

An Archaeology of Victorian Newspapers

PAUL FYFE

As Robert Darnton sees it, we are living amidst the most recent of “four fundamental changes in information technology” that have patterned human history.¹ Coming just before our age of digital transition, he claims, was the Victorian epoch, which was transformed by a dramatic expansion of texts and readers. Darnton’s sketch is at once debatable and elegant, offering an attractive homology to nineteenth-century scholars who would connect these moments of change, past and present.² In the forms of digitized historical resources and computational research methods, the connections between media shifts then and now can not only be theorized but also operationalized, with the nineteenth century’s prolific sources serving as the materials for twenty-first century digital humanities research. This essay results from just such an initiative—a content-mining project focused on the digital collection *British Nineteenth-Century Newspapers* from the commercial publisher Gale Cengage. Yet my goal is not to explain the contemporary challenges of computational approaches, the scale of digital materials, or a culture of proliferating information that connects us to the Victorians. Instead, this essay calls attention to the gaps in that story: the largely hidden history of how Victorian data gets to now. It argues that our justifiable enthusiasm for linking past and present has effectively erased the interval between—the twentieth-century transmission histories that established the parameters for scholarly resources in digital forms. New media is always in the process of constituting itself as new, erasing the legacies of its entanglements and the continuous work of its propagation.³ This essay follows the lead of several scholars in media studies and critical bibliography to outline—and then pursue—a method for investigating these material histories, an “archaeology” of data, to better grasp the historiography of our research objects, which are expressed, for the moment, in digital form. Such an approach enables us not only to understand the mediated

conditions of scholarly materials but also to engage in informed critique of how they are created, sold, accessed, and used, whether by casual users or scholars interested in computational techniques.

From Media Literacy to Media Archaeology

On November 11, 2014, the North Carolina State University (NCSU) Libraries announced that they had reached an innovative agreement with the commercial publisher Gale Cengage to license data-mining rights to its digital collections.⁴ Within a week, Gale Cengage announced in its own press release that it would generally “make available content from its *Gale Digital Collections* to academic researchers for data mining and textual analysis purposes.”⁵ In other words, institutional subscribers would be able to request the source files behind Gale’s web-based interfaces. This blanket agreement, the first of its kind, was the result of long negotiations between Gale Cengage and Darby Orcutt of the NCSU Libraries.⁶ The terms of the agreement with Gale included “content mining rights” for digital collections, which, unlike “textual analysis” or “data mining,” allow for the fullest range of analytical approaches, such as computer vision and image analytics. The license includes the NCSU Libraries’ standard subscription to Gale’s web-based search interface for these collections, as well as physical hard drives containing the source data (for a nominal “cost recovery” charge). Soon the drives arrived by mail (figure 1).

Excited by this seemingly direct access to materials that had been screened by web interfaces and varying institutional subscriptions, I began to explore what the drives contained, dreaming of the new prospects they opened. However, my elation quickly changed to stupefaction as I began sorting through prosaic directories, title manifests, XML files, and image files. My excitement further degraded upon closely examining what the files encoded. Gale’s product does offer solid metadata, article segmentation, page facsimiles, and searchability. It is perhaps unfair to assess a database based on the first page of a historical newspaper, which often contains advertisements that are incredibly challenging for OCR. Alas, that is where I first looked, randomly perusing a front page of the *Huddersfield Chronicle* from April 6, 1850 (figure 2). I compared this scanned image with its XML rendering: within the layers of metadata provided for the newspaper, the issue, and the first article, each word appeared wrapped in page coordinates along with a 46 percent confidence score judging the OCR’s accuracy in recognizing the text (figure 3).⁷ In using data from Gale Cengage’s *British Nineteenth-Century Newspapers* collection, what exactly am I dealing with? As Lisa Gitelman and Virginia Jackson argue, “raw data” is an oxymoron; the phrase obscures the discursive conditions and technical



Figure 1. Hard drives of *19th-Century British Library Newspapers*, February 2, 2015.
Photo by Markus Wust, NCSU Libraries.

particularities of its own production.⁸ Data is always already cooked or, as Johanna Drucker suggests, “captured” within a set of epistemological presumptions.⁹ How did it get this way? If commercial publishers like Gale Cengage are increasingly offering access to source data, as well as selling their own subscription-based portals for analyzing digital texts, scholars must learn what the data comprises and how it has evolved.¹⁰ The efficacy of our scholarship depends upon a largely missing source history of these digital collections.

The sense of awe summoned by big data, along with the rhetoric of paradigmatic discovery in digital research, is a symptom of a digital sublime better understood as sublimation. In physics, sublimation occurs when substances skip a physical state change, as when dry ice goes directly from solid to gas without an intermediate liquid stage. It is a useful metaphor for digital resources, which can seem to erase any intermediary state between source object and digital surrogate in the cloud. Thanks to the work of many scholars, today we seldom approach digital scholarly resources so naively, and we have become increasingly sensitive to the formal changes of scholarly objects as they move through different mediated states. Jim Mussell, for instance, has called for critical bibliographies of *contemporary* digital objects and delivery platforms to match our study of *historical* text technologies and their materiality. He calls this “historically reflexive media literacy”—a phrase that enables much productive crossover between new media, past and present.¹¹ Yet its very diachronicity may overlook what has happened in between, the entanglement of digital scholarly resources



Figure 2. Image of the front page of the *Huddersfield Chronicle*, April 6, 1850, in *19th-Century British Library Newspapers*, part II.

with time frames other than nineteenth-century history. I would argue that scholars of Victorian digital media have tended to focus on form, or on the formal repercussions of encoding, which privileges media literacy at the expense of materialist histories of how Victorian media evolved into its digital states. In literary studies, scholars are certainly comfortable talking about the material transmission of books and periodicals, but they are less comfortable, perhaps, understanding the life cycle of digital resources, particularly those generated by commercial scholarly publishers, whose history reaches back one hundred years. This history weaves through

```

1 <?xml version="1.0" encoding="UTF-8"?>
2 <!DOCTYPE issue SYSTEM "bl_nonp_issue_apex.dtd">
3 <issue>
4   <id>HUCE-1850-04-06</id>
5   <jn>Huddersfield Chronicle</jn>
6   <titleAbbreviation>HUCE</titleAbbreviation>
7   <ba>2FWZ</ba>
8   <rtag>BLD</rtag>
9   <module>19th Century British Library Newspapers: Part II</module>
10  <is1</is1>
11  <printedDate>April 06, 1850</printedDate>
12  <da>April 06, 1850</da>
13  <pf>18500406</pf>
14  <dw>Saturday</dw>
15  <lp>8</lp>
16  <cp>Copyright ©x00A9; The British Library Board</cp>
17  <dc:Language>eng</dc:Language>
18  <newspaperId>NCBL0089</newspaperId>
19  <sourceLibrary>British Library</sourceLibrary>
20  <citation>
21    <titleGroup>
22      <marcTitle>The Huddersfield Chronicle and West Yorkshire Advertiser</marcTitle>
23    </titleGroup>
24    <datesOfPublication></datesOfPublication>
25    <typeOfPublication>Newspaper</typeOfPublication>
26    <conversionCredit></conversionCredit>
27  </citation>
28  <article>
29    <article zc="3210185924">
30      <id>HUCE-1850-04-06-0001-001</id>
31      <ocr relevant="yes">46.40</ocr>
32      <sc>A</sc>
33      <pi pgpref="1">HUCE-1850-04-06-0001</pi>
34      <ci pgpref="1" clip="1">HUCE-1850-04-06-0001-001</ci>
35      <pa>[1]</pa>
36      <pc>1</pc>
37      <ti>The Huddersfield Chronicle AND WEST YORKSHIRE ADVERTISER</ti>
38      <ct>News</ct>
39      <text>
40        <text.title>
41          <pg pgpref="1" clippref="1" pos="96,375,5096,1027"/>

```

Figure 3. Initial XML of the front page of the *Huddersfield Chronicle*, April 6, 1850, in *19th-Century British Library Newspapers*, part II.

variously sold and reconfigured companies, links big data to microfilming and micropublishing projects from the 1920s onwards, and blends labor practices from library acquisitions to the technical work outsourced to the global economy.

