the controlled realm of archival spaces. Because of their already established epistemic power within these spaces, states, Mbembe argued, are able to not only exercise an unquestionable interpretative authority, but also—through this interpretative authority—to tame “the demon that they carry” as colonizers. This act of “taming” through archival processes of destruction and inscription, Mbembe pointed out, was premised on archival processes of discrimination and selection, leading him to importantly also describe the archive as “not a piece of data, but as status.” Yet, Mbembe (2015) also noted that even if the Western archive disregard other epistemic traditions, “it also contains within itself the resources of its own refutation.” While acknowledging the archival institution as a site of power, Mbembe and Nuttall (2004) thus also points out the potential of epistemic resistance within archival institutions, encouraging people to both work with new archives and even with old archives in new ways.

The poststructuralist and deconstructive approaches to archives described above each in their way reconfigure perspectives on archives and archival interpretation from a positivist understanding of archives as neutral sites of knowledge retrieval to political and contingent sites of world-making (Thylstrup 2022). Such perspectives are particularly crucial for criminologists, because it speaks to archival records as relating as much to the study of crime control as to crime itself. As scholars of historical colonial and slavery archives show, for instance, colonial archival regimes not only shaped crime as a predominantly Black issue, but also engendered the imaginary of “white innocence.” Saidiya Hartman (1996) mentions the case of *State of Missouri v. Celia* (1855); here the records of the case shows that Celia is raped repeatedly by her owner from the moment she’s purchased. Resisting her pleas to stop, Celia eventually resorts to killing him, making her crime the crime on record and her the culpable agent. Situating the contemporary significance of these archival records, Simone Browne (2015) has since forcefully showed the material and social entanglements between contemporary surveillance technologies and longer histories of racialized practices of policing under slavery, including branding, runaway slave notices, and lantern laws. Moreover, mobilizing Foucault’s notion of the carceral, Tonia Sutherland (2023) shows how the fundamental link between archival power structures and archival subjectivities persist in new digital iterations with significant implications for social justice.

Contemporary critical engagements with archives not only articulate archives as sites of unidirectional exercise of power, however; they also enable an understanding of the archive as a form of technology that is always haunted by internal contradictions and external dissent (Agostinho et al., 2021; Al-Badri and Nelles, 2021). A wide-ranging and heterogenous formation of scholarly works drawing on feminist, queer, Black, post-colonial, and critical anti-racist theories thus both challenge the authority of the ar-

chive as a site of neutral storage, and harness the insights of critical and intersectional feminist theory to highlight how archival engagements might be mobilized as forms of feminist and Black resistance (Agostinho, 2019; Caswell and Cifor, 2016; Odumosu, 2020; Sutherland, 2023).

## **Analyzing digital criminology through the lens of archival theories: two cases**

In the following we offer two different examples of how emergent issues in digital criminology can be analyzed through the lens of archival theories. The first is technology-facilitated domestic abuse, the second is machine learning models. We attend in particular to how digital transformations create new patterns of crime data governance; new forms of data collection; new infrastructures of data accessibility; and new ontologies and epistemologies of crime and crime control. We thus use the examples to show how digital criminology crystallizes issues raised in contemporary feminist, post-, and anticolonial archival scholarship about access and consent, and how archival technologies are structured around, and shaped by gendered, racialized, and class-based regimes of power.

### **Digital archives of domestic abuse**

Digital technologies are often used to inflict old harms and assert power in new ways. Smartphones, GPS trackers, social media, and connected devices in the home (such as thermostats, locks, and loudspeakers) offer new avenues for what has come to be known as “technology-facilitated domestic abuse” (Yardley, 2021). As such they are also central to instances of “coercive control” (Stark, 2007) where abuse victims are monitored, controlled, and restricted in their everyday movements (see Abuse by McLinden). Significantly, the sense of a presence of a digital archive that can be retrieved and ‘found’ despite attempts to delete it is striking in the responses of survivors of intimate partner abuse. One of the interviewees recounts: “I have a friend that I was emailing and telling about the situation, and [my abuser] found out about it [...] it was deleted but it didn’t delete out of my phone like that. He went to the archives. He went through something, and found it” (Matthews et al., 2017). The archive referred to here describes a practice that is at the same time a security measure (continuously sharing accounts of the abuse with a friend), but also a repository that can be weaponized against the person being abused if ‘found.’ Since the digital traces can also be used as evidence in court against the perpetrator, the ability to master what is and what is not deleted becomes an important part of the power relation between the abuser and those subjected to abuse. Looking at such concrete criminal cases through the lens of critical archive theory allows us to scrutinize the intricate relation between power and

resistance at work and to understand how digital forms of data collection, storage, and access means that the same data can be operationalized to vastly different purposes that play out on the spectrum of care and control.

