

INTO THE STACKS

A Tour of the Virtual Stacks

Cameron Blevins

Rows upon rows of “virtual stacks” now stretch as far as the eye can see. From JSTOR to the Library of Congress to Ancestry.com, unprecedented quantities of historical material are being added to the digital ether. In fact, you are probably reading these words on a screen right now.¹ Search-box interfaces allow historians to instantly query vast quantities of historical material in order to pull out information about individuals, events, institutions, and locations. With just a few strokes of a keyboard, a historian can sift through millions of digitized pages of newspapers, government documents, or books. A process that would have once taken a lifetime of flipping through microfilm or archival folders can be conducted in just a few minutes. As historian Lara Putnam notes, this now “feels as revolutionary as oatmeal.” But, she argues, the “mass digitized turn” has nevertheless had a profound impact on the practice of history in ways that the discipline is only beginning to understand.² This is especially true for a field like modern American history, where an abundance of easily scannable English-language sources has generated a wealth of online material.

In the past few years, a much more critical conversation has emerged around the limitations of the virtual stacks. Putnam describes an array of problems that come with text search. Digitization can often amplify preexisting disparities: an overrepresentation of Anglophone material, for instance, or the inability for scholars from resource-poor countries or institutions to access expensive paywalled databases. Keyword searching in the virtual stacks influences our choice of historical subjects and the kinds of stories we tell about them, often subtly nudging us toward topics or people that are easiest to locate in an online database. The silences in the traditional archive—of people of color, women, the impoverished or illiterate, the Global South—are further magnified by the seemingly limitless scope of digitization. When it feels like you are searching everything, it is easy to forget just how much is actually missing.³

Moreover, many of the institutions that house the virtual stacks are private companies like Google, LexisNexis, or Readex that are interested in generating profits rather than providing permanent, stable, or free access to the public. The virtual stacks might be vast, but many of them are remain closed off. Google Books—the company’s starry-eyed ambition to build an online collection of every book that has ever been published—offers a cautionary tale. Since 2011, it has been bogged down with a lawsuit focusing on copyright law and the public domain.

¹Journals like *Law and History Review*, *Modern Intellectual History*, *Journal of Sport History*, and *The Public Historian* have all published roundtables, overviews, or special issues about the impact of digital history on their particular subfields. See Elizabeth Dale, ed., “In This Issue: Digital Law and History,” *Law and History Review* 34, no. 4 (Nov. 2016): v–vi; Dan Edelstein, “Intellectual History and Digital Humanities,” *Modern Intellectual History* 13, no. 1 (Apr. 2016): 237–46; Jennifer J. Sterling, Murray G. Phillips, and Mary G. McDonald, “Doing Sport History in the Digital Present,” *Journal of Sport History* 44, no. 2 (Summer 2017): 135–45; William Bryans et al., “Imagining the Digital Future of The Public Historian,” *The Public Historian* 35, no. 1 (Feb. 2013): 8–27.

²Lara Putnam, “The Transnational and the Text-Searchable: Digitized Sources and the Shadows They Cast,” *The American Historical Review* 121, no. 2 (Apr. 2016): 377–402, here 380.

³Putnam, “The Transnational and the Text-Searchable.” See also Julia Laite, “The Emmet’s Inch: Small History in a Digital Age,” *Journal of Social History*, doi: 10.1093/jsh/shy118 (accessed Mar. 29, 2019).

The company has since all but shuttered its book-scanning operations, still approximately 100 million volumes short of its original goal.⁴

Copyright law is especially pertinent for historians of the twentieth- and twenty-first-century United States. Prior to January 1, 2019, the year 1922 was a dividing line. Anything published after this date was still under copyright, and therefore illegal to make freely available online in full-text form. The virtual stacks fell off a virtual cliff from 1923 forward. On January 1, that line inched forward to 1923—the first of what will be an annual extension of the public domain by one-year increments. It was a landmark change to copyright. Just three months earlier, in September 2018, there had been a quieter landmark. HathiTrust, a publicly available, nonprofit, digital library, announced that it was making the entirety of its 16.7 million items available to researchers, *including* post-1922 material still under copyright. It came with one important caveat: the collection was only available for “non-consumptive research.” This legalistic language means that somebody cannot just go onto HathiTrust and download a copy of Sylvia Plath’s *The Bell Jar* to read on their couch. That same person can, however, use text mining or data visualization tools to analyze lexical patterns across the roughly 65,000–70,000 words in *The Bell Jar*, or compare Plath’s vocabulary to hundreds of other twentieth-century novelists.⁵