We should look at Gale's hard drives as relatively recent artifacts of that transmission or as a visual metaphor for the black boxing of commercial workflows, which supply so much of our digitized historical resources. Rather than focusing on whether the data they contain is either intentionally structured or messy, we should study them as residues of procedures, technologies, and decisions that the data does not sufficiently disclose. This record may be difficult to access. It may have been intentionally erased. Yet it remains crucially important for our work as scholars. Critics of big data used for cultural analytics have pointed out the importance of metadata.¹² But the legacy and functionality of digital scholarly resources also deeply depend on something else—"paradata." Loosely defined, "paradata" refers to the procedural contexts, workflows, and intellectual capital generated by groups throughout a project's life cycle.¹³ Paradata might include commentary, rationale, process notes, and records of decisions about projects recording what its participants chose to include or exclude. As Tim

Sherratt elegantly explains, “Big data is made up of many small acts of living.”¹⁴ Those acts are imperfect, ephemeral, and (particularly in cases of big scholarly data) embedded in a complex corporate world often careless of its history and jealous of its secrets.

This essay, then, concerns the largely invisible corporate histories of digital scholarly resources and the question of how (or even if) we might recover, reconstruct, and interpret them. To study the contingencies of remediated scholarly resources is, in a way, to engage with what Matthew Kirschenbaum calls the “.txtual condition,” a critical stance that prompts us to seriously consider the platform dependencies of electronic texts.¹⁵ Importantly, as Kirschenbaum has shown in *Mechanisms*, these dependencies are inextricable from the material circumstances of digital media, which require us to take a forensic approach to digital materiality and to collaborate with well-trained archivists in the process.¹⁶ For humanities scholars, critical attention to media processes and material histories has developed into “media archaeology,” an academic domain that is difficult to define but can perhaps be summed up as an attention to materialities and the different notions of history that they embody. Media archaeology attempts to distinguish itself from media history by exchanging chronologies for historical strata and exploring the deep, plural temporalities in which media become manifest and construct their newness. In this context, a project in “data mining” shifts to “data archaeology,” changing metaphors from the extraction of meaning toward trying to recover and reconstitute media objects within their changing ecologies. In her brilliant examination of *Early English Books Online*, Bonnie Mak proposes one method we might adopt in our study of digitized nineteenth-century resources: “An archaeology of a digitisation . . . should understand the digitally-encoded entity as a cultural object, produced by human labour, and necessarily shaped by—and consequently embodying—historical circumstance.”¹⁷ So, rather than doing book history with a database of Victorian newspapers, a researcher might do book history on the database itself, looking to its anterior forms, along with their social circumstances and mechanisms. Wolfgang Ernst prefers the term “archaeography,” which acknowledges the importance of machines in the historiography that Mak attributes to human actors and cultural circumstances.¹⁸ Somewhere on this spectrum, though, we must reckon with the archaeology of digitized scholarly resources. Victorian studies needs a neo-formalist history of the digital present, a historiography of Victorian data.

Jim Mussell knows exactly how crucial and difficult this problem is, especially for digital materials created or distributed by commercial vendors:

At present, publishers do not see the value in documenting their methodology (or at least making it available), nor do they provide accessible histories of the content they republish beyond introductory essays about the source material. There is little or no information about where this material is from (which archives?), what has been omitted (multiple editions? supplements?), any intermediary forms (microfilm?), let alone the various transformations that underpin the production of the image and metadata delivered over the web. Given the stake scholars have in this digital material and the fact that it will be used, in one form or another, as it is republished in new resources into the foreseeable future, it is vital that we can account for its history.¹⁹

Mussell underscores the importance of simply knowing what you are dealing with. I might ask of my own project: On what basis can I use data from *British Nineteenth-Century Newspapers* to make claims about the nineteenth-century press? From one perspective, this is a question about the evidentiary status of a digital collection or the validity of quantitative research. But from another perspective, it becomes a question about new opportunities for critical bibliography and cultural studies, in line with the compelling case studies furnished by scholars such as Lisa Gitelman, Bonnie Mak, Ryan Cordell, and Natalia Cecire.²⁰ To track the witnesses of Victorian media through the twentieth century is to approach our digital scholarly resources neither as historical surrogates nor as remediated objects but as potentially distinctive works.

In his initial bibliographic analysis of digitized nineteenth-century US newspapers in the *Chronicling America* collection, Cordell argues that such remediations can be properly conceived as new editions and that “acknowledging digitised historical texts as new editions is an important step . . . to developing media-specific approaches to the digital that more effectively exploit its affordances; more responsibly represent the material, social, and economic circumstances of its production; and more carefully delineate with its limitations.”²¹ This is as much a practical as an intellectual challenge, as scholars tend not to “[acknowledge] digitised historical texts” in published research. Gale publisher Ray Abruzzi and historian Tim Hitchcock have each complained, for different reasons, that academics prefer to cite primary sources and mask the uses of digital collections as unscholarly intermediaries.²² But these collections are distinct objects with histories of their own. To analyze their intermediation is, as Paul Eggert puts it, “to understand the text’s successive discursive inscriptions, its ideological absorbency.”²³ In the case of nineteenth-century British newspapers, their discursive absorptions have included twentieth-century political economies of global conflict, the intelligence community’s alliances with scholarly associations and research libraries, gendered and outsourced

labor, and commercial techno-futurism. Exploring these histories takes critical bibliography to another remove, requiring a method constituted at the intersections of book history, media archaeology, and investigative scholarly journalism.

From Accession to Micropublishing

This section pursues such a method on the hardware and data objects that arrived on my desk in early 2015 as “19th Century British Library Newspapers.” It seeks to offer a glimpse at the historical contingencies that shaped the passage of such materials from the nineteenth century to the present moment. This is not meant to be an exhaustive survey but a starting point for how we might study Victorian periodicals against the *longue durée* of their mediation—and an argument for why we might need to do so. That time frame includes roughly four phases in the life of a newspaper: its initial production, accession into library volumes, remediation through micropublishing, and digitization. These phases do not necessarily follow discrete moments of change, nor do they offer chronological scenes in which to detect traces of some durable information commodity. Instead, they mark domains of reproduction, more properly conceived as editorial frameworks for the creation of new archives, each activated at a different moment yet still impinging on the others. Nonetheless, we can put them into a simplified historiographical narrative to better understand their relations. Newspapers were printed. Some copies were accessioned. Some of those were microfilmed. Some of those were digitized. Some of those were included in the *British Nineteenth-Century Newspapers* database, a derivative of which was mailed to the NCSU Libraries. All the rest is complicated.

The production of nineteenth-century newspapers is well documented in histories and cultural studies of the press and continues to interest scholars of Victorian periodicals. I will pass over it here to focus on the transmission of newspapers through the evolving architectures of scholarly resources, beginning with the accessioning of newsprint by the British Museum in 1822. That date marks the first systematic and institutional attempt to collect British newspapers as such. At that time, the British Museum negotiated a deal with the Board of Commissioners for Stamps (which later amalgamated with the Commissioners of Inland Revenue). The Stamp Office oversaw the longstanding “taxes on knowledge,” which, especially after duties increased on periodicals in 1819, profoundly shaped the landscape of newspapers, including what it passed on to the British Museum.²⁴ The Stamp Office kept copies of stamped newspapers for two years in case they were needed as legal evidence, especially in cases of libel.

At the request of the principal librarian, the Stamp Office would “gift” these copies to the museum, a practice that continued until 1869. Already shaped by the political implications of the newspaper tax, these collections were often incomplete and did not include provincial newspapers until 1832. Meanwhile, Irish and Scottish titles were not regularly acquired until after 1849.²⁵

While books subject to the laws of copyright were deposited into the British Museum, this was not enforced for newspapers until 1869, when a different agreement was reached. Publishers would now be required to deposit newspapers within seven days of printing. The Inland Revenue Office would pay for them, keep them for the legal statute of three years, and then deliver them to the British Museum. These papers often show hand annotations in pencil, suggesting they included editorial proofs or even publishers’ in-house “receipts” confirming payment for advertisements and some local news.²⁶ The ongoing gifts and legal deposits over the nineteenth century were supplemented by donations and the British Museum’s purchase of sets to cover gaps in the record. But all of these consolidating efforts were subject to the pressures of storage space and the expense of binding newspapers into volumes (figure 4). The museum’s accessioning policies changed as a result. For example, beginning in 1879, the museum collected only the first edition of any given newspaper and for a brief time in the 1880s only bound the “most used newspapers.”²⁷ Each of these initiatives favored the metropolitan press: library usage statistics from 1896 show 3,000 provincial papers consulted as opposed to 40,000 metropolitan titles, with nearly all newspapers subpoenaed for legal use coming from London. The space issue prompted even more drastic proposals, including handing over the entire collections enterprise to another institution, redistributing the provincial collections to “local bodies” (which did not even exist), and simply disposing of provincial papers.²⁸

