

2022

## Review of Uncertain Archives: Critical Keywords for Big Data

Marissa Friedman

University of California, Berkeley, [marissa.friedman@berkeley.edu](mailto:marissa.friedman@berkeley.edu)

Follow this and additional works at: <https://elischolar.library.yale.edu/jcas>



Part of the [Archival Science Commons](#)

---

### Recommended Citation

Friedman, Marissa (2022) "Review of Uncertain Archives: Critical Keywords for Big Data," *Journal of Contemporary Archival Studies*: Vol. 9 , Article 5.

Available at: <https://elischolar.library.yale.edu/jcas/vol9/iss1/5>

This Book Review is brought to you for free and open access by EliScholar – A Digital Platform for Scholarly Publishing at Yale. It has been accepted for inclusion in *Journal of Contemporary Archival Studies* by an authorized editor of EliScholar – A Digital Platform for Scholarly Publishing at Yale. For more information, please contact [elischolar@yale.edu](mailto:elischolar@yale.edu).

**Nanna Bonde Thylstrup et al., eds. *Uncertain Archives: Critical Keywords for Big Data*. Cambridge: MIT Press, 2021.**

The impetus for *Uncertain Archives: Critical Keywords for Big Data* emerged from a series of workshops held by the Uncertain Archives research group, founded in 2014 by members of the Department of Arts and Cultural Studies, University of Copenhagen.<sup>1</sup> The Uncertain Archives collective has since grown to encompass a wide range of “practice-based scholars, archivists, artists, designers, activists, and computer scientists” based in Denmark and abroad (3). The editors make clear that the book, while by no means comprehensive, attempts to mirror the interdisciplinary engagements fostered in those original workshops. In its entirety this volume—a hefty 624 pages—is enormously dense and rich in scope, introducing readers to a wide range of issues related to the creation, ownership, storage, preservation, dissemination, ethical concerns, and impact of big data in our contemporary society. Its five editors and seventy-three authors span an impressive array of different academic disciplines, artistic practices, and professional or activist backgrounds.

*Uncertain Archives* is prefaced by an introductory chapter from the editors that situates the book within a larger discourse around the nature of exclusion, omission, and uncertainty in the archives. In particular, it draws heavily on poststructuralist, critical, decolonial, queer, and feminist theory. According to the editors, these theoretical lenses reveal big data archives as embodied, risky, at-risk, and affective sites of both oppression and potential liberation. The editors define “big data” as “the sheer volume of data . . . that are too numerous to be processed by human minds, and are therefore subject to analysis, and archiving, by smart machines” (1). Their definition, however, as well as much of the analyses found in the book, are primarily concerned with vast bodies of information collected by corporate and governmental agencies for proprietary, surveillance, or other purposes that are not strictly archival, at least not in a more traditional sense, although references to other kinds of digital cultural heritage collections are present in the text. The age of big data, in the editors’ minds, represents a significant shift in “the notion of the archive . . . from regime of knowledge about the past to a regime of future anticipation” or prediction (1). While the introduction presents a compelling framework for larger debates about modern (digital) archival praxis and theory, it also raises questions about the exact relationship between big data and “the archives.”

Readers will likely wonder where big data archives live; how they are collected, preserved, and managed; and who they are for. The editors posit an expansive, if not illusive, definition of what counts as an archive in the age of big data, challenging us to consider the ways in which unstable, fragile, and often proprietary big data archives break down traditional notions of archives as collections governed by standard appraisal, preservation, and description practices. Many of the data archives contextualized within the text are collected and activated as parts of vast, distributed, and often proprietary surveillance systems. Governed by opaque human-machine processes, these archives are utilized in ways contrary to traditional archival paradigms and are concerned primarily with profit. And considering how central the concept is to the entire project, the text exhibits a noticeable fuzziness around what exactly constitutes a big data archive—is it the data itself or the sociotechnical infrastructures and tools that rely on data, or either or both? And is it sufficient to

---

<sup>1</sup> Uncertain Archives, <https://artsandculturalstudies.ku.dk/research/focus/uncertainarchives/>.

