

Abstract

This paper traces the little-known history of the *London Stage Information Bank*, a digital initiative that ran from 1970 to 1978 under the direction of Professor Ben R. Schneider, Jr. at Lawrence University. With support from the National Endowment for the Humanities, the American Council of Learned Societies, and the Mellon Foundation, Schneider's team produced a database from the multi-volume reference work *The London Stage 1660-1800* (Southern Illinois University Press, 1960-68). Today, however, most of the project's outputs are lost or damaged, and its history has been largely forgotten in both theater studies and eighteenth-century studies. This essay traces the history of the Information Bank and my efforts to recover its damaged data and code, offering the project as an object lesson in questions of access, preservation, and institutional memory that digital humanities practitioners continue to confront in 2017. I argue that the project faded into obscurity, not only because of technological obsolescence, but also because the development team was unable to promote the kinds of research questions and behaviors that would enable their tool's widespread adoption and survival. The indifference of literary and theater scholars to the *Information Bank* throughout the late 1970s and early 1980s demonstrates how vital it is that digital and computational humanities work articulate its meaningfulness within existing intellectual and disciplinary traditions. While digital scholars build new avenues for inquiry that expand and transform humanities research, the survival of these approaches depends on their relationship to current humanities questions, methods, commitments, and epistemologies.

The year is 1970. Scholars of British theater are celebrating the recent publication of *The London Stage, 1660-1800*, an 11-book, 8000-page calendar of theatrical performances in London over a 140-year period. After more than a decade conducting the archival and editorial work necessary to produce this reference work, the advisory board seeks outside help constructing an index, which they hope will allow researchers to cross-reference information about a given play title or actor across volumes. The theater scholar they choose for the job, a dabbler in the new field of humanities computing, is confident that a program can perform this task better than a person. He assembles an international team of humanities scholars, technologists, and students, garners hundreds of thousands of dollars in private and federal funding, and spends the better part of a decade producing a database and a suite of tools for accessing and manipulating it. In the end, however, his team's work is lost to technological obsolescence, as well as to the indifference of a scholarly community unconvinced of the need for such digital and computational approaches to humanities research.

1

This is the story of the *London Stage Information Bank*, an early humanities computing project that presents irresistible parallels to many issues facing digital humanities today. The project ran from 1970 to 1978 under the direction of Professor Ben Ross Schneider, Jr. at Lawrence University in Appleton, Wisconsin, and it was enabled by grants from the National Endowment for the Humanities, the American Council of Learned Societies, the Mellon Foundation, and other major funders. Today, however, most of the project's outputs are lost or damaged, its history largely forgotten. This essay presents a study of the archival record and material artifacts of this project, positioning the *Information Bank* as an object lesson in issues of access, preservation, and institutional memory that digital scholars continue to confront in 2017.

2

More than a mere cautionary tale about data loss, however, the history of the *Information Bank* illuminates the need for researchers to articulate the relationship of digital research methods to existing intellectual and disciplinary traditions. I will argue that the project faded into obscurity, not only because it was ahead of its time or because the rapid advancement of computing technology took the team by surprise, but also because Schneider made key assumptions about the intellectual dispositions of his user base. Proceeding from those assumptions, his team did not promote the kinds of research practices that would ensure their tool's widespread adoption and survival. Their presentations and publications exhibited the capabilities of the database but largely assumed that its usefulness would be self-evident—a tendency that continues today in many demos of digital humanities datasets and tools. While digital scholars can and should build new avenues for inquiry that expand and transform humanities research, the survival of these approaches depends on their relationship to

3

current humanities questions, methods, commitments, and epistemologies. Digital humanities practitioners must model the modes of inquiry our work enables and demonstrate that these modes allow us to produce new, exciting knowledge that is legible to humanists who do not identify as digital scholars.

