Technical Literature and the Text-Searchable: The History of Technology and the Digitized Turn

Jeffrey T. Manuel*   DOI: 10.15763/jou.ts.2021.01.05.04

_____

We historians of technology are constantly assessing technology and placing it in historical context. But from time to time, it is useful to consider how the technologies we rely on shape the stories we tell about technology. Specifically, how has the so-called digitized turn within the broader discipline of history affected the history of technology? By this, I mean more than asking whether historians of technology are doing digital history. Certainly, fascinating and high-profile digital history projects challenge historians of technology to consider new ways of making and disseminating knowledge.[1]

Digital history projects are important, but there has been a deeper, more pervasive change in historical practice that is usefully called the digitized turn. It includes the torrent of source digitization underway and the ever-widening availability of

---

* Copyright 2020 Jeffrey T. Manuel. Jeffrey T. Manuel is an associate professor in the Department of History at Southern Illinois University Edwardsville.

[1] Examples of digital history projects that focus on the history of technology include Honghong Tinn, Tyson Vaughan, and Lisa Onaga's online resources on Fukashima, teach311.org, and the team of scholars, led by Gregg Mitmann, documenting the rubber industry in Liberia via the web exhibit, "A Liberian Journey." Lisa Onaga and Hanna Rose Shell, "Digital Histories of Disasters: History of Technology through Social Media," *Technology and Culture* 57, no. 1 (2016): 225–30; Gregg Mitman et al., *A Liberian Journey: History, Memory, and the Making of a Nation*, http://liberianhistory.org/ (accessed March 24, 2019).

digitized, text-searchable sources accessible via the web. Although self-identified digital historians are now common, nearly every historian working today is "applying technology to their historical work through database software and online source repositories, making them what Ian Milligan calls "unwitting digital historians."[2] Lara Putnam likewise describes how the digitized turn is reshaping historical practice in her 2016 *American Historical Review* article, "The Transnational and the Text-Searchable: Digitized Sources and the Shadows They Cast."[3] Practicing history amid the digitized turn, Putnam argues, "makes new realms of connection visible, new kinds of questions answerable. At the same time, the new topography of information has systematic blind spots. It opens shortcuts that enable ignorance as well as knowledge." Putnam's central concern is how the digitized turn is affecting transnational history, where, she argues, "digital search offers release from place-based research practices that have been central to our discipline's epistemology and ethics alike."[4] Historians of technology ought to ask similar questions about how the now-ubiquitous tools we use to tell technology's stories—digitized books and journals, scanned and OCRed documents and publications, source databases, and word processors—are shaping the stories we tell. No field of the historical enterprise has thought longer or harder about the relationship between tools and stories than historians of technology. It is worth considering the tools we use to tell those histories and how new digital tools enable certain stories about technology but make others more difficult to tell.

Technical Literature and Trade Journals

Technical trade journals are one set of sources that illustrate how the digitized turn affects the history of technology. It is no coincidence that so many historians of technology have worked at technical universities with rich library holdings in technical journals and trade publications. There are many reasons why the history of technology,

---

[2] Ian Milligan, "Illusionary Order: Online Databases, Optical Character Recognition, and Canadian History, 1997–2010," *Canadian Historical Review* 94, no. 4 (November 2013): 540–69.
[3] Lara Putnam, "The Transnational and the Text-Searchable: Digitized Sources and the Shadows They Cast," *American Historical Review* 121, no. 2 (April 2016): 377–402.
[4] Putnam, "The Transnational and the Text-Searchable," 379.

as a field, evolved symbiotically with technical universities. But certainly, one reason why these settings have proven so generative for historians of technology is that historians working in these universities have incredible opportunities to use technical trade journals as sources. I imagine a great many outstanding books and articles in our field began with a historian browsing the stacks in an engineering or scientific library where they worked.