In 1905, the storage crisis resulted in the construction of the museum’s newspaper repository at Colindale (which was filled within twenty years) and the subsequent construction of the British Museum Newspaper Library in 1932.²⁹ These were also formative years in the development of microphotography as a storage solution and preservation medium. But microfilm emerged in a significantly different context. Its application to nineteenth-century newspapers at Colindale entwines the British Museum’s history with commercial and strategic intelligence interests in microfilming more generally. In the early part of the century, these efforts were significantly focused by the work of companies such as Eastman Kodak and University Microfilms Incorporated (UMI), whose cultural preservation efforts were, in turn, mandated and funded by the interests of American cultural elites and intelligence services. The history of UMI (now a part of ProQuest)



Figure 4. “Newspaper library #3,” August 14, 2013, photo by Luke McKernan, Flickr, <https://www.flickr.com>.

provides a useful starting place for investigating the complicated relationships that drove such efforts. While UMI may be more familiar to scholars as a vehicle for consuming and archiving dissertations, its history entangles with Gale’s *British Nineteenth-Century Newspapers* database through its extensive connections to twentieth-century microphotography efforts at the British Museum’s newspaper archive.

An American company, UMI began in 1938 when its founder, Eugene Power, began photographing rare materials from the Clements Library at the University of Michigan. Power was inspired by a 35 millimeter camera developed by Captain R. H. Draeger, an officer in the US Navy, who had devised the technology as a means to photograph books to read while on assignment in China. While abroad, Draeger simply reprinted the positives from the film and enjoyed the same sort of transportable library which today energizes the marketing of Kindles and other eBook readers. Back in Ann Arbor, Eugene Power cobbled together “parts of two movie and still cameras into the second microfilm-book camera in existence.”³⁰ His first office comprised two rooms in the back of a funeral home where he filmed old books from the English Short Title Catalog, which would eventually become part of *Early English Books Online*, now owned by ProQuest (figure 5).³¹ In his autobiography, Power tells the origin story, which reads like



Figure 5. Postcard of the Majestic Theater and Dolph's Funeral Home on Maynard Street, Ann Arbor, 1929. Dolph's Funeral Home Records, Bentley Historical Library, University of Michigan.

the incarnation scene from *Frankenstein*: “I can see that those nights of feverish work amid the caskets and embalming odors of Dolph's Funeral Parlor were the real beginning of University Microfilms.”³² Out of old books made of animal skins and rags, something is born anew, a tortured acetate body for the eloquence of Enlightenment texts. Though Power characterizes UMI as his own Promethean invention, his work was made possible by a unique confederation of library interests, commercial micro-publishing, and intelligence services at the outset of World War II. It was, according to Seth Cayley, now Director of Research Publishing at Gale, a massively coordinated effort that would never again be possible. This point may be disputed, especially considering the phenomenon of Google Books. Nevertheless, it is an argument worth considering—that twentieth-century microfilms are not simply the accidental intermediaries for commercial digital objects but rather serve as their institutional precondition.

In 1940, an incendiary bomb destroyed a significant portion of the King George III's collection in the British Museum, whose missing contents are still indexed in the British Library's catalog.³³ That same year, a high-explosive bomb hit the museum's original 1905 newspaper repository building at Colindale, affecting 40 percent of its volumes of nineteenth-century English provincial newspapers, of which about 6,000 volumes were completely lost.³⁴ The specter of war damage and book burnings in Europe created a wave of interest in microfilming as a preservation strategy

that would distribute surrogate copies of newspapers in worldwide depositories. In the United States, institutions such as the Library of Congress, the American Council of Learned Societies, and the Council of Library and Information Resources formed committees which agreed that preservation of European and especially British materials was of the utmost importance. The committee generated a wish list from American scholars that featured rare books, manuscripts, and periodicals; it then secured funding from the Rockefeller Foundation to carry out the work abroad. In her remarkable study of these developments, Kathy Peiss argues that American preservation efforts during the war were motivated by a broadly shared sensitivity to an endangered European culture, as well as by the opportunism of cultural elites in shaping government policy.³⁵ As Lester Born would later describe such initiatives, “American scholars see the bringing of the resources of the Old World to the New as the principal role for scholarly photocopying.”³⁶

At the same time, the Office of the Coordinator of Information, soon to become the Office of Strategic Services and later the Central Intelligence Agency, had similar designs on the old world, especially German materials that might be microfilmed. Peiss argues that the “government’s need for intelligence had a greater impact on the fate of books than did the organizations whose mandate was cultural protection. The war brought librarians squarely into a relationship with the intelligence-gathering arm of the state through the Office of Strategic Services (OSS), as well as the intelligence units of the armed forces.”³⁷ This relationship not only included librarians but also the commercial agents who supplied the technology, expertise, and ultimately the saleable products of preserving cultural heritage which, in remediated forms, scholars continue to use today. The Office of the Coordinator of Information and the Office of Strategic Services backed Eugene Power on several trips to Europe for copying caches of documents and recent periodicals. For this work, Power turned to a different camera, the Kodak Model D Microfile, otherwise known as a Recordak 35mm. Figures 6 and 7 depict microfilm technology from later decades, but they fairly represent this machine and its predominantly female operators, who, like nineteenth-century “typewriters” and twentieth-century computing staff, are mostly absent from micrographic histories.³⁸ As Power describes it, the Microfile “had a large bed with a glass platen which would press bound newspapers flat and hold them in place to be photographed. Its cradle would shift back and forth in order to get a complete page on each exposure.”³⁹ The look of these setups is familiar to anyone working with a modern overhead scanner, but the Microfile’s technique of shifting its cradle may have informed the decision to identify the page—rather than the opening or two-page spread—as the fundamental unit of display and



Figure 6. "Microfilming with a Recordak 35mm," National Library of Medicine, ca. 1969, <https://ihm.nlm.nih.gov>.

analysis. Around 1941, Power shipped three Microfile cameras to England for on-site work, but the cargo ship was sunk by a German submarine.⁴⁰

Power would soon successfully develop relationships with the British Museum, Bodleian Library, and Cambridge University Library, installing his cameras and operators to generate the installment-based products which became *Early English Books Online* and similar collections for sale. Power and UMI also facilitated the micropublishing of British newspapers. During a visit to the United States in 1944, the director and principal librarian of the British Museum, Sir John Forsdyke, became convinced of the value of microphotography as a solution to the museum's escalating newspaper problem. The museum sponsored Forsdyke's visit to UMI in Ann Arbor where he consulted with Power about what equipment the museum would need. According to Power, he then sketched the microfilming facilities that Forsdyke would soon build.⁴¹ When the Colindale facility rebuilt temporary structures in 1950, it opened a microfilm annex with equipment the Rockefeller Foundation donated for the purpose. In 1948, the foundation also sponsored three-month fellowships for British Museum staff to train at UMI on large-scale microfilm production, training that was designed to prepare them for photographing the sprawling collections of British newspapers back in Colindale. These visiting staff included D. A. Wilson, who was in charge of the Colindale facilities and



Figure 7. “Microfilming at Colindale began in the 1950s,” Preservation Advisory Centre, British Library, <http://britishlibrary.typepad.co.uk>.

soon oversaw the four microfilm cameras from Rockefeller, one of which remained in use through the 1990s.⁴² Setting out on the monumental task of microfilming the collections, Wilson prioritized current British provincial newspapers, important sets of London newspapers, and then historical London newspapers in chronological order from 1801.⁴³

It is difficult to say how much of the British Museum’s extant print collection of newspapers was filmed. The estimates I have received suggest that the British Library newspapers on microfilm represent somewhere between 5 and 30 percent of its print collection. If, as Patrick Leary suggests, there is an “offline penumbra” of nineteenth-century materials that have not been digitized, that shadow lengthens with the “unfilmed penumbra” of those periodicals that remain in volume form.⁴⁴ Beginning with Wilson, the microfilming program necessarily had to make decisions about what to photograph based on the fragility of materials, the coverage of the collection, the interests of potential users, and the budget available for the project. Those decisions both did and did not have downstream impacts on the digitization programs to come. The complex question of what microfilm represents and excludes anticipates the very problematics of digitized periodicals resources.⁴⁵ Just as importantly, microfilm collections are also characterized by the erasure of the institutional memory that attends their transmission of a preserved past. In considering newspapers’ remediation

through micropublishing, I would emphasize thinking not so much about the formal properties of newspapers in print, microfilm, or digital format but about the vanishing institutional histories which condition their mediation in any form.