This essay first presents a history of the *London Stage Information Bank* and my ongoing effort to recover it, then synthesizes the lessons that can be learned from the project and suggests future applications of those lessons. In doing so, it contributes to the growing body of work that recognizes the importance of a reciprocal exchange of perspectives between theater and performance studies and digital humanities. As Debra Caplan has recently pointed out, the historical lack of interaction between the two communities has led to serious issues of accessibility and preservation in digital theater studies: "Without disciplinary-wide best practices for creating, funding, disseminating, or reviewing digital humanities work, projects are often idiosyncratic in the technologies and approaches they choose, and projects are not always equally accessible" [Caplan 2015, 358–59]. At the same time, as Sarah Bay-Cheng has argued and as Caplan echoes, theater studies researchers also have much to offer to the digital humanities, particularly in their long tradition of attention to the ephemerality of interactions—a problem that is inherent both to performance history and to digital media and culture [Bay-Cheng 2012].^[1] This interplay between ephemerality and durability, dramatized in the ability of a fleeting performance event to leave traces and create impacts that ripple out across time, resonates with recent media archaeological work within digital humanities. Jussi Parikka reminds us that digitally encoded information, often imagined in terms of its "immaterial virtuality," is in fact dependent on "hardware, software, and other material contexts" that are "prone to deterioration;" for Parikka, "[t]he digital is not eternal, nor is it simply ephemeral" [Parikka 2012, 118–19]. Matthew Kirschenbaum similarly highlights the tension between the mutability and resilience of digital inscriptions on physical media in his book *Mechanisms: New Media and the Forensic Imagination* [Kirschenbaum 2008], his emphasis often falling on the curious survival of seemingly evanescent digital texts.^[2] This is a tension that accords especially well with theater studies, where an interest in ephemerality must be balanced with an awareness of the mediated afterlife of performance.

In both theater studies and media archaeology, then, poststructuralist notions of the "always already disappearing" object of study are confronted by a growing awareness of the stubborn material residue of those objects' transmission [Bloom et al. 2013, 167]. The *London Stage Information Bank* exemplifies this tension: it was lost to technological change, yet it survives in at least partially recoverable forms. Furthermore, notions of residue and resonance drive us to recognize not only how the material stuff of hardware, software, and data persist, but also how more abstract artifacts—data models, for instance, that reflect particular orientations toward the objects under consideration—are passed down and inherited from collection to collection, from tool to tool.^[3] My attempt to recover both the history and the material artifacts of this forgotten project illustrates the productive interplay between the theoretical orientations of theater history, performance studies, and media archaeological approaches within digital humanities.

History of the *Information Bank*

The eleven books of *The London Stage, 1660-1800: A Calendar of Plays, Entertainments & Afterpieces, Together with Casts, Box-Receipts and Contemporary Comment. Compiled from the Playbills, Newspapers and Theatrical Diaries of the Period* were published by Southern Illinois University Press between 1960 and 1968. These books contain extensive information about nearly all recorded performances thought to have taken place in London over the course of the long eighteenth century, based on archival evidence held in libraries throughout the U.S. and the U.K. The volumes are organized by theatrical season, with a typical entry representing a performance at a specific theater on a specific evening. It usually gives the date or approximate date of the performance, the title of the main play staged, the cast list if known, and any other entertainments that accompanied the main attraction. Some entries specify the amount of money that the theater made that evening, mention prominent audience members, or detail the provenance of the information in the entry (Figure 1).

THE BEGGAR'S OPERA. Macheath – Beard; Peachum – Yates; Lockit – Berry; Saturday 14
 Beggar – Winstone; Player – Burton; Lucy – Mrs Clive; Mrs Peachum – Mrs DL
 Cross; Filch – Raftor; Diana Trapes – Mrs Havard; Polly – Miss Norris; Mat-o-
 Mint – Blakes; Mrs Slammekin – Mrs Bennet. In Act III a *Hornpipe* by Mathews.
 To conclude with a *country dance* by the characters of the opera. Also QUEEN
 MAB. As 12 Oct.
 DANCING. I: The *Matelot Basque*, as 5 Oct.
 COMMENT. Receipts: £100 (Cross).