But working in scientific and engineering libraries led to unintended consequences that have been significant for the history of technology. Simply spending time in scientific and engineering libraries to read articles, take notes, review microfilm, and do the things historians used to do, often created productive friction between historians of technology and other users of these spaces. While searching for old issues of an electrical engineering journal, you ran into electrical engineering students. In reading technical conference reports from a century ago, it was impossible to ignore the other, non-technical issues raised by engineers and scientists in their professional publications. In short, accessing these sources in this way led to a peripheral awareness of the broader context in which technical issues developed. Such contextual awareness has been central to the history of technology; it certainly was the case for my dissertation research, which began just before the digitized turn. I wrote a dissertation on the history of iron ore mining, specifically the transition between soft hematite ores and engineered pelletized ores in the United States steel industry. I vividly remember trekking to the University of Minnesota's Walter Library, where the engineering collections were held—a long, cold walk from Wilson Library, which housed the non-technical history collections—and navigating the books and students I encountered there. I began my research with a cynical view of the mining and metallurgical engineers who developed pelletized iron ore, assuming they negligently ignored the environmental consequences that resulted from their work. Yet as I watched the engineering students around me—reading, talking, buying the same unhealthy snacks at the vending machine—I had a hunch that the engineers I was studying from a century earlier were surprisingly like these young people: bright, complex humans who were eager to solve problems and often surprised by the consequences of what they made and did. I

suspect other historians of technology have similar ones to tell. Put simply, the frictions involved in accessing technical sources have been an important source of contextual awareness for historians of technology.

Figure 1. U.S. National Archives staff photographer Donald C. Holmes using a microfilm camera to copy a bound volume, 1/15/1937. (Source: United States National Archives and Records Administration, Series: Historic Photograph File of National Archives Events and Personnel, 1935–75
Record Group 64: Records of the National Archives and Records Administration, 1789–ca. 2007.)

Papers of Inventors and Other Technologists

Manuscript collections of inventors, engineers, and technologists offer another example of how frictions generated by the research process emphasized technology's context. Working in the personal papers of inventors has been an important method for historians of technology since the field's creation. Projects to make technologists' papers available to researchers, such as the Thomas A. Edison Papers Project, are invaluable. For researchers who made use of such collections, it was impossible to ignore how technical matters were interspersed with non-technical context. Lab reports and flow charts sit side-by-side with personal correspondence, handwritten notes, and receipts. The haphazard jumble was even more pronounced in unprocessed collections since archivists impose order and logic on the messy complexity of the paperwork individuals leave behind. I remember searching through an archival box and finding a mortgage deed for a house purchased by a mining engineer at the beginning of his

career with the Minnesota Mines Experiment Station.[5] I was looking for evidence of the engineer's role in iron ore development, but the mortgage humanized him, especially since I had just purchased a home and was nervous about my new financial responsibilities. It is a truism among historians of technology that invention and innovation are social processes carried out by specific people in particular historical contexts. It is less widely acknowledged but equally true that the frictions of researching in technologists' papers strongly reinforce this contextual awareness. Yet, the digitized turn is changing historians' relationship to archival collections. Manuscript collections held in archives are far from obsolete, but it is no longer a given that researchers will travel to them to access their riches. And when in the archive, historians now act as human scanning machines, processing as much material as they can for later analysis. Stays in the archives are short, and vast troves of unprocessed images on our hard drives are now a running joke among academic historians.

Historical Newspapers

Historical newspapers offer another example of how the digitized turn has rearranged historical scholarship and the technological infrastructure that has shaped the digitized turn. Today, vast, text-searchable newspaper archives make it easy to find stories that mention a certain term or phrase across years and even decades. (At least for those scholars who have access to an institutional subscription or pay the digital toll.) Previously, newspaper archives required access to an index—available only for the largest newspapers or, in rare cases, created by a heroic local librarian or archivist—microfilm, and countless hours squinting as the negatives rolled in front of our eyes. Microfilm was the primary technology used to preserve newspapers during the twentieth century. Unlike bound volumes of old papers, microfilm archives were compact, and the work could be outsourced to companies like Bell and Howell.[6] Although, as any user of microfilm collections could tell you, microfilm only worked within a system of supporting

---

[5] This research was published as Jeffrey T. Manuel, "Mr. Taconite: Edward W. Davis and the Promotion of Low-Grade Iron Ore, 1913–1955," *Technology and Culture* 54, no. 2 (April 2013): 317–45.