In his 1982 history of scholarly micropublishing, Alan Meckler explains how librarians did not know about the source histories of microfilm materials and neither, in many cases, did the companies which produced them: "Several, particularly those entrepreneurs who entered the field solely as a commercial venture, simply did not bother to keep papers that might have had some historic value. Others lost or misplaced papers during the course of their careers, especially as business changed hands."⁴⁶ For his book, Meckler had to undertake extensive interviews, tracking the oral histories of an apparently durable cultural heritage medium. A 1964 article in *PMLA*, later reprinted in the 1972 first volume of *Microform Review*, raises an important concern: "The history of scholarship is consequently in large measure a history of the diffusion of the materials for scholarly research."⁴⁷ This article—a report prepared for the American Council of Learned Societies—attempted to survey the landscape of scholarly micropublishing, concluding that the "plain truth is that we have accumulated scattered, incomplete, and almost random collections to which there are few guides and of which there is insufficient use."⁴⁸ Ironically, the article's author is actually missing from the metadata in JSTOR and miscataloged in the MLA International Bibliography. In fact, he was Lester Born, a classics professor and archivist, who in 1950 would coordinate the microfilm holdings at the Library of Congress. He was also a soldier in the Allied Command's special Monuments, Fine Arts, and Archives section, known as the "Monuments Men," who worked in wartime Europe to secure historic buildings, collections, and books.⁴⁹

From Scanning to Global IT

From one perspective, we are a long way in decades and context from nineteenth-century newspapers. Yet, for today's scholars, nineteenth-century newspapers only exist because of the elaborate transmission history in the intervening years between then and now. With a nod to library accession volumes, Laurel Brake has characterized the "ephemera" of Victorian periodicals as being much more durable than we typically think.⁵⁰ But as media studies scholar Wendy Chun clarifies, the "enduring ephemeral" is rooted in how mediums constitute themselves as "new," promising durable storage that belies the contingencies of memory.⁵¹ Memory—and digital memory especially—is not storage but an active process of constantly refreshed electronic charges as well as a continuum of institutional decisions about



Figure 8. “So farewell, then, British Library newspaper archive Colindale,” March 27, 2015, photo by Clare Newsome, Twitter, <https://twitter.com/ClareNewsome>.

preservation, transmission, and retrieval. For nineteenth-century newspapers, understanding the relatively recent shift from microfilm to digitization involves just such complexities, in which the military-informational complex cedes to the interests of a global IT economy which makes digital Victorian periodicals scholarship possible.

The British Museum—whose library departments became the British Library in 1973—ended its microfilming program with the demolition of the Colindale facility in 2010 to make way for real estate development (figure 8). When it closed, the facilities at Colindale held about thirty miles of shelved periodicals with over 750 million pages of newspapers, materials that were gradually relocated to the British Library’s state-of-the-art facility in Boston Spa, with its dark, low-oxygen storage stacks which one staffer recently called the “void.”⁵² The British Library’s plans for digitizing its newspaper collections began with a grant application to the Joint Information Systems Committee (JISC), whose funding competition already set the terms for the proposed collection: it had to be large scale, include significant geographical coverage, and be broadly useful. The proposed British Newspapers 1800–1900 project had an initial target of 2 million pages (about 0.3 percent of the print collection). An advisory group of library staff and

scholars was assembled to establish a “framework of national titles and countrywide coverage with the breadth and depth to form a virtual key to many other provincial newspapers of the same date.”⁵³ In this description, the forty-eight selected titles represent a “virtual key,” an indexical link or representative sample. The funding period aimed to deliver the “scanning of the entire microfilmed content; article zoning and page extraction; OCR of the page images; and the production of the required metadata.”⁵⁴ This part of the story is admirably documented by project managers and department heads at the British Library, including Jane Shaw and Ed King, who continually shared process lessons about large-scale newspaper digitization with the library community in the 2000s.⁵⁵ Their published conference papers include an assessment of the image quality and material durability of microfilm slated to become digital facsimiles. Decades of microfilm had been developed on acetate rolls, subject to acidic decay, before microfilm preservation standards changed to more stable polyester substrates. Shaw observed that the collection of nineteenth-century newspapers on microfilm was actually in relatively good shape, with only 2 percent unfit for the library’s new Zeuschel microfilm cameras, which used large book cradles and page de-skewing software. Interestingly, the British Library decided to improve its speed and image consistency by, whenever necessary, making *new* microfilms from print copies, especially to update acetate-based rolls with National Preservation Standard microfilm. Estimates rose from 50 percent to 90 percent for new filming over the course of the project. In other words, microfilm was not simply a historical intermediary between print and digitization; it was the immediate step in contemporary digitization practices created for the making of new media. The British Library would later recommend direct scans from original sources, but in the 2000s digital cameras did not yet supply sufficient megapixel capture and, considering the heavy bound volumes of its newspapers, flat-bed scanning was potentially damaging as well as prohibitively expensive. Thus, to avoid “gutter shadow” and to keep pages evenly lit, newspapers were held in book cradles with each page photographed to microfilm.⁵⁶ The single page remained the primary unit of image production.

The British Library’s newspaper digitization work occurred in two distinct phases, each of which corresponds to a segment of Gale’s collection. The first phase—which would become known internally as “JISC I,” also known as Gale’s “Part I”—occurred from 2004 to 2007. A second funding period followed in 2008–9, “JISC II,” Gale’s “Part II,” which expanded the British Newspapers 1800–1900 project to include more regional and local news, as well as extending existing titles back through the eighteenth century. This project added a million more pages to the digital collection, though it slightly changed the workflows for the project. Both phases used



aside, and asked him if he wanted
 Witness then noticed that the ot
 for the door, and, missing his t
 placed on the top of his bundle c
 hold of the prisoner. The oth
 Martin was given into the cust
 The telescope was not recover
 robbing the prosecutor, and Mr. I
 was not sufficient evidence to
 magistrate) advised Coomber to g
 if the police recovered the teles
 municated with.

The Sunday Band

At Lambeth Police-court Henr
 Reynolds, both living at East D
 for refusing to cease playing cert
 of a house after being request
 away out of a Salvation Army

Figure 9. Detail of a hybrid service image from JISC I showing greyscale illustrations and bitonal text, *Illustrated Police News*, January 28, 1893, *19th-Century British Library Newspapers*.

scans predominantly from new microfilm, though in JISC II one complete paper—the *Standard*—was directly scanned from print at Boston Spa. For each phase, the British Library sent its in-house scanned images or new microfilm reels to external vendors for digital scanning and processing. Scans from the films, new as well as old, would result in multiple digital objects for each newspaper page. In JISC I, this included an “archival master file for each page, in TIFF format, version 6.0 . . . at a resolution of 300 dpi, 8-bit greyscale,” as well as service images generated “after the process of article zoning and OCR . . . [and] delivered as greyscale hybrids.”⁵⁷ The “hybrid” images combine 1-bit bitonal scans (for example, simple black and white) of textual components with 8-bit greyscale scans of any illustrated content, all to conserve on the eventual file size and to avoid “problems for 56k modem users” (see figure 9).⁵⁸ In JISC II, the British Library changed its requirements to greyscale scans at 400 dpi. It received TIFF images of raw scans (one per page, unedited), a lossless JP2 or JPEG2000 master image (cropped and lightly corrected), a compressed JPEG derivative image for service copies, and associated XML files.⁵⁹ Images from JISC II also look a little different from their earlier counterparts, as “many of the local newspapers are in poorer condition with uneven printing within a run and across individual pages. . . . This has resulted in deliberately chosen lighter looking scanned images in order to improve the OCR word accuracy.”⁶⁰ In addition to the British Library’s documentation, the digital files can tell their own stories. Ryan Cordell has demonstrated how EXIF metadata extraction software can help reveal the conditions under which such images were produced.⁶¹ However, that depends on access to the archival master files, which are often distinct from a derivative master,

or “mezzanine copy,” of the improved image from which other service files are made.⁶² Running EXIF software against the derivatives of JPEG and TIFF files on Gale’s hard drives only reveals the *absence* of evidence for their production. Indeed, as the British Library documents, the initial TIFF images from microfilm scans were “destroyed at the end of production,” being far too large for the library to ingest in its own long-term object management system.⁶³ Even the masters are already derivatives.