Figure 1. A typical entry from *The London Stage, 1660-1800* (Avery et al 1960-68, Part 4, Vol. 1, p. 325). This entry represents a performance at the Theatre Royal–Drury Lane ("DL") on Saturday, October 14, 1752. The main play performed was *The Beggar's Opera*, a ballad opera by John Gay that was popular throughout the century. The cast list is provided, showing that, for example, the leading role of Macheath was played by an actor named Beard. The description of the hornpipe and country dance featured in this performance are assumed to come from the newspaper advertisement or playbill on which this entry is based. *Queen Mab* is the title of the afterpiece, a shorter entertainment that would have been performed after the conclusion of *The Beggar's Opera*. "As 12 Oct". indicates that a cast list is provided in the advertisement but omitted here because it is identical to the one that appears in an entry for a performance two days earlier. (See Note 8 for additional information on the abbreviation of cast lists in *The London Stage*.) An additional dance performance titled *Matelot Basque* was featured. The theater sold approximately £100 worth of tickets that evening, according to the receipt book maintained by prompter Richard Cross.

The London Stage is clearly a valuable resource for theater history, but it also has serious limitations. One major issue is that its indices are volume-specific, making it difficult to trace particular actors or play titles across decades. This obstacle has been partially redressed by the recent release of searchable full-text scans of the reference books through HathiTrust, but one must still perform a single keyword search on each of the eleven books individually.^[4] Importantly, neither the reference books nor their digitized counterparts can readily support complex queries that go beyond indexing to describe relationships among objects and persons, a limitation that has been recognized from the beginning. Schneider and his programmer partner, Will Daland, detailed the books' limited operability in 1971, in an article that appeared in the journal *Computers and the Humanities*: "If, for example, one wished to determine how many times actor X and actress Y performed in the same play together during their careers, it might be necessary to scan a period of fifteen or twenty years (possibly 800 to 1,000 pages) to exhaust all the possibilities of intersection" [Schneider and Daland 1971, 209]. The previous year, Schneider had been approached by the editorial board of the series to create a computerized index of all of the volumes.^[5] By 1979, Schneider's team had published *The Index to The London Stage*, which contained entries for the entities they thought researchers would be most interested in, such as actor names; in the process, they created the *Information Bank*. Schneider's *Index*, now available on HathiTrust, can provide some guidance for researchers, but it represents only a fraction of all possible questions one could ask of the database.

It is possible to reconstruct the workflow involved in the transformation of the reference book into data, using the introduction to the *Index* along with a memoir Schneider wrote about the process titled *Travels in Computerland* [Schneider 1974]. Schneider recruited a small group of editors, mostly PhD students in eighteenth-century British theater, who were promised first use of the new tool in exchange for their contributions to the database. The editors, along with Schneider, created a marked-up copy of the printed reference book, using colored pencils to denote items that were to be coded in specific ways, such as performance headers, cast lists, or extraneous text that should be delimited from the main entry.^[6] The hand-annotated pages were then shipped to China Data Systems in Hong Kong, where professional typists transcribed the marked-up text. They simultaneously standardized elements like punctuation according to rules defined by Schneider, and they also coded the editorial markups using a pre-defined custom schema.

The transcribed and coded text was sent to Information Control Incorporated in Kansas City, where it underwent Optical Character Recognition (OCR). The results were stored on magnetic tapes that were returned to Appleton, where an interactive program called ICIFIX, created by Daland, was used to perform additional correction and standardization.^[7] Daland also developed a suite of programs, using the PL/1 language, to perform a variety of tasks: translating the markup used by the typists into human-readable tags, expanding the cast lists that were abbreviated by the reference book editors to save on production costs, and sorting or querying the data according to predetermined fields (date, theatre, title, role, actor, type of act) as well as tagged named entities (historical people and places, textual sources).^[8] These programs were

combined into a system called GWSJR1 (after George Winchester Stone, Jr., one of the original *London Stage* editors and patron of the *Information Bank* project) and stored on an IBM 2311 disk.