call any collection of digital information and/or digital objects an “archive” if it exists solely within proprietary or other spaces in which preservation and access are not necessarily attended to? For example, the editors posit that YouTube’s collections of user-generated content constitute a big data archives even while admitting that it is a highly unstable and fragile archive due to the lack of preservation functions built into the platform. But are web platforms archives? In her 2020 article “Big Questions: Digital Preservation of Big Data in Government,” Emily Larson raises a salient critique worth nothing here: “Big Data is sometimes defined as an archives in and of itself, but these claims often emphasize storage over archival principles, particularly authenticity and long-term preservation.”<sup>2</sup> At times, *Uncertain Archives* risks falling into this trap of attributing any large-scale compilation of digital objects as an “archive.” Clearly, this is a useful reminder to attend to the complexity and contested nature of terminology when it comes to the (digital) archive.<sup>3</sup> The tension over where the cutoff lies for determining what constitutes a big data archive is left unresolved and open to further debate.

*Uncertain Archives* is presented as an encyclopedia featuring sixty-one critical terms, each accompanied by brief essays with bibliographies for further exploration. Each entry is written as a stand-alone piece, and while there are clearly related terms, due to the alphabetical sorting of terms and the diversity of viewpoints presented, there is not as much repetition as you might expect across the volume. The book is framed as a “centrifugal force” that unveils the multiplicities in approaches and ways of thinking about uncertainty in big data archives by including “not only . . . coexisting but also . . . conflicting voices, linked to different cultural meanings, ideologies, and materialities within the same linguistic space” (13). The resulting cacophony of perspectives included clearly drives home this point, but the editors’ choice to publish the entries alphabetically rather than in any sort of thematic grouping also significantly hinders the accessibility of the book. Due to the array of disciplinary approaches included, and because many of the entries assume varying levels of prior familiarity with specific disciplinary terminology or bodies of literature, the volume would have benefited from a more structured organization with greater editorial guidance for readers. Reading straight through the book is not particularly useful and may in fact be confusing, given that some entries are more explicitly linked to the overall theme of the book than others. The choice to present the book as an encyclopedia, while clearly aligned with the editorial intentions for the work, may leave many readers feeling adrift. The editors could have provided thematic roadmaps to link together related entries and help make the material more comprehensible to a greater audience.

The volume proposes two main overarching theses. First, it argues that it is the uncertainties of these archives which are in dire need of analysis, and that it is critical to interrogate the presumptions of certainty, precision, and universality baked into big data regimes. Second, the book makes the case for a big tent, interdisciplinary, and humanities-based approach to interrogating big data archives “because big data interact at every level with the human”—it is material actions and events that sustain seemingly disembodied machine intelligence (1). The editors also outline several broad thematic connections among the disparate entries in the book’s introduction. Major areas covered in the text include the opaque, disjointed, and often

---

<sup>2</sup> Emily Larson, “Big Questions: Digital Preservation of Big Data in Government,” *American Archivist* 83, no. 1 (2020): 6.

<sup>3</sup> Contributor Sara T. Roberts notes the “slippage and application of metaphor” (11), describing how, for example, the boundary between the archive and the database appears to be increasingly fuzzier.

indecipherable natures of big data archives; representation, perception, and power; the complex mechanical and epistemological systems (including knowledge organization, material processes, cultural values, and logistical structures) that support big data systems; where authority, expertise, and control reside in these systems (and who is allowed to describe, access, and act on this data, and to what ends); the fragility and instability of big data, as information is created, consumed, and potentially lost within increasingly rapid and opaque computing processes; the uncertain ethical terrain big data archives pose; and the dangers minoritized communities face when corporations and governments employ predictive and proprietary data algorithms to make decisions. One of the major premises of the book is that in their production and activation, big data archives are not precise, neutral, transparent, or disembodied. Rather, the book suggests that a growing reliance on big data in the realm of government, business, and culture has produced a huge array of intended and unintended consequences. These arguments may not necessarily be new to many information professionals, but given the number of contributors and disciplinary approaches here, it is likely that readers will encounter unfamiliar or thought-provoking ideas.