## Criminal records as training data in machine learning regimes

Our second example is how criminal records attain new afterlives as training data for machine learning technologies. Machine learning relies on massive data sets where the data used for one purpose are repurposed for others purposes (Thylstrup, 2022). A closer look at how such technologies are built also shows a contentious relation between existing archives of policing and new machine learning technologies. The United States' National Institute of Standards and Technology's (NIST) facial recognition-testing programme from 2017 offers a good example. The original purpose of the programme was to "assess facial recognition systems on an on-going basis," focusing on how the tested systems perform with respect to "accuracy, speed, storage and memory consumption, and resilience" (NIST, 2019). Yet, as Os Keyes et al. (2019) showed, the United States' government's Facial Recognition Verification Testing programme depended on

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. Additional images are drawn from the Department of Homeland Security documentation of travelers boarding aircraft in the U.S. and individuals booked on suspicion of criminal activity.

As Nina Toft Djanegara (2023) notes, such use of, for example, "border detainee booking photos" evidences a surveillance creep where facial photos of, for example, millions of immigrants become raw material for the refinement of facial recognition software not only without their knowledge or consent, but also for purposes that extend far beyond the scope of border control purposes (see Borders and Border Control by Jeandesboz). NIST's reuse of criminological archives is thus an illustrative example of how machine learning technologies become entangled (Thylstrup et al., 2022) with criminology, even in cases where the link between machine learning and criminology is not immediately apparent. Such algorithmic assemblages not only raise questions about emergent work patterns in crime control; they also facilitate new ethical relations beyond the immediate parameters of crime (Amoore, 2021).

## Conclusion

The unprecedented capacity to store, retain and share data requires critical assessment of the ethics and politics of digital archival technologies in criminology. The study of crime control must embrace how the rise of new archives ties in with broader tendencies in the tech world such as "surveillance capitalism" (Zuboff, 2019). As this chapter

shows, digital technologies offer both new usage for crime and crime control as well as new challenges. Archival theories offer a good vantage point from which to gauge how these advances and challenges are shaped by both emergent dynamics and historic trajectories. These trajectories can be found in how technologies facilitate 'access' and 'governance' as well as in the consent structures made available to archival subjects and the broader extraction processes they are made subject to. But they can also be found in questions about how to square democratic principles of access and data sharing with the ethics of care and consent (Agostinho et al., 2021). Critical archival approaches thus bring out not only how digitized crime/control is ingrained in deeper affective, temporal, and political structures, but also how the authority of these structures is also always haunted by uncertainty and instability. Ultimately, an understanding of critical archival theories empowers the researcher to navigate the complex landscape of digital phenomena relevant to criminology and contribute to the ongoing conversations about access, governance, power, and ethics in relation to emergent digital technologies.

## Suggested reading

Derrida, J. (1996). *Archive Fever: A Freudian Impression*. Chicago: University of Chicago Press.

Foucault, M. (1972). *The Archaeology of Knowledge*. New York: Pantheon Books.

Hamilton, C., Harris, V., Taylor, J., Pickover, M., Reid, G., & Saleh, R. (eds.) (2002). *Refiguring the Archive*. Dordrecht: Springer Science & Business Media.

Stoler, A. L. (2010). *Along the Archival Grain*. Princeton, NJ: Princeton University Press.

Sutherland, T. (2023). *Resurrecting the Black Body: Race and the Digital Afterlife*. Berkeley: University of California Press.

## References

Al-Badri, N. & Nelles, N. (2021). Technoheritage. In N. B. Thylstrup, D. Agostinho, A. Ring, C. D'Ignazio, & K. Veel (eds.). (2021). *Uncertain Archives: Critical Keywords for Big Data* (pp. 499–512). Cambridge, MA: MIT Press.

Agostinho, D. (2019) Archival encounters: Rethinking access and care in digital colonial archives. *Archival Science*, 19(2), 141–165.

Agostinho, D., Gade, S., Thylstrup, N. B., and Veel, K. (2021). *(W)archives: Archival Imaginaries, War, and Contemporary Art*. Berlin: Sternberg Press.

Amoore, L. (2021). Ethics. In N. B. Thylstrup, D. Agostinho, A. Ring, C. D'Ignazio, & K. Veel (eds.), *Uncertain Archives: Critical Keywords for Big Data* (pp. 201–208). Cambridge, MA: MIT Press.

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

Caswell, M. (2016). 'The archive' is not an archives: Acknowledging the intellectual contributions of archival studies. *Reconstruction: Studies in Contemporary Culture*, 16(1).

Caswell, M., & Cifor, M. (2016). From human rights to feminist ethics: Radical empathy in the archives. *Archivaria*, 81(1), 23–43.

Derrida, J. (1998). *Archive Fever: A Freudian Impression*. Chicago: University of Chicago Press.

Egbert, S. (2019). Predictive policing and the platformization of police work. *Surveillance & Society*, 17(1/2), 83–88.

Farge, A. (2013). *The Allure of the Archives*, transl. Thomas Scott-Railton. New Haven, CT: Yale University Press. [Original edition, Arlette Farge (1989). *Le goût de l'archive* (Paris: Seuil)]

Farge, A., & Foucault, M. (1982). *Le Désordre des familles: Lettres de cachet des Archives de la Bastille au XVIIIe siècle*. Paris: Gallimard.