During the project's second phase—outlined in a second memoir by Schneider, titled *My Personal Computer and Other Family Crises* [Schneider 1984]—a programmer named Reid Watts developed a word-processing and concordance program called SITAR specifically for the project. Completed in 1974, SITAR then was used by as many as eighteen undergraduate student assistants to iteratively edit the underlying data for errors and inconsistencies before it was run through GWSJR1 to produce the *Index*^[9] (Figure 2).

10

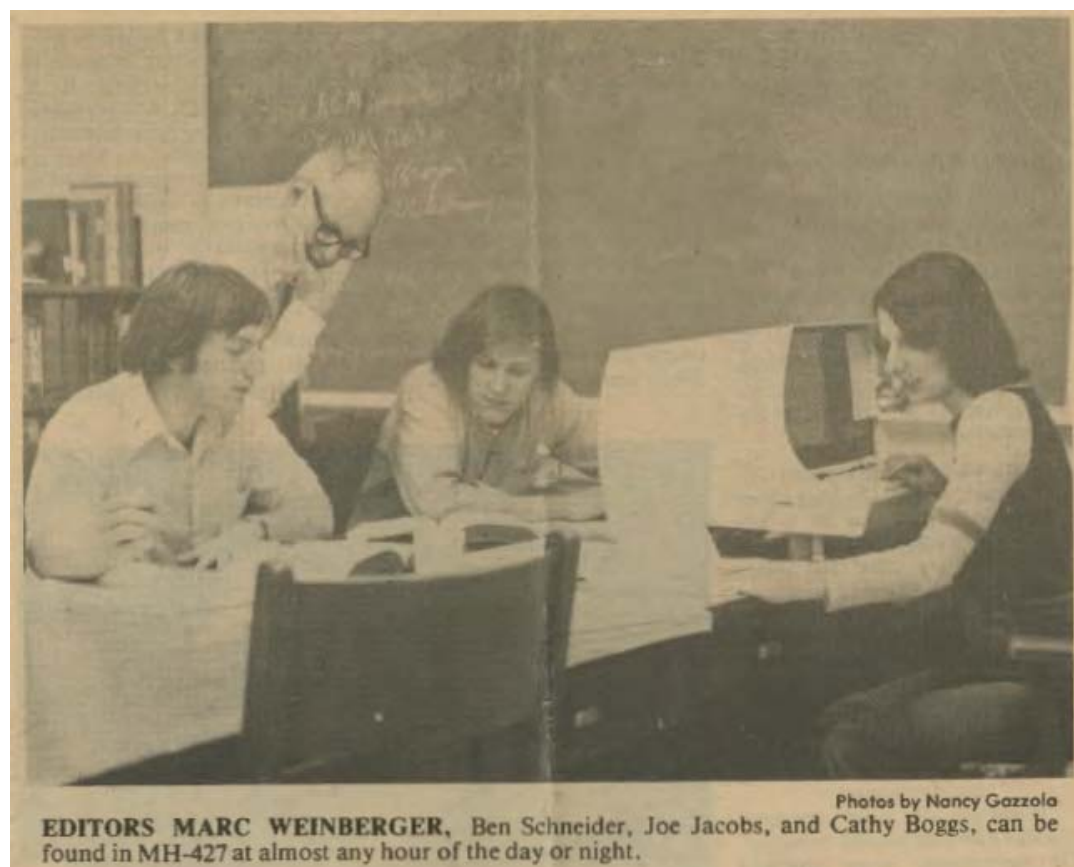


Figure 2. Schneider with three of the undergraduate students who edited and corrected the *Information Bank* data after the project's pilot phase. Source: *The Lawrentian*, 30 April 1976, courtesy of Lawrence University Archives.