Remediation is not a one-way street, however. As Ed King explains, “One of the most fascinating aspects of ‘first time’ large scale digitisation of 19th century newspapers was how little librarians and archivists in the UK knew about the *actual* printing run of any given newspaper in the UK.”⁶⁴ As the British Library logged pages for newspapers microfilmed and scanned for JISC I and JISC II, they were often improving the library’s catalog information about those very materials, whose existence in print was frequently unknown or otherwise complicated by variously timed or regionally distributed editions. (When faced with duplicates or multiple editions, the JISC projects selected the last timed edition of a paper for digitization.) Furthermore, the JISC projects included physically stabilizing and repairing original newspapers when necessary, aiding their material preservation in the service of getting good pictures. Thus, digitization ironically regenerates the memory of nineteenth-century newspapers as preserved in other storage media, including print as well as new microfilm.

The British Library employed several third-party vendors for scanning microfilm and processing digital images. This work included segmenting the pages into articles; creating OCR of the text; encoding page divisions, content, and metadata description into a standardized XML schema; and then delivering service images and XML files back to the library. The specific involvement of these companies is difficult to trace because they do not have records of their own histories, they ascribe to contractual non-disclosure rules, and they typically avoid discussions of outsourcing. For all the British Library’s admirable transparency, its relations to third-party vendors can be vague: “The allocation of work between in-house operations and third parties is based on where value can best be achieved, balancing the cost effectiveness of competitive tender with the optimum deployment of experience and expertise from the library.”⁶⁵ The corporate language exemplifies the abstract characterizations of labor and costs which render invisible and palatable the conditions of outsourced work, especially in global contexts. Having taken the “view that the use of human intelligence combined with software applications would give the best quality result,” the British Library chose a company called Apex CoVantage, which the metadata in the XML files for JISC I (Gale Part I) still names in its <conversionCredit> element.⁶⁶ In snapshots of its website

from 2006, Apex celebrates its “dual-shore advantage” and “truly global solutions,” boasting that it possesses “one of the lowest employee turnover rates in the outsourcing industry.”⁶⁷ Nineteenth-century newspapers were processed through a digitization product called “isaac” for which Apex still furnishes a corporate video. The video, which includes the British Library’s logo at its conclusion, begins with the promise of “unlocking [the] treasures” of historic newspapers, creating cultures of tolerance and peace, offering rich resources for libraries, and increasing revenue streams for content providers.⁶⁸ The video illustrates the digitization workflow, highlighting the software’s unique processes while also clarifying (explicitly and implicitly) its uses of human labor. The gentle voice-over intones, “Between these steps, a human insures that article boundaries are accurate, adjusting them when necessary.” “Article components,” it continues, “are then sent through different human-driven workflows.” In the background, the video offers glimpses of the “human-driven workflows” as brown-skinned men work at computer terminals in cubicles, cleaning “dirty” materials and validating software suggestions for eventual reassembly and sale in the Anglophone West. Natalia Cecire has pointed to the enforced invisibility of marginalized labor in the scanning operations of the Google Books project, which are occasionally glimpsed as the accidental photography of workers’ hands and fingers.⁶⁹ Bonnie Mak has recently argued for a long history of exploitative transcription practices which connect early modern scriptoria to offshore companies currently supplying cheap transcriptions for the *Eighteenth-Century Collections Online*–Text Creation Partnership project.⁷⁰ At the very least, Mak suggests, we need to be aware of the enabling conditions of our scholarship, which may increasingly rely upon the segmentation and relative invisibility of global labor practices in digitizing historical materials.

The British Library switched vendors for JISC II, choosing Olive Software, a company based in Santa Clara, California, with a research development office in Israel.⁷¹ However, during an early pilot to assure quality, the company failed to produce scans of a similar or better quality than the British Library. It was soon replaced with the runner-up, Content Conversion Specialists. Although the company is headquartered in Hamburg, Germany, it subcontracts its large digitization projects with companies such as Digital Divide Data, then located in Cambodia.⁷² This subcontractor claims to practice socially responsible “impact sourcing,” hiring disadvantaged high school graduates as workers and offering them opportunities for higher education (after one year of employment). None of this is recorded in the XML metadata for JISC II files.

For both of its digitization phases, the British Library needed Gale to assume responsibility for transforming, hosting, and serving data files on

an accessible website. The library initially planned to create its own web services and interface but soon realized that this would be unfeasible. And here—finally—the sharper outlines of *British Nineteenth-Century Newspapers* begin to come into focus. Sometimes building on and sometimes overwriting legacy codes and processes, Gale's product was based on procedures developed at its corporate offices in the United Kingdom, with skilled labor outsourced to India and web development services located at Gale's US headquarters in Farmington, Michigan. Gale further subcontracted work to the company HTC Global Services (also uncredited in the metadata), with its team of 400 people in Chennai, India. According to Gale, these were English-speaking workers with computer science backgrounds who ran the Abby FineReader OCR software, entered and verified metadata, visually mapped distinct articles, and validated the XML. Launched in October 2007, the resulting product was called *19th-Century British Library Newspapers*.

Jane Shaw describes the project as an “innovative and challenging example of a public/private partnership between Gale Cengage Learning, CCS and the British Library,” each of which has its own “cultural emphases.”⁷³ Indeed, Gale's emphases would shape the project in important ways.⁷⁴ Shaw notes that the British Library had already developed its own standards for digitizing microfilm in the face of inconsistencies in similar projects worldwide. A project manager at Gale characterizes the early 2000s as a “wild west” where projects and standards proliferated, and another called it the “gold rush to digital.”⁷⁵ In either case, Gale was soon using its own workflows and proposing additional requirements such as using subject categories for articles (with twenty-six options).⁷⁶ These decisions are consolidated in a master file called the “document type definition,” a set of rules that all of the project's XML files must obey. In simpler terms, it constitutes the set of editorially accepted categories for which subjects, genres, and features the digital archive will record. Establishing these parameters was apparently the most contentious aspect of the project, as developers and scholars attempted to model the staggering heterogeneity of formats and content across a century of periodical publishing. The British Library project managers made sure that the project remained in touch with the Metadata Encoding and Transmission Standard (METS) and Dublin Core standards, including four structural levels: title, issue, page, and article.⁷⁷ For its part, Gale was following the procedures it developed for its recently completed *Eighteenth-Century Collections Online* and, more directly, the *Times Digital Archive*. These near ancestors may have passed on their genetic materials. Andrew Hobbs has recently criticized scholars for reflexively gravitating to the *Times* and overlooking the provincial press.⁷⁸ Not coincidentally, the *Times* was the first British newspaper made

commercially available as a digitized scholarly resource.⁷⁹ Furthermore, the same workflows and article categories Gale developed for the *Times Digital Archive* quickly came to structure the experience of *19th-Century British Library Newspapers* as a whole. Scholars of the Victorian press have generally noted how metropolitan news was copied throughout the country. Gale's database has cemented that legacy at the level of its document type definition, against which all of its XML must be validated.

For each "work product," Gale took a snapshot of their file production and backed it up. We can see from these file histories that the copy of the data we received at North Carolina State University was derived from 2007 (part I), and 2008–9 (part II).⁸⁰ Gale also uses an "iron mountain" backup service called Portico, a non-profit affiliated with the Library of Congress and JSTOR, which will maintain Gale's data even if the company goes bust. But which version will be preserved? In 2008, in the midst of my graduate research, I serendipitously took a couple of screenshots of my access to the collection, then titled *19th-Century British Library Newspapers*, which featured the British Library's icon in the upper-left-hand corner. At the time, I was quite interested in Victorian accidents, the subject of my first book, so I began browsing the 18,524 results returned for a simple keyword search. Run in early 2015, that same search generated 1,145,071 results. Scholars such as Charles Upchurch have alerted us to pay attention to the versioning of such sources, which can be silently and periodically updated.⁸¹ Even within a given dataset, researchers ought to verify coverage. Bob Nicholson has demonstrated how Gale's content swells towards the second half of the nineteenth century and sometimes contains significant gaps within any given issue's run.⁸² These challenges of "bibliographic control" further expand when scholars are using different commercial or institutional versions of particular databases. Jim Mussell identifies "Gale Cengage's *19th Century UK Periodicals* (2007), Gale Cengage and the British Library's *British Newspapers, 1800–1900* (2007) (also known as *19th-Century British Library Newspapers*), ProQuest's *Historical Newspapers* (2001–) and *British Periodicals* (2007), and Brightsolid's *British Newspaper Archive* (2011–)" as all constituting a variable corpus of digitized periodicals that researchers, depending on their level of access, might utilize in their scholarship.⁸³