Foucault, M. (1972). *The Archaeology of Knowledge*. New York: Pantheon Books.

Hacking, I. (1986). Making up people. In T. Heller, M. Sosna, & D. E. Wellbery (eds.), *Reconstructing Individualism: Autonomy, Individuality, and the Self in Western Thought* (pp. 222–236). Stanford: Stanford University Press.

Hartman, S. (1996). Seduction and the ruses of power. *Callaloo*, 19(2), 537–560.

Hayward, K. J., & Maas, M. M. (2021). Artificial intelligence and crime: A primer for criminologists. *Crime, Media, Culture*, 17(2), 209–233. <https://doi.org/10.1177/1741659020917434>

Kaufmann, M., Egbert, S., & Leese, M. (2019). Predictive policing and the politics of patterns. *The British Journal of Criminology*, 59(3), 674–692.

Kaufmann, M., Thylstrup, N. B., Burgess, J. P., & Sætnan, A. R. (2020). Engaging the data moment. *STS Encounters*, 11(1), 227–254. [https://dasts.dk/wp-content/uploads/9\\_Data\\_Criticality\\_FV\\_1.pdf](https://dasts.dk/wp-content/uploads/9_Data_Criticality_FV_1.pdf)

Keyes, O., Steven, N., & Wernimont, J. (2019). The government is using the most vulnerable people to test facial recognition software. *Slate Magazine*, 17 March. Available at: <https://slate.com/technology/2019/03/facial-recognition-nist-verification-testing-data-sets-children-immigrants-consent.html> (accessed 29 April 2019).

Matthews, T., O'Leary, K., Turner, A., Sleeper, M., Woelfer, J. P., Shelton, M., Manthorne, C., Churchill, E. F., & Consolvo, S. (2017). Stories from survivors: Privacy & security practices when coping with intimate partner abuse. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (pp. 2189–2201).

Mbembe, A. (2015). Decolonizing Knowledge and the Question of the Archive. Public lecture held at given at the Wits Institute for Social and Economic Research (WISER), University of the Witwatersrand (Johannesburg). Available at: <https://wiser.wits.ac.za/system/files/Achille%20Mbembe%20-%20Decolonizing%20Knowledge%20and%20the%20Question%20of%20the%20Archive.pdf>. [accessed July 26, 2024)

Mbembe, A. (2002). The power of the archive and its limits. In C. Hamilton, V. Harris, J. Taylor, M. Pickover, G. Reid, & R. Saleh (eds.), *Refiguring the Archive* (pp. 19–27). Dordrecht: Springer Netherlands.

Mbembé, J. A., & Nuttal, S. (2004). Writing the world from an African metropolis. *Public Culture*, 16(3), 347–372.

NIST (National Institute of Standards and Technology) (2019). FRVT 1:1 verification. Available at: [www.nist.gov/programs-projects/frvt-11-verification](http://www.nist.gov/programs-projects/frvt-11-verification) (accessed 29 April 2019).

Odumosu, T. (2020). The crying child: On colonial archives, digitization, and ethics of care in the cultural commons. *Current Anthropology*, 61(S22), S289–S302.

Ring, A. (2014). The (W) hole in the Archive. *Paragraph*, 37(3), 387–402.

Smith, Gavin J. D., Bennett Moses, L., & Chan, J. (2017). The challenges of doing criminology in the big data era: Towards a digital and data-driven approach. *The British Journal of Criminology*, 57(2), 259–274. <https://doi.org/10.1093/bjc/azw096>

Stark, E. (2007). *Coercive Control: How Men Entrap Women in Personal Life*. Oxford: Oxford University Press.

Stoler, A. L. (2002). Colonial archives and the arts of governance. *Archival Science*, 2, 87–109.

Sutherland, T. (2023). *Resurrecting the Black Body: Race and the Digital Afterlife*. Berkeley: University of California Press.

Thylstrup, N. B. (2022). The ethics and politics of data sets in the age of machine learning: Deleting traces and encountering remains. *Media, Culture & Society*, 44(4), 655–671.

Thylstrup, N. B., Agostinho, D., Ring, A., D'Ignazio, C., & Veel, K. (eds.). (2021). *Uncertain Archives: Critical Keywords for Big Data*. Cambridge, MA: MIT Press.

Thylstrup, N. B., Hansen, K. B., Flyverbom, M., & Amoore, L. (2022). Politics of data reuse in machine learning systems: Theorizing reuse entanglements. *Big Data & Society*, 9(2), 20539517221139785.

Toft Djanegara, N. (2023). Ghosts in the machine. *Media-N*, 19(1), 42 – 67.

Trouillot, M.-R. (1995). *Silencing the Past: Power and the Production of History*. Boston: Beacon Press

Yardley, E. (2021). Technology-facilitated domestic abuse in political economy: A new theoretical framework. *Violence Against Women*, 27(10), 1479 – 1498. <https://doi.org/10.1177/1077801220947172>

Zuboff, S. (2019). *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. New York: PublicAffairs.