The *Index* was completed in 1978 and published in 1979, signaling the end of the grant-funded phases of the project. However, Schneider continued to try to establish an ongoing maintenance and preservation plan for its various products. The underlying data for the *Information Bank* was stored on tapes that Lawrence intended to host in perpetuity, fulfilling scholarly inquiries at the base cost of computing power, in accordance with NEH guidelines (Figure 3). However, at the end of the 1970s, Lawrence stopped paying to time-share the computer they were using at a nearby research institute. In 1980, approaching retirement, Schneider tried to find a new home for the database where it would have ongoing technical support. Unfortunately, no one wanted to host it—in part because the technology was becoming gradually more obsolete, and in part because there had turned out to be very limited scholarly interest in the *Information Bank*.^[10] In 1983, the tapes were transferred to the Harvard Theatre Collection, as a storage repository rather than as a new base for operations. Shortly thereafter, the curators lent out the tapes to be migrated onto a new medium, at which point they appear to have been lost.^[11]

11

LU to house *The London Stage* data bank

A computerized information bank, apparently the first of its kind in the field of literary history, is being established at Lawrence through the efforts of Ben Schneider, professor of English.

Within three years, Project Director Schneider plans to have on computer tape the entire 11-volume, 8,000 page, three-million word reference work entitled *The London Stage 1660-1800*.

The London Stage is an exhaustive calendar of plays and other stage activities from 1660-1800. Since its publication, scholars have been able to turn to its pages with the confidence that whatever is missing concerning the performance of plays in London during the period covered almost certainly does not exist.

one serious drawback

The magnitude of the reference work, ironically, has one serious drawback. Without mechanical methods, the search for information in its pages is sometimes so extensive that the researcher can't justify the investment of time required.

Once the reference work is put on tape, however, scholars will be able to obtain compilations of data that otherwise would take weeks or months to assemble by turning pages and filling out filing cards.

"Sorting, counting and tabu-

acters in 83 plays from 1660-1730; a feat which would have been extremely difficult if attempted without mechanical help.

Schneider has received substantial financial support for the *London Stage* Data Bank project from several foundations, but the cost of the total project is still to be met. The largest anticipated expense, he said, will be the cost of transferring data from the printed page to magnetic tape. Estimating the cost of a number of available methods is a major task of the pilot project.

"In recent years, literary journals have published an increasing number of articles based on information gathered from *The London Stage*," he said.

Once established, Schneider is certain the information bank will be used.

He estimates that if all the scholars who used *The London Stage* for articles in 1969 had access to an information bank such as he is developing, there would have been about three inquiries per week.

"I'm sure that researchers in history, economics, and sociology would also find the information bank useful. It could, for example, answer questions about political trends, class mobility, availability of leisure time, and many other areas of interest as revealed in the tastes of theatre-goers of the period."



THE CONTENTS OF THOSE 11 VOLUMES of *The London Stage* 1660-1800, drawing the attention of John Church, associate director of the Institute of Paper Chemistry's computer center, Will Daland computer programmer-analyst, and Ben Schneider, professor of English, will eventually be transferred to computer tape. Project Director Schneider is currently working on the pilot phase of the

Figure 3. John Church (left) of the Institute for Paper Chemistry poses for a photo with Schneider (right) and programmer Will Daland (center). The Institute hosted the computers on which the *Information Bank* was developed and on which it was always intended to be maintained. After Lawrence University stopped paying for time on the Institute's computers in 1978, this was no longer a sustainable hosting solution. Source: *Lawrence Alumnus*, March 1971, courtesy of Lawrence University Archives.

As this brief sketch of the *London Stage* Project's history suggests, it was a highly ambitious endeavor involving dozens of personnel, including professional typists, OCR experts, programmers, graduate student editors, and numerous undergraduate assistants. It was awarded a total of \$200,000 in funding over eight years, the equivalent of roughly \$750,000 today.^[12] It was considered a success insofar as it produced a number of significant deliverables. One of these was a flat-file database of the original reference book that categorized the data within the performance entries. Other products included a system (GWSJR1) that expanded the abbreviated cast lists and allowed the data to be queried based on pre-defined categories; an interactive program (SITAR) that allowed the data to be edited and updated iteratively by non-programming specialists, and that also enabled concordance-style sorting and searching of the data; and the printed *Index to the London Stage*, which contains about 500,000 references to over 25,000 items.