In 2010, when the British Library began to expand its newspaper digitization project once again, Gale lost the bid to a genealogy company called Findmypast (along with its hosting subsidiary Brightsolid), which now owns the rights and operations to the renamed and still expanding *British Newspaper Archive*. Genealogy companies are currently among the most aggressive commercial players in the digitization of historical newspapers.⁸⁴ However, the specific reasons for the British Library's decision

to change vendors were not disclosed. Having lost the rights, Gale had to rename its product *British Nineteenth-Century Newspapers*. It now packages this in four separate parts to make it saleable to libraries with differently sized budgets. According to Ray Abruzzi, then vice-president and publisher of Gale's digital collections, the company tries to "right size" its products.⁸⁵ Gale could create a five-million-page database, but it would be too expensive for anyone to afford. So we get two million pages as a way of balancing Gale's production costs against library acquisitions budgets, which are themselves shaped by a library's own sense of how much coverage is adequate or representative. In its partnership with the British Library, Gale consulted with academics and expert advisors on what to include in its databases. To their great credit, Gale's managers are remarkably accessible to scholars. Abruzzi was recently interviewed for the journal *C19: Interdisciplinary Studies in the Long Nineteenth Century*.⁸⁶ Seth Cayley, in the United Kingdom, has published in *Victorian Periodicals Review* and regularly attends the Research Society for Victorian Periodicals conference.⁸⁷ Compelled by the scholars he has encountered, Cayley has provided process histories and perspectives for some of Gale's collections, such as his terrific essay on the *Daily Mail* archive.⁸⁸

That kind of valuable paradata does not exist for nineteenth-century newspapers. Gale does not have its own archivist or historian. On the phone, Cayley and Abruzzi were both candid about Gale's lack of institutional memory. They pointed out the irony that a company based on archives and history so often lacks its own. Much of this record exists only by word of mouth, and staff sometimes leave. In fact, Abruzzi himself left Gale in the interval since this article was first submitted for review. An emeritus worker could take on the job of company historian in an unpaid capacity, but the corporate masters see this as a luxury. There is no obvious place for Gale to store this institutional memory and no financial benefit in doing so. In early 2015, the online "corporate history" page for Gale Cengage had a broken style sheet and a non-functional timeline (figure 10). As of early 2016, that history page no longer exists. Even the British Library's history can be surprisingly contingent. In the introduction to his massive *History of the British Museum Library*, Philip Rowland Harris explains that he began the project because he was concerned that common knowledge would be lost when the library moved to St. Pancras: "The way in which the library operated for over two centuries should be recorded before memories fade."⁸⁹ Now-retired British Library employee Ed King, without whose extensive help this paper could not have been written, pointed me to the British Library's internal reports for JISC about *British Newspapers, 1800–1900* which are now only web-accessible through PDFs on the *Internet Archive*.⁹⁰

Progress and partnership: Gale history

The individual strengths of three legacy companies — Gale, Primary Source Media and Information Access Company — were combined in 1998 to produce the wide-reaching and robust company that today is known as Gale, part of Cengage Learning.

Roll over the blue boxes on the timeline below to learn more about Gale's history of progress and partnership.



Figure 10. “Progress and Partnership: Gale History,” screenshot taken April 1, 2015, *Gale Cengage Learning*, <http://www.gale.cengage.com/publishers/history.htm>.

Much of this essay has relied on the *Internet Archive* and its Wayback Machine to glimpse the near history of the web. If corporate records and project paradata are already elusive, their status online may be even more fugitive—just as “ethereal, ephemeral, unstable, and unreliable” as perhaps the web itself.⁹¹ Yet though the *Internet Archive* seems like a tenuous and contingent solution to web preservation, it adds another chapter to the unfolding story of our data’s continuous regeneration. The British Library has itself adapted the open-source web-collection technology that drives the *Internet Archive* and, since 2013, has extended legal-deposit law to include everything posted to the United Kingdom’s web domain that resembles a publication. The historiography of nineteenth-century newspapers, including the digitization reports missing from the British Library’s own site, now entangles with the formidable problems and necessary partnerships needed to sustain new media online.

In one of those documents, the “Final Report” from JISC II, Jane Shaw asks some compelling questions about digital preservation: “What are we trying to sustain/preserve? Is it the exact project outputs, the digital experience, the digital skills, knowledge transfer, or the website?”⁹² Such questions pertain to any moment of shifting media. However answered or deferred, they shape the legacy of what we refer to as historical materials. This essay offers just a glimpse of the long legacy too easily called “digitization.” There remain glaring holes in this story—not just the history of the digital archive but the history of its production. And still the question remains: What can be done with this data? At North Carolina State University, we are beginning the work officially known as “analytics.” But inherited data prompts other forms of analysis and storytelling, which trained literary critics or cultural analysts may be uniquely suited to undertake. The simple fact is that no one else will.

North Carolina State University

NOTES

An early version of this essay was delivered at the CUNY Annual Victorian Conference in May 2015. I am deeply grateful for the email exchanges and telephone conversations in which the following persons shared histories, resources, and expertise: Ed King, retired head of the British Library's newspaper collection; Seth Cayley and Ray Abruzzi at Gale; and Markus Wust, Brian Dietz, and Jason Groth of the NCSU Libraries. This essay takes particular inspiration from Bonnie Mak and Ryan Cordell, who have pioneered an archaeological approach to the study of digitized historical materials. Any mistakes or misrepresentations in this complicated story are my own.

1. Darnton, "Library in the New Age."
2. Examples of such promising transhistorical work include Mussell's *Nineteenth-Century Press in the Digital Age* and Alfano and Stauffer's *Virtual Victorians*.
3. Gitelman, *Always Already New*; Chun, "Enduring Ephemeral."
4. Hiscoe, "NCSU Libraries."
5. Massari, "Gale Leads to Advance Academic Research."
6. In Victorian periodicals studies, precedents for working with commercial data include Gibbs and Cohen, "Conversation with Data"; Liddle, "Reflections on 20,000 Victorian Newspapers"; and Pionke, "Excavating Victorian Cuba."
7. I do not mean to mischaracterize this collection as being error prone. Front pages with advertisements offer special challenges for OCR software. For a thorough discussion of these processes and the historical reliability of OCR, see Milligan, "Illusionary Order."
8. Gitelman and Jackson, "Introduction."
9. Drucker, "Humanities Approaches to Graphical Display," 6.
10. For a glimpse of Gale's plans for data access and digital humanities research portals, see Abruzzi, Calè, and Vadillo, "Gale Digital Collections."
11. Mussell, *Nineteenth-Century Press in the Digital Age*, 152.
12. For a classic glimpse of this, see Nunberg, "Google Books."
13. Baker, Bentkowska-Kafel, and Denard, *Paradata and Transparency*.
14. Sherratt, "Stories for Machines."
15. Kirschenbaum, "The .txtual Condition."
16. Kirschenbaum, *Mechanisms*.
17. Mak, "Archaeology of a Digitisation," 26.
18. Ernst, *Digital Memory and the Archive*.
19. Mussell, "Teaching Nineteenth-Century Periodicals," 205–6.
20. Gitelman, *Paper Knowledge*; Cecire, "Ways of Not Reading Gertrude Stein."