One major critique of the “archival turn” among humanities scholarship is that it has prioritized the theorizing of “the archive” in ways that perpetuate fundamental misunderstandings of what constitutes an archive (for example, archivists would likely agree that a website is not an archive, nor is the internet in and of itself one). B. M. Watson’s blog post “Please Stop Calling Things Archives” aptly describes the implications of this “tendency towards the over-casual use of the word ‘archive’ as a shorthand to refer to, well, just about anything.”<sup>4</sup> As previously noted, *Uncertain Archives* does, at times, stray into this territory of potentially over-theorizing the “big data archive” along these lines. Crucially, the tendency to misconstrue “the archive” in some humanities scholarship also has another effect: to erase or neglect considerations of practices, material processes, and human labor that actually maintain archives. Where do archivists who work in the cultural heritage and galleries, libraries, archives, and museum (GLAM) sector fit into this context? While not meant to be a practical resource guide for archivists, information workers seeking more concrete insights into how the ideas and theories raised in the book have been or can be put into practice by archivists will likely need to turn to more LIS-specific literature. While critical archival studies and related disciplines have attempted to fill in these gaps, it is important to think about how archivists can engage with, contribute to, and resist some of the tendencies and observations raised in this book. What role do archives workers play in identifying, mitigating, and actively resisting the worst abuses and threatening features of big data regimes? Several entries provide interesting theoretical perspectives on these and related issues, which are discussed below. While there are many concepts raised in the collection, this review focuses only on contributions that are of particular relevance and interest to information professionals. These are presented alphabetically to mirror the organizing principles of the book.

### **Critical Theory Approaches**

Readers are introduced to key theoretical lenses and methodological approaches that may be familiar to those versed in critical archival scholarship and who will appreciate seeing their applications to big data archives. In “Affect,” Marika Cifor draws on a rich body of literature from

---

<sup>4</sup> B. M. Watson, “Please Stop Calling Things Archives,” *Perspectives on History* (blog), American Historical Association, January 22, 2021, <https://www.historians.org/publications-and-directories/perspectives-on-history/-january-2021/please-stop-calling-things-archives-an-archivists-plea>.

critical archival studies to help readers understand the significance of affect for evaluating the risks, impacts, and possibilities of digital archives and big data. First, Cifor argues that digital archives are evocative, embodied, and affective entities capable of generating liberatory and radical relations, which is part of why they are so compelling. And yet, her entry cautions readers to consider the ways affect becomes an exploitable resource in privatized digital spaces, to be harnessed and deployed by business interests against vulnerable populations.

In the entry on “Abuse,” Sarah T. Roberts also calls on critical archival theory to evaluate an archive of user-generated content found in PhotoDNA, a digital database that deploys computer automation to capture and store disturbing materials related to child sexual exploitation. Roberts notes that content moderation tools like PhotoDNA raise a number of uncertainties related to whether they constitute an archive, how they reflect power relations and privilege interests of profit over moral or ethical ones, and what the removal of user-generated content in this manner means in social and political terms. And in “Flesh,” Romi Ron Morrison draws on critical data studies and Black feminist scholarship to explore resistance to algorithmically generated violence and data management practices designed to circumscribe and control unruly, vulnerable bodies. Morrison’s work is a valuable reminder of the longstanding archival discourse on the objectifying and disciplining tendencies of recordkeeping on minoritized bodies and offers some interesting insights about how the archival field might begin to redress this.

### **Digital Humanities**

In the entry “Digital Humanities,” Roopika Risam makes the case for a more global and critical digital humanities that has “a core responsibility . . . to build an inclusive digital cultural record” (162). This should be a collective endeavor among cultural heritage practitioners, librarians, archivists, scholars, and students to identify whose places are most precarious in the archive and rectify the silences and injustices present in prevailing trends in digitization; according to Risam, the digital archival record up to this point merely replicates existing canons of representation to privilege the identities, values, and worldviews of those who have had the most power to shape the historical record. Her overview of different critically minded movements within the digital humanities, including global digital humanities, #transformDH and feminist digital humanities, U.S. Latinx digital humanities (#usLdf), and Indigenous digital humanities, raises questions about representation, access, and ethics that are contextually and culturally interdependent, and pushes back against “overdetermining influence of values of the Global North” (167). Her entry explicitly calls for information professionals to attend to discrepancies in representation within the digital cultural record, noting that critical and diverse perspectives can reshape approaches to the creation and dissemination of digital cultural heritage archives.