Despite the successes of the project, however, there are serious material and technological barriers to accessing it today. Not only were the data tapes lost after being transferred to the Harvard Theatre Collection, but even if they could be found, and even if an appropriate machine could be located to read them, it is likely that they would be materially degraded and might be damaged by being run through a machine. Given these realities, it is possible to see the *London Stage Information Bank* as a cautionary tale of a near-decade of work lost. The next section explores in more detail why the scholarly community was uninterested in the *Information Bank* and what lessons the project has to offer for digital humanities work today.

Lessons from the *London Stage* Project

In order to discover why the *Information Bank* was lost to history, I turned to Schneider's own writings about the project—the two memoirs mentioned above, along with a series of articles published in scholarly journals and edited collections—as well as to the records of the project, now housed in the Lawrence University Archives. Taken together, these documents reveal several factors that contributed to the *Information Bank's* fate: the difficulty of developing tools that speak to current humanities research questions while also opening up new avenues of inquiry; the need to work within the incentive

structures of academe, which are sometimes ill-suited to new forms of intellectual productivity; the tendency of scholar-programmers to develop custom software and datasets from scratch rather than consulting the past or the community for models and precedents; and the relentless pace of advancement in computing technology, which all but ensures that projects developed on the time scales of humanities scholarship will be obsolete by the time they are complete. This list is likely to strike a chord with digital humanities practitioners today grappling with many of the same issues.

The archive of the London Stage Project immediately reveals a central tension between the novelty of what computers could do with humanities information and the assumption that computers were self-evidently useful tools for pursuing existing humanities research questions. In *Travels in Computerland*, Schneider extolls the virtues of the *Information Bank*, boasting that it will enable researchers to see new kinds of patterns in theater history:

15

We can study trends (the rise of pantomime; the interest in Shakespeare; the rise and fall of theatres; the decline of the drama); we can look for patterns (In what ways is one season like another? What is a typical stage career like? To what extent do actors specialize? What is the effect of the repertory system?). There's too much information about 18th century theatre; without computer help we can't see the forest for the trees. [Schneider 1974, 230]

Yet despite all the new questions a database of *The London Stage* could potentially raise and answer, Schneider's *Information Bank* garnered little attention from eighteenth-century scholars. Five years after his optimistic assessment in *Travels in Computerland*, he wrote an essay for an edited collection on *Data Bases in the Humanities and Social Sciences* lamenting the lack of interest. As the article explains, the *Information Bank* was publicized widely to the research community as a search service; in the database's first three years of availability to the public, it was advertised in seven newsletters sent to 1000 subscribers, as well as fourteen scholarly journals. In that time Schneider received 126 requests for information, including 34 requests for price quotes, but by 1979 not one researcher had followed through and paid for the results of his or her queries [Schneider 1980, 31–34].

While Schneider states bitterly and with only minimal irony that "the failure of the world to beat a path to my door is truly a mystery to me," he does offer a provisional explanation of his colleagues' indifference to the new tool:

16

Most of the research that goes on in theatre history today is precisely the kind of thing one can do just as well without the computer. At the point where a good computer printout of the repertoires of all the actors who played Shylock might reveal a great deal about the staging of *The Merchant of Venice*, it would never occur to a writer on that subject (or to his reader) that the question deserved further research in the form of a complete list of all the roles of all the actors of Shylock. The kind of thing you can do now by computer is not the kind of thing that anyone ever did, or felt the need to do. [Schneider 1980, 34]

This assertion that the database can answer unprecedented, unintuitive questions for humanities researchers is somewhat at odds with Schneider's earlier insistence that the database was of obvious relevance to current scholarship. Indeed, the questions about generic trends, acting careers, and theater finances that Schneider listed off in 1974 were relevant to theater historians in the mid-twentieth century, and they remain relevant today; the database altered the scope and speed of inquiry into these topics, but not necessarily the range of possible queries. Yet by 1980, Schneider had also identified a mismatch between the kinds of questions that interested scholars at the time and the kinds of questions a computer could answer. For Schneider, the fundamental issue was that his database was ahead of its time—and certainly, it was an uphill battle to ask