21. Cordell, "Q I-Jtb the Raven." Laurel Brake has made a similar claim—that the "digital edition is a copy of a version that is unique among many possible editions." See "London Letter," 248.
22. Abruzzi, Calè, and Vadillo, "Gale Digital Collections"; Hitchcock, "Academic History Writing."
23. Eggert, *Securing the Past*, 157.
24. Weiner, "Newspaper Taxes."
25. King, "Legal Deposit of Newspapers," 161; Harris, *History of the British Museum Library*, 31, 70, 131, 271.
26. Shaw, "British Newspapers," 31n37.
27. Harris, *History of the British Museum Library*, 276, 345.
28. *Ibid.*, 376.
29. Gillies, "History of British Library Newspapers." For an overview of the storage crisis and building needs, see Harris, *History of the British Museum Library*, 376–78.
30. Power, *Edition of One*, 29.
31. For detailed studies of the emergence of *Early English Books Online*, see Mak, "Archaeology of a Digitisation"; and Gadd, "Use and Misuse of *Early English Books Online*."
32. Power, *Edition of One*, 31.
33. For the story, see Edwards, "Destroyed, Damaged and Replaced."
34. Harris, *History of the British Museum Library*, 555.
35. Peiss, "Cultural Policy in a Time of War."
36. Born, "Planning for Scholarly Photocopying," 86.
37. Peiss, "Cultural Policy in a Time of War," 377.
38. See Mak, "Archaeology of a Digitisation," 15; Marvin, *When Old Technologies Were New*; and Terras, "Father Busa's Female Punch Card Operatives."
39. Power, *Edition of One*, 112.
40. *Ibid.*, 128.
41. *Ibid.*, 154.
42. Ed King, personal communication.
43. Harris, *History of the British Museum Library*, 601–2.
44. Leary, "Googling the Victorians."
45. For a good overview, see Brake, "Half Full and Half Empty."
46. Meckler, *Micropublishing*, xii–xiii.
47. Born, "Planning for Scholarly Photocopying," 77.
48. *Ibid.*, 86. For another institutional history, see Bourke, "Scholarly Micropublishing."
49. Bradsher, "Archivist Monuments Man"; Peiss, "Cultural Policy in a Time of War," 372.
50. Brake, "Longevity of 'Ephemera.'"

51. Chun, "Enduring Ephemeral."
52. McKernan, "Into the Void." See also McKernan's wonderfully illustrated "Leaving Colindale."
53. King, "Digital Historic Newspapers Online," 61. They also had to take copyright restrictions into account in cases where newspapers are still in print. For details of how the British Library addressed such concerns, see King, "British Library Digitisation."
54. Shaw, "10 Billion Words," 3.
55. King, "Digitisation of Newspapers at the British Library."
56. Shaw, "British Newspapers," 12, 14.
57. King, "Digital Historic Newspapers Online," 62.
58. Shaw, "10 Billion Words," 10.
59. Shaw, "British Newspapers," 21-22.
60. *Ibid.*, 11.
61. Cordell, "Q I-Jtb the Raven."
62. Many thanks to Jason Groth of the NCSU Libraries for help on these points.
63. Shaw, "British Newspapers," 16, 30.
64. Personal communication.
65. King, "Digitisation of Newspapers at the British Library," 179.
66. King, "Digital Historic Newspapers Online," 63.
67. See the Wayback snapshot at http://web.archive.org/web/20070717194234/http://www.apexcovantage.com/about_us.aspx.
68. See <http://fast.wistia.net/embed/iframe/8aouo435t1?popover=true>.
69. Cecire, "Visible Hand."
70. Mak, "Confessions."
71. These days, the website for Olive Software shows signs of age, with several broken links; however, the *Internet Archive* offers snapshots of the site during the company's partnership with the British Library. At that time, the company promoted its "ActivePaper Archive" product for historical newspaper digitization, which offered automated strategies for generating XML. See the Wayback Machine's snapshot, <http://web.archive.org/web/20060316022837/http://www.olivesoftware.com/products/technology.asp>.
72. Shaw and Fleming, "JISC Project Plan," 8.
73. Shaw, "British Newspapers," 8.
74. Abruzzi offers a glimpse of Gale's cultural emphases in Abruzzi, Calè, and Vadillo, "Gale Digital Collections."
75. Personal communication.
76. Shaw, "British Newspapers," 15.
77. King, "Digital Historic Newspapers Online," 65. Incidentally, the Library of Congress's *Chronicling America* collection of nineteenth-century US newspapers does not segment by article.

78. Hobbs, "Deleterious Dominance." Coincidentally, Andrew Hobbs (then a PhD student) was quoted in the promotional materials for the JISC I Project Plan as saying it "could well change the face of British historiography." See Shaw and Fleming, "JISC Project Plan," 36.
79. In "Illusionary Order," Milligan makes a similar argument about how the availability of digital resources privileges certain papers and modes of Canadian history.
80. The file structures are also different. Part I separates XML files from TIFF image collections. Part II includes XML and JPEG files together for each day's issue.
81. Upchurch, "Full-Text Databases."
82. Nicholson, "Counting Culture," 243–44. See also the extensive documentation of coverage and gaps in Shaw, "British Newspapers," 34.
83. Mussell, "Teaching Nineteenth-Century Periodicals," 201.
84. Coincidentally, among the listed partners of Digital Divide Data is Ancestry.com, <http://www.digitaldividedata.com/about>.
85. Abruzzi, personal communication.
86. Abruzzi, Calè, and Vadillo, "Gale Digital Collections."
87. Cayley, "Digitisation in Teaching and Learning"; Cayley and Horrocks, "*Punch Historical Archive*."
88. Cayley, "Creating the *Daily Mail* Historical Archive."
89. Harris, *History of the British Museum Library*, xv.
90. In a personal communication, King qualifies my claims, suggesting that "in the context of a national library such as the British Library, institutional memory and files of archived papers relating to the work of previous generations can remain quite strong."
91. Lepore, "Cobweb."
92. Shaw, "British Newspapers," 30.

BIBLIOGRAPHY

- Abruzzi, Ray, Luisa Calè, and Ana Parejo Vadillo. "Gale Digital Collections: Ray Abruzzi Interviewed by Luisa Calè and Ana Parejo Vadillo." *19: Interdisciplinary Studies in the Long Nineteenth Century* 21 (December 10, 2015). <http://19.bbk.ac.uk/articles/753/>.
- Alfano, Veronica, and Andrew M. Stauffer, eds. *Virtual Victorians: Networks, Connections, Technologies*. New York: Palgrave Macmillan, 2015.
- Baker, Drew, Anna Bentkowska-Kafel, and Hugh Denard, eds. *Paradata and Transparency in Virtual Heritage*. Farnham: Ashgate, 2012.
- Born, Lester. "Planning for Scholarly Photocopying." *PMLA* 79, no. 4 (1964): 77–90.

- Bourke, Thomas A. "Scholarly Micropublishing, Preservation Microfilming, and the National Preservation Effort in the Last Two Decades of the Twentieth Century: History and Prognosis." *Microform Review* 19, no. 1 (1990): 4-16.
- Bradsher, Greg. "Archivist Monuments Man: Lester K. Born." *Text Message*, April 8, 2014. <http://text-message.blogs.archives.gov/2014/04/08/monuments-man-lester-k-born>.
- Brake, Laurel. "Half Full and Half Empty." *Journal of Victorian Culture* 17, no. 2 (2012): 222-29.
- . "London Letter: Researching the Historical Press, Now and Here." *Victorian Periodicals Review* 42, no. 2 (2015): 246-53.
- . "The Longevity of 'Ephemera': Library Editions of Nineteenth-Century Periodicals and Newspapers." *Media History* 18, no. 1 (2012): 7-20.
- Cayley, Seth. "Creating the *Daily Mail Historical Archive*." Gale Cengage, 2013. <http://gale.cengage.co.uk/images/Creating%20the%20Daily%20Mail%20Historical%20Archive.pdf>.
- . "Digitization in Teaching and Learning: The Publisher's View." *Victorian Periodicals Review* 45, no. 2 (2012): 210-14.
- Cayley, Seth, and Clare Horrocks. "*The Punch Historical Archive, 1841-1992: A Sustainable Brand for the Digital Age*." *Victorian Periodicals Review* 48, no. 2 (2015): 238-43.
- Cecire, Natalia. "The Visible Hand." *Works Cited*, May 3, 2011. <http://natalia-accire.blogspot.com/2011/05/visible-hand.html>.
- . "Ways of Not Reading Gertrude Stein." *ELH* 82 (2015): 281-315.
- Chun, Wendy Hui Kyong. "The Enduring Ephemeral, or the Future Is a Memory." *Critical Inquiry* 35, no. 1 (2008): 148-71.
- Cordell, Ryan. "'Q I-Jtb the Raven': Taking Dirty OCR Seriously." Paper presented at the MLA Convention, 2016, Austin, Texas. <http://ryancordell.org/research/qijtb-the-raven/>.
- Darnton, Robert. "The Library in the New Age." *New York Review of Books*, June 12, 2008. <http://www.nybooks.com/articles/archives/2008/jun/12/the-library-in-the-new-age/>.
- Drucker, Johanna. "Humanities Approaches to Graphical Display." *Digital Humanities Quarterly* 5, no. 1 (Winter 2011). <http://digitalhumanities.org/dhq/vol/5/1/000091/000091.html>.
- Edwards, Adrian S. "Destroyed, Damaged and Replaced: The Legacy of World War II Bomb Damage in the King's Library." *Electronic British Library Journal*, 2013. <http://www.bl.uk/ebj/2013/articles/article8.html>.
- Eggert, Paul. *Securing the Past: Conservation in Art, Architecture and Literature*. Cambridge: Cambridge University Press, 2009.
- Ernst, Wolfgang. *Digital Memory and the Archive*. Edited by Jussi Parikka. Minneapolis: University of Minnesota Press, 2012.
- Gadd, Ian. "The Use and Misuse of *Early English Books Online*." *Literature Compass* 6, no. 3 (2009): 680-92.