### **Field/Cooling**

In my mind one of the most urgent issues this book addresses is the uncertainties of producing and maintaining big data archives in terms of climate change. The environmental impacts of maintaining digital archives are of increasing importance among the archival community. Shannon Mattern’s essay on “Field” reminds us that maintaining archives of any sort of records requires an increasing amount of energy and natural resources, and the extraction of these resources may in turn threaten to increase the severity of environmental crises into the future. In “Cooling,” Nicole

Starosielski examines the extreme fragilities and vulnerabilities of digital record infrastructures to heat, noting how digital archives, much like their physical counterparts, are dependent on cold or cool storage even as they generate incredible amounts of heat. Data centers are not intangible entities in the “cloud”; rather, data are stored in material infrastructures that consume large amounts of energy to stave off self-destruction. Archives are thus rife with “existential uncertainties,” which have only been amplified by the increasing size and scope of digital collections (231). What does it look like, particularly in the age of big data, to manage archives responsibly and to contribute to a more sustainable and resilient world? Throughout it all, the specter of loss haunts many of the entries in this book. How much of our cultural and informational record will be lost to the physical ravages of a world in crisis, to profit-driven corporations and governmental surveillance regimes? How much can, or should, be saved? These are not new questions for archivists, but there is great value in thinking through these questions from the variety of different disciplinary standpoints provided in the book.

### **Hashtag Archiving**

Just as the text challenges what counts as an archive, it also challenges traditional conceptions of who counts as an archivist. In her entry on hashtag archiving, Tara L. Conley introduces readers to this concept as both a research methodology and archival practice. According to Conley, hashtag archiving refers to the process of capturing and building publicly accessible data repositories of born-digital social media content. Her framework posits researchers and community members working in the vein of ethnography as archivists. From an archival perspective, this entry might be better understood as offering points of convergence where archivists can work with ethnographers and other researchers to preserve digital collections into the future.

In another vein, Shannon Mattern considers the perception of the terrestrial plane as an archive within the geosciences in her entry “Field,” treating climate scientists and geologists tasked with collecting, recording, classifying, storing, and analyzing vast amounts of extracted terrestrial and oceanic material as archivists in charge of maintaining complex geological archives. According to Mattern, through their specific archival practices, climate scientists “make rocks into records” (229). Her entry laments the lack of professional collaboration or dialogue between geologic archives and paper-based archives. This suggests there is a need for LIS professionals to adopt a more expansive understanding of what constitutes an archive and archival record. Archives workers should remain open to more collaboration with other fields that are preoccupied with maintaining records in various forms, and that are facing similar sources of uncertainty, from the logistical challenges of managing vast collections to lack of funding and support, faulty infrastructures, shifting standards, and the threat of global climate change.

### **Metadata**

If disrupting the presumption of neutrality, transparency, and clarity in big data and digital archives is a continuing refrain in this book, Amelia Acker’s entry on “Metadata” extends this theme in relation to the function of metadata in big data regimes. She raises particularly interesting questions about how metadata is generated, by and about whom this data is collected, who has access to this data, and how feasible it is to opt out of these infrastructures “of data creation collection, control, access, and reuse through metadata technologies” (327). Acker asks us to consider: in a big data

culture that posits data sharing as the single most important means of social and cultural belonging, what is the stake of representation in big data archives that are controlled and exploited for corporate gain? Metadata is not neutral and reflects the perspectives and values of those who have the power to name, define, classify, and organize. Archivists and other information professionals can help to untangle these complex webs of data extraction and management, uncover the problematic categories and representational structures that govern so much of our individual experiences within big data systems, and continue to engage in efforts for “metadata justice” by changing descriptive standards and practices when it comes to physical and digital archival collections (323). Additionally, this framing could also apply to the kinds of systems and platforms used within archives and other cultural heritage repositories to track patrons, users, and collections alike, prompting a consideration of where ownership of that data lies.