[6] Kathleen A. Hansen and Nora Paul, *Future-Proofing the News: Preserving the First Draft of History* (Lanham, MD: Rowman & Littlefield, 2017), 173.

technologies, including a newspaper index and microfilm readers. And even though accessing newspapers via microfilm was certainly different from reading old, bound copies, scrolling through page after page of microfilm built a similar type of contextual awareness of the era. Which historian has not spent hours intrigued and amused by advertisements in the sources they examined?

But starting around the turn of the twenty-first century projects to digitize old newspapers changed how researchers accessed these sources. The *Toronto Star* was the world's first newspaper to be fully digitized. Just before the turn of the century, a company called Cold North Wind, Inc. scanned the entire 110-year run of the *Star*. This took four months of work. (Cold North Wind, Inc., was better known for its affiliated web site, Paper of Record. Both were acquired by Google in 2008 and incorporated into Google's News Archive, which Google shut down in 2011.)[7] Cold North Wind's primary motivation in scanning the *Star* was financial, not archival. The company's founder claimed, "online digital archives show great promise in generating new profit from old content for newspapers." As the company gained experience in digitizing newspapers, it began to automate much of the work, lower costs, and turn a profit. The costs to digitize the Toronto *Globe* were recouped within ninety days.[8] Yet, in an era of modem connections, small monitors, and slow personal computers, accessing large image files was not necessarily faster than the older microfilm technology. Experienced researchers seeking a particular day's newspaper were more likely to fire up the microfilm reader than to wait for a slow-loading image on their computer screen. Thus, Cold North Wind turned to keyword searching as a technological fix for moving through a huge corpus of image files.[9]

Yet translating image files into a keyword-searchable database required an intermediate technology, optical character recognition (OCR). OCR is a complex technology developed to allow computers to read printed text. It was, Milligan notes, "originally and primarily designed for the efficient digitization of reams of corporate and

---

[7] Hanson and Paul, *Future Proofing the News*, 175.
[8] "All the News That's Fit to Scan," *Financial Post*, July 2003.
[9] Milligan, "Illusory Order," 559–60.

legal documents, conventionally formatted. Applying these tools, initially designed for specific commercial applications, to historical documents yields mixed results."[10] OCR proved far from accurate when reading old newspapers—rife with column breaks, uneven typography, and far less consistency than legal documents—and thus, the databases created by scanning newspapers were riddled with errors. Despite the errors and omissions built into the digitized newspaper archives, historians came to rely on them. Citations of those newspapers that had been digitized skyrocketed between 1998–2010, while newspapers that remained unscanned saw no corresponding increase in citations. Canadian historians increased their citation of the Toronto *Globe and Mail* by an astonishing 926 percent in this period.[11] The ease of finding information via keyword search in these newspaper archives has been irresistible for historians. The underlying technologies—scanning, OCR, and the database—hardly offer a transparent window into the underlying print archive.

The technologies that enable digitized, text-searchable sources also engage with older technologies in complex ways that deserve more attention from historians of technology. The invisibility of handwritten sources in the digital archive is one example. OCR has thus far been limited to printed materials, so the vast body of handwritten sources still requires tedious and expensive transcription before it can be made keyword searchable. This technical limitation has begun to skew the research of historians who are ever-more reliant on digital access and keyword searches, overrepresenting some traces of the past and hiding others. As Rebecca Wright describes in a recent article, digitization of the Mass Observation Archive, a premier record of modern British social history, has skewed historians' access to the archive and, thus, our understanding of British social history. Since the Mass Observation Archive was digitized by Adam Matthews in 2007, most researchers engage with the archive's online database rather than the paper archive. Before digitization, researchers were guided through the collection by curators. Today, keyword search is the primary entrance to the