- Gibbs, Frederick W., and Daniel J. Cohen. "A Conversation with Data: Prospecting Victorian Words and Ideas." *Victorian Studies* 54, no. 1 (2011): 69–77.
- Gillies, Stewart. "History of British Library Newspapers." *British Library*, 2008. <http://www.bl.uk/reshelp/findhelprestype/news/historicalblnews/>.
- Gitelman, Lisa. *Always Already New: Media, History, and the Data of Culture*. Cambridge: MIT Press, 2006.
- . *Paper Knowledge: Toward a Media History of Documents*. Durham: Duke University Press, 2014.
- Gitelman, Lisa, and Virginia Jackson. "Introduction." In "*Raw Data*" is an Oxymoron, edited by Lisa Gitelman, 1–14. Cambridge: MIT Press, 2013.
- Harris, Philip Rowland. *A History of the British Museum Library, 1753–1973*. London: British Library, 1998.
- Hiscoe, David. "NCSU Libraries Opens Pioneering New Possibilities for Data Mining Historical Content." *NCSU Libraries*, November 11, 2014. <http://www.infodocket.com/wp-content/uploads/2014/11/final-Gale-data-mining-press-release-1103142.pdf>.
- Hitchcock, Tim. "Academic History Writing and Its Disconnects." *Journal of Digital Humanities* 1, no. 1 (2011). <http://journalofdigitalhumanities.org/1-1/academic-history-writing-and-its-disconnects-by-tim-hitchcock/>.
- Hobbs, Andrew. "The Deleterious Dominance of *The Times* in Nineteenth-Century Scholarship." *Journal of Victorian Culture* 18, no. 4 (2013): 472–97.
- King, Ed. "British Library Digitisation: Access and Copyright." World Library and Information Congress. Quebec, 2008.
- . "Digital Historic Newspapers Online: Prospects and Challenges." In *Newspapers Collection Management: Printed and Digital Challenges. Proceedings of the International Newspaper Conference, Santiago de Chile, April 3–5, 2007*, edited by Hartmut Walravens, 59–78. München: Saur, 2008.
- . "Digitisation of Newspapers at the British Library." *Serials Librarian* 49, no. 1–2 (2005): 165–81.
- . "Legal Deposit of Newspapers at the British Library: Past, Present and Future." In *Newspapers: Legal Deposit and Research in the Digital Era*, edited by Hartmut Walravens, 161–66. Berlin: Walter de Gruyter, 2011.
- Kirschenbaum, Matthew. *Mechanisms: New Media and the Forensic Imagination*. Cambridge: MIT Press, 2008.
- . "The .txtual Condition: Digital Humanities, Born-Digital Archives, and the Future Literary." *Digital Humanities Quarterly* 7, no. 1 (2013). <http://www.digitalhumanities.org/dhq/vol/7/1/000151/000151.html>.
- Leary, Patrick. "Googling the Victorians." *Journal of Victorian Culture* 10, no. 1 (2005): 72–86.
- Lepore, Jill. "The Cobweb: Can the Internet Be Archived?" *New Yorker*, January 26, 2015. <http://www.newyorker.com/magazine/2015/01/26/cobweb>.
- Liddle, Dallas. "Reflections on 20,000 Victorian Newspapers: 'Distant Reading' *The Times* Using *The Times Digital Archive*." *Journal of Victorian Culture* 17, no. 2 (2012): 230–37.

- Mak, Bonnie. "Archaeology of a Digitisation." *Journal of the American Society for Information Science and Technology* 65, no. 8 (2014): 1515–26.
- . "Confessions of a Twenty-First-Century Memahib: The Offshore Sweatshops of the Digital Humanities." Paper presented at the MLA Convention, Austin, Texas, 2016.
- Marvin, Carolyn. *When Old Technologies Were New: Thinking About Electric Communication in the Late Nineteenth Century*. New York: Oxford University Press, 1988.
- Massari, Kristina. "Gale Leads to Advance Academic Research by Offering Content for Data Mining and Textual Analysis." *Cengage Learning*, November 17, 2014. <http://news.cengage.com/higher-education/gale-leads-to-advance-academic-research-by-offering-content-for-data-mining-and-textual-analysis/>.
- McKernan, Luke. "Into the Void." *Newsroom Blog: British Library*, January 27, 2015. <http://britishlibrary.typepad.co.uk/thenewsroom/2015/01/into-the-void.html#>.
- . "Leaving Colindale." *Luke McKernan*, October 9, 2013. <http://lukemckernan.com/2013/10/09/leaving-colindale/>.
- Meckler, Alan M. *Micropublishing: A History of Scholarly Micropublishing in America, 1938–1980*. Westport: Greenwood Press, 1982.
- Milligan, Ian. "Illusionary Order: Online Databases, Optical Character Recognition, and Canadian History, 1997–2010." *Canadian Historical Review* 94, no. 4 (2013): 540–69.
- Mussell, James. *The Nineteenth-Century Press in the Digital Age*. New York: Palgrave Macmillan, 2012.
- . "Teaching Nineteenth-Century Periodicals Using Digital Resources: Myths and Methods." *Victorian Periodicals Review* 45, no. 2 (2012): 201–9.
- Nicholson, Bob. "Counting Culture; Or, How to Read Victorian Newspapers from a Distance." *Journal of Victorian Culture* 17, no. 2 (2012): 238–46.
- Nunberg, Geoffrey. "Google Books: A Metadata Train Wreck." *Language Log*, August 29, 2009. <http://languagelog.ldc.upenn.edu/nll/?p=1701>.
- Peiss, Kathy. "Cultural Policy in a Time of War: The American Response to Endangered Books in World War II." *Library Trends* 55, no. 3 (2007): 370–86.
- Pionke, Albert D. "Excavating Victorian Cuba in the British Periodicals Database." *Victorian Periodicals Review* 47, no. 3 (2014): 369–97.
- Power, Eugene B. *Edition of One: The Autobiography of Eugene B. Power, Founder of University Microfilms*. Ann Arbor: University Microfilms International, 1990.
- Shaw, Jane. "10 Billion Words: The British Library Newspapers 1800–1900 Project. Some Guidelines for Large-Scale Newspaper Digitisation." In *Libraries—A Voyage of Discovery*. Oslo: IFLA, 2005. http://www.webarchive.org.uk/way-back/archive/20140615090156/http://www.jisc.ac.uk/uploaded_documents/IFLA_2005.pdf.

- . “British Newspapers 1620–1900: Final Report.” JISC and the British Library, August 2009. <http://www.webarchive.org.uk/wayback/archive/20140614080134/http://www.jisc.ac.uk/media/documents/programmes/digitisation/blfinal.pdf>.
- Shaw, Jane, and Patrick Fleming. “JISC Project Plan.” JISC and the British Library, October 2008. http://www.webarchive.org.uk/wayback/archive/20140614080145/http://www.jisc.ac.uk/media/documents/programmes/digitisation/bl_newspapers_public_plan.pdf.
- Sherratt, Tim. “Stories for Machines, Data for Humans.” *Discontents*, April 10, 2015. <http://discontents.com.au/stories-for-machines-data-for-humans/>.
- Terras, Melissa. “Father Busa’s Female Punch Card Operatives.” *Melissa Terras’ Blog*, October 15, 2013. <http://melissaterras.blogspot.com/2013/10/for-ada-lovelace-day-father-busas.html>.
- Upchurch, Charles. “Full-Text Databases and Historical Research: Cautionary Results from a Ten-Year Study.” *Journal of Social History* 46, no. 1 (2012): 89–105.
- Weiner, Joel. “Newspaper Taxes, Taxes on Knowledge, Stamp Taxes.” In *The Dictionary of Nineteenth-Century Journalism* (online edition), edited by Laurel Brake and Marysa Demoor. Ghent and London: Academia Press and the British Library, 2009.