HathiTrust’s 2018 announcement was a milestone for scholars using computational text analysis, a method of looking for empirical patterns across large collections of text. This approach expands the available source base (or, in more scientific terms, the sample size). For instance, if a historian wants to know how newspapers covered the 1918 influenza epidemic, he or she might have spent several years traveling to different archives and poring over a few dozen microfilmed newspapers. Or he or she could follow the example of a team of scholars at Virginia Tech who applied computational techniques to thousands of digitized newspaper pages from across the country in order to unearth patterns in where and how coverage about the epidemic spread.⁶ Similarly, the historian Michelle Moravec has used a range of computational approaches to study the history of women’s suffrage, feminist artists, and gender disparities in Wikipedia.⁷ In the field of diplomatic history, Micki Kaufman has done groundbreaking computational research using some 18,600 telephone conversations and memoranda from former Secretary of State Henry Kissinger.⁸ And perhaps not surprisingly, literary scholars have been some of the most enthusiastic adopters of computational text analysis in the humanities. Intellectual and cultural historians of the modern United States should peruse the *Journal of Cultural Analytics*, which has rapidly become a leading outlet for scholarship in this field. Several recent articles in the journal have revealed important literary patterns about gender

⁴James Somers, “Torching the Modern-Day Library of Alexandria,” *The Atlantic*, Apr. 20, 2017, <https://www.theatlantic.com/technology/archive/2017/04/the-tragedy-of-google-books/523320/> (accessed Mar. 29, 2019).

⁵Jessica Rohr, “HathiTrust Research Center Extends Non-Consumptive Research Tools to Copyrighted Materials: Expanding Research through Fair Use,” *Perspectives from HathiTrust* (blog), Sept. 20, 2018, <https://www.hathitrust.org/blogs/perspectives-from-hathitrust/hathitrust-research-center-extends-non-consumptive-research-tools> (accessed on Mar. 29, 2019).

⁶E. Thomas Ewing et al., “An Epidemiology of Information: Data Mining the 1918 Influenza Pandemic,” White Paper (Washington, DC, 2014), <https://securegrants.nih.gov/publicquery/main.aspx?f=1&gn=HJ-50067-12> (accessed Mar. 29, 2019).

⁷Michelle Moravec, “‘Under This Name She Is Fityly Described’: A Digital History of Gender in the History of Woman Suffrage,” *Women and Social Movements* 19, no. 1 (Mar. 2015), <http://womhist.alexanderstreet.com/moravec-full.html>; Michelle Moravec, “Network Analysis and Feminist Artists,” *Artl@s Bulletin* 6, no. 3 (Nov. 2017), <https://docs.lib.purdue.edu/artlas/vol6/iss3/5>; Michelle Moravec, “The Endless Night of Wikipedia’s Notable Woman Problem,” *Boundary 2*, Aug. 1, 2018, <https://www.boundary2.org/2018/08/moravec/> (accessed Mar. 29, 2019).

⁸Micki Kaufman, “‘Everything on Paper Will Be Used Against Me’: Quantifying Kissinger,” <https://blog.quantifyingkissinger.com/> (accessed Mar. 29, 2019).

and race by analyzing thousands of English-language novels spanning the twentieth- and twenty-first centuries.⁹

As the “virtual stacks” expand, computational methods give scholars a means of grappling with this new archival scale. But of course textual sources are only one kind of historical evidence. Some of the most exciting digital work has coalesced around non-textual sources, including maps, photographs, film, music, architecture, and other kinds of material long used by historians of the modern United States. Digital mapping is one of the most established of these approaches that emerged with Historical Geographical Information Systems (HGIS) in the late 1990s and early 2000s.¹⁰ In recent years, mapping has helped scholars make major inroads into studies of race, segregation, and social justice in the twentieth-century United States. The Mapping Inequality project, for instance, has overlaid the Home Owners’ Loan Corporation’s notorious “redlining” maps from the 1930s onto contemporary maps of some 150 American cities, driving home the enduring impact of racist federal housing practices on the modern urban landscape.¹¹ Other mapping projects have studied racial segregation in specific American cities or cataloged a landscape of racial violence during the early twentieth century.¹²