---

[10] Milligan, "Illusory Order," 542.
[11] Milligan, "Illusory Order," 541–42, 549–50.

collection.[12] Digitization of the archive was uneven, however, since the daily logs at the heart of the archive switched from handwriting to typewriting in the early twentieth century. Only the typewritten logs could be "read" by OCR and made searchable. Thus, researchers relying on the digital archive are seeing only a portion of the total archive. "Reliance on typewritten material thus elevates particular social experiences above others," Wright argues. And switching from handwriting to typewriting reshaped respondents' daily logs in subtle but significant ways. Switching from pen and paper to a typewriter often changed what observers recorded about their days. Daily reports became more bureaucratic and streamlined when typed.[13] New technologies of digital access to sources are always layered atop earlier technologies of information—writing, the archival box, the typewriter—and each additional layer rearranges our view of the past.

Figure 2. A researcher uses a microfilm machine in the U.S. National Archives in 1972.
(Source: National Archives and Records Administration, item 35810070, Series: Historic Photograph File of National Archives Events and Personnel, 1935–75, Record Group 64: Records of the National Archives and Records Administration, 1789–ca. 2007.)

---

[12] Rebecca K. Wright, "Typewriting Mass Observation Online: Media Imprints on the Digital Archive," *History Workshop Journal* 87 (February 2019): 118, 121.
[13] Wright, "Typewriting Mass Observation Online," 127, 130–31.

Friction and Encounter in the Digitized Turn

The digitized turn has rearranged the frictions of research and the encounters inherent in doing scholarship. Accessing many sources is different today than it was twenty years ago. Google's search bar is often the first stop for our students (and most scholars, if we're honest). This change has been enabled by technologies of automated scanning, optical character recognition, and databases that make it easier than ever to locate references to specific terms across a vast corpus. Certain frictions inherent in research have undoubtedly lessened. Finding and tracking down isolated references in primary sources used to involve time-consuming treks around the library and, occasionally, around the world. Today, Google Books and HathiTrust bring these references to your browser window in milliseconds.

Although the digitized turn is reshaping how historians work, we need to be wary of telling a deterministic story about these technologies and their effects on historians. Indeed, pushing back against seamless narratives of technological determinism has been a great strength of the history of technology. Historians of technology, therefore, have an especially important role to play in reminding the larger historical enterprise that new information technologies have not revolutionized our work but rather rearranged it. An important task for historians of technology is reminding our colleagues of earlier technologies of scholarship and examining how those tools shaped the stories we told about the past. The humble index, mentioned above, was one such tool. Library classification systems have long allowed historians to scan up and down a shelf of books seeking titles that caught our attention. But the frictions of using such technologies were significant. Relying on magazine and newspaper articles for one chapter of my dissertation, I spent months in my university library's windowless periodicals room shuttling between worn, old copies of *Readers Guide to Periodical Literature* and bound magazines arranged by title. Today, many of these articles come to my web browser in a fraction of the time (provided that I have access to the expensive online index).

Historians of technology are uniquely positioned to examine the choices—political, social, and cultural—baked into the technologies we now rely on to produce scholarship. The algorithms that return hits from the search bar are technologies made by individuals in specific historical contexts, although we often have to look harder for the traces of this historical context than we did when facing the archival box. Yet, a shelf of excellent recent work by historians of technology has shown how those algorithms were shaped by gender, race, capitalism, and many other "contextual" matters. And despite the ease and satisfaction of finding isolated references in keyword-searchable archives, great gaps exist in the digitized record, especially beyond the west or for marginalized groups. We have to remember this as it becomes ever more tempting to treat the available digitized sources as the total archive. The digitized turn has given historians of technology many new tools, but, as Ian Milligan argues, "a critical methodology is needed if historians are to use these tools responsibly.