### **Misgendering**

In “Misgendering,” Os Keyes offers a compelling plea for critical data studies that consider “the contextuality of gender specifically and identity more broadly, the way data systems work to *strip* contextuality and how we might preserve and rebuild that. It must consider not only the flows of data but the eddies: the *dead* data, left static and in place until [it] is reanimated in a temporal context where it can do no harm” (344). Strategic alignments between those goals and the professional responsibilities and expertise of archivists can help push back against these conditions, restore context to aggregates of data, and engage more proactively with the ethical dimensions of access to and discovery of data. In particular, these entries point to the ways in which archivists will increasingly be called on to practice archival activism in mediating and advocating for the needs and concerns of data subjects.

Archivists can work toward an “ethics of care” as outlined by Daniela Agostinho in her entry on “Care,” which prioritizes the ethical responsibility archivists have to documented subjects as well as to users. Agostinho reminds us of the colonialist underpinnings of archival stewardship, which have caused and continue to cause harm on some bodies more than others through dispossessing, excluding, silencing, or misattributing them in the archive. What does it mean for archivists to participate in reparative or restitutive action/intervention in our digital world? What does a decolonial ethics of care look like in practice for archivists managing digital collections? The field continues to have much to learn from decolonial and Black feminist theory in this regard, as scholars and information professionals continue to explore what archival activism entails in the age of big data.

### **Remains**

Tonia Sutherland’s excellent entry on “Remains” demonstrates not just a thought-provoking application of critical and Black feminist theory to understanding digital remains as embodied records but also raises pivotal questions concerning ethics and the duties of care archivists may have in relation to preserving the digital afterlives of Black lives. Her description of digital records as embodied is a crucial distinction here, as it returns the focus to the material bodies that produce digital records and the physical stakes of these human lives being commodified, exploited, or otherwise circulated digitally, often without consent or context. As Sutherland notes, how Black lives are lived, died, remembered, and reconstituted through their digital traces “is forever

intimately linked to systemic and structural practices of anti-Black (and often state-sponsored) violence, a violence too frequently reinscribed and reified in—and then justified by—the archival record” (461). Sutherland’s work asks readers to consider who is granted the right to be remembered and/or the right to be forgotten.

When considered alongside entries such as “Algorithmic Racism” by Alana Lentin—which uses decolonial computing theory to unravel the ways racism and white supremacy are baked into digital tools and spaces, thus reproducing harm on racialized bodies—these perspectives push back against the presumed neutrality of (digital) archives. *Uncertain Archives* compellingly demonstrates the value of adopting intersectional methodologies informed by Black feminist thought, as advocated by Brooklyn Gipson, Frances Corry, and Safiya Umoja Noble in their entry on “Intersectionality,” when considering digital archival practices. Intersectional frameworks of analysis allow information workers operating in the trenches of information infrastructures to better visualize “the normative value systems of whiteness, patriarchy, heteronormativity, and beyond at work in the mass collection, storage, and organization of data,” and resist those systems that “facilitate erasure” of minoritized, racialized, (mis)gendered bodies (306). This work also helps to make the case for principled archival activism in the face of clear structural inequities and the ongoing perpetuation of violence against particular peoples in and beyond the (digital) archive.