Scholars are also branching out into other sorts of historical source material. Digital sound studies has emerged as a coherent field of historical inquiry.¹³ Dance historians have begun to use digital techniques to study the lives and contributions of past performers.¹⁴ More broadly, Lauren Tilton recently issued a call for a “visual turn” in digital history, or harnessing computational tools to process and analyze photographs, film, and other visual media.¹⁵ As Tilton notes, this “visual turn” will increasingly draw from computer vision, a subset of machine learning. This computational approach has exploded in recent years, in which a computer uses “training sets” of pre-processed data to build models and predictions that it can then apply to future sets of raw data. It is the technology behind facial recognition: with enough photographs that have been identified as, say, Rosa Parks, an algorithm can “teach” itself to identify *other* photographs of Parks. The implications for historians are profound, not just in terms of retrieving information from media archives but also surfacing patterns across those same sources. For instance, Tilton and her collaborators are using computer vision to analyze gender and narrative arcs across tens of thousands of hours of television sitcoms from the 1950s and 1960s.¹⁶

If facial recognition and artificial intelligence give you pause, you are not alone. This is, perhaps paradoxically, an area where historians can and should offer much-needed expertise and perspective to our colleagues in computer science. “Data” are never neutral; they are collected

⁹Ted Underwood, David Bamman, and Sabrina Lee, “The Transformation of Gender in English-Language Fiction,” *Journal of Cultural Analytics*, Feb. 13, 2018, doi: 10.22148/16.019; Eve Kraicer and Andrew Piper, “Social Characters: The Hierarchy of Gender in Contemporary English-Language Fiction,” *Journal of Cultural Analytics*, Jan. 30, 2019, doi: 10.22148/16.032; Richard Jean So, Hoyt Long, and Yuancheng Zhu, “Race, Writing, and Computation: Racial Difference and the US Novel, 1880–2000,” *Journal of Cultural Analytics*, Jan. 11, 2019, doi: 10.22148/16.031 (accessed Mar. 29, 2019).

¹⁰Anne Kelly Knowles, “GIS and History,” in *Placing History: How Maps, Spatial Data, and GIS Are Changing Historical Scholarship*, eds. Amy Hillier and Anne Kelly Knowles (Redlands, CA, 2008), 1–27.

¹¹Robert K. Nelson, LaDale Winling, Richard Marciano, Nathan Connolly, et al., “Mapping Inequality,” in *American Panorama*, ed. Robert K. Nelson and Edward L. Ayers, <https://dsl.richmond.edu/panorama/redlining> (accessed Mar. 29, 2019).

¹²Sarah Bond, “How Is Digital Mapping Changing The Way We Visualize Racism and Segregation?,” *Forbes*, Oct. 20, 2017, <https://www.forbes.com/sites/drsarahbond/2017/10/20/how-is-digital-mapping-changing-the-way-we-visualize-racism-and-segregation/>; Mara Cherkasky, Sarah Jane Schoenfeld, and Brian Kraft, “Mapping Segregation in Washington DC,” Prologue DC, <http://www.mappingsegregationdc.org/#about>; Monica Martinez, “Mapping Violence,” <http://mappingviolence.org/> (accessed Mar. 29, 2019).

¹³Mary Caton Lingold, Darren Mueller, and Whitney Trettien, eds., *Digital Sound Studies* (Durham, NC, 2018).

¹⁴Harmony Bench and Kate Elswit, “Mapping Movement on the Move: Dance Touring and Digital Methods,” *Theatre Journal* 68, no. 4 (Dec. 2016): 575–96, doi: 10.1353/tj.2016.0107.

¹⁵Lauren Tilton, “Towards a Visual Turn in (Digital) History” (Quantitative Analysis and the Digital Turn in Historical Studies, Fields Institute, 2019), <http://laurentilton.com/files/visualturnv3.pdf> (accessed Mar. 29, 2019).


¹⁶Lauren Tilton and Taylor Arnold, “Distant Viewing: Analyzing Large Visual Corpora,” *Digital Scholarship in the Humanities*, doi: 10.1093/digitalsh/fqz013 (accessed April 12, 2019).

and preserved by people and institutions that operate within particular historical settings and societal contexts. For historians, this is a rudimentary observation. But it is vital for understanding the technology that defines so much of our world. To take one example: try typing “history professor” into Google Image’s search-box and you will find yourself awash in an ocean of older white men standing behind lecterns or in front of bookshelves. As Safiya Noble details in *Algorithms of Oppression*, these kinds of search results (and much more harmful ones) are not explicitly programmed to be racist or sexist. But algorithms based on “training datasets” consisting of billions of prior searches that have been shaped by structural racism and sexism are, in turn, going to generate search results that are racist and sexist. The problem is magnified by the particular institutional context of Google, a corporation with a non-diverse workforce whose decisions will subtly reflect the values and worldview of a social elite.¹⁷

Once we reframe “data” in terms of sources and archives, it turns out that historians have quite a bit to contribute to this topic. Marisa Fuentes and Jessica Marie Johnson, for example, have both detailed how the archive of the Atlantic World was shaped by slavery’s violence and the commodification and erasure of black bodies, and how modern scholars’ use of these colonial documents has often reinforced, in Johnson’s words, the ongoing “thingification of black women, children, and men.”¹⁸ To take a more modern example, researchers Os Keyes, Nikki Stevens, and Jacqueline Wernimont recently described a government program that uses a database of millions of images in order to help private companies evaluate the accuracy of their facial recognition technology. They discovered that this database includes police mugshots and images of U.S. visa applicants (especially those from Mexico).¹⁹

Mugshots and visa photos do not make up some objective, neutral dataset. They are a quite particular archive of the American state, the product of decades worth of racist incarceration and immigration policies that have equated blackness with criminality and Mexican immigrants as “illegal aliens.” Khalil Gibran Muhammad, Kali Nicole Gross, Kelly Lytle Hernández, and Mae Ngai are just a few of the historians whose work can (and must) inform our understanding of how the use of such photographs will only reinscribe this racist history into today’s facial recognition software.²⁰ This kind of historical and humanistic approach is exactly the sort of perspective that is so vital in contemporary debates about technology. Whether or not historians learn how to write code or publish interactive maps, the discipline needs to build a more sophisticated understanding of the virtual stacks and their implications.

Cameron Blevins is assistant professor of history at Northeastern University and a core faculty member of the NULab for Texts, Maps, and Networks, where he teaches and researches digital history and the nineteenth-century United States. He is completing a book project that presents a spatial interpretation of the American state and its territorial expansion through the infrastructure of the U.S. Post.”

Author ORCID.  Cameron Blevins, 0000-0002-5272-5770

¹⁷Safiya Umoja Noble, *Algorithms of Oppression: How Search Engines Reinforce Racism* (New York, 2018).

¹⁸Marisa J. Fuentes, *Dispossessed Lives: Enslaved Women, Violence, and the Archive* (Philadelphia, 2016); Jessica Marie Johnson, “Markup Bodies: Black [Life] Studies and Slavery [Death] Studies at the Digital Crossroads,” *Social Text* 36, no. 4 (Dec. 2018): 57–79, doi: 10.1215/01642472-7145658.

¹⁹Os Keyes, Nikki Stevens, and Jacqueline Wernimont, “The Government Is Using the Most Vulnerable People to Test Facial Recognition Software,” *Slate Magazine*, Mar. 17, 2019, <https://slate.com/technology/2019/03/facial-recognition-nist-verification-testing-data-sets-children-immigrants-consent.html> (accessed Mar. 29, 2019).

²⁰Khalil Gibran Muhammad, *The Condemnation of Blackness: Race, Crime, and the Making of Modern Urban America* (Cambridge, MA, 2010); Kali Nicole Gross, “Policing Black Women’s and Black Girls’ Bodies in the Carceral United States,” *Souls* 20, no. 1 (Jan. 2018): 1–13; Kelly Lytle Hernández, *City of Inmates: Conquest, Rebellion, and the Rise of Human Caging in Los Angeles, 1771–1965* (Chapel Hill, NC, 2017); Mae M. Ngai, *Impossible Subjects: Illegal Aliens and the Making of Modern America* (Princeton, NJ, 2005). See also the 2015 special issue “Historians and the Carceral State” in the *Journal of American History*: Kelly Lytle Hernández, Khalil Gibran Muhammad, and Heather Ann Thompson, “Introduction: Constructing the Carceral State,” *Journal of American History* 102, no. 1 (June 2015): 18–24.