## Reparative

Reparations and redress are presented as key interventions in disrupting the normalization and maintenance of racialized, gendered, and sexualized violence within big data regimes. It may be useful for readers to consider the “Reparative” entry, in which Katrine Dirckinck-Holmfeld applies the “notion of reparative practice/reparative reading,” first presented by American queer feminist and literary theorist Eve Kosofsky Sedwick, to digital archives (443). What would it look like for archivists to adopt a “reparative critical praxis that does not seek to repair or restore a preexisting whole” but rather embraces “radical, creative, decolonial, and technological imagination?” (452). Dirckinck-Holmfeld’s entry implies the need for new and radical archival imaginaries and a productive engagement with arts-based modes of inquiry to help grapple with these issues. It is worth noting that the concept of reparative archival work has been raised in many other contexts, from Lae’l Hughes-Watkins’s excellent article on “creating inclusive spaces for marginalized voices” in academic repositories and the development and application of antiracist description practices (as seen in the efforts of groups such as Archives for Black Lives in Philadelphia) to the ongoing work in the realm of (digital) repatriation of Indigenous archival materials, a process that has been written about by scholars such as Kimberley Christen.<sup>5</sup>

## Conclusion

---

<sup>5</sup> Lae’l Hughes-Watkins, “Moving toward a Reparative Archive: A Roadmap for a Holistic Approach to Disrupting Homogenous Histories in Academic Repositories and Creating Inclusive Spaces for Marginalized Voices,” *Journal of Contemporary Archival Studies* 5 (2018), <https://elischolar.library.yale.edu/cgi/viewcontent.cgi?article=1045&-context=jcas>; Archives for Black Lives in Philadelphia, “Anti-Racist Description Resources” (2019), [https://-archivesforblacklives.files.wordpress.com/2019/10/ardr\\_final.pdf](https://-archivesforblacklives.files.wordpress.com/2019/10/ardr_final.pdf); Kimberly Christen Withey, “Opening Archives: Respectful Repatriation,” *American Archivist* 74, no. 1 (2011): 185–210.

For those interested in LIS approaches to big data, *Uncertain Archives* is part of a growing body of scholarship concerning the growth of digital archives and the correlation between digital collections and big data. For example, in “Critical Questions for Archives as (Big) Data,” Devon Mordell explores some potential ramifications for what she terms the emerging “archives-as-data paradigm,” arguing that this paradigm tends toward reinforcing traditional and highly problematic notions of archival neutrality.<sup>6</sup> Mordell’s argument intersects with some of the conclusions drawn in *Uncertain Archives*; both share a common concern with disrupting the presumptions of precision, universality, and completeness of big data (archives). While *Uncertain Archives* is concerned with questions of data creation and access—who controls the production of and availability of data and for what ends—others have tackled questions of preservation, description, and maintenance. Emily Larson has written on the digital preservation needs for big data collected by governmental bodies and the ways in which archivists need to critically intervene through both technical systems development and documentation to ensure the long-term preservation of this data and its complex contexts. These and other articles provide a welcome application of archival principles to managing big data as specific types of archival records.<sup>7</sup> And proponents of the emerging “transdiscipline” of computational archival science such as Richard Marciano and Victoria Lemiux are concerned with developing new tools, techniques, and training for archivists managing digital records at scale.<sup>8</sup>

While its encyclopedic organization does hinder the book’s overall accessibility, *Uncertain Archives* presents some useful and important theoretical frameworks for archivists working with digitized and born-digital collections. In its entirety, the book provides a complex analysis of present and possible future impacts of big data across many aspects of human life and organization. It raises thought-provoking questions and areas of inquiry for archivists tasked with collecting, preserving, describing, and providing access to exponentially growing digital collections. It may also be of interest to archivists and special collections librarians participating in the collections as data movement and other projects that seek to make special collections material available for computational analysis in the form of data.<sup>9</sup>

---

<sup>6</sup> Devon Mordell, “Critical Questions for Archives As (Big) Data,” *Archivaria* 87 (May 2019): 140–61.

<sup>7</sup> Larson, “Big Questions.”

<sup>8</sup> For example, see Richard Marciano et al., “Archival Records and Training in the Age of Big Data,” in *Re-Envisioning the MLS: Perspectives on the Future of Library and Information Science Education*, ed. Johanna Percell et al. (Bingley, U.K.: Emerald, 2018), 179–99, <https://doi.org/10.1108/s0065-28302018000044b010>.

<sup>9</sup> Always Already Computational—Collections as Data, 2020, <https://collectionsasdata.github.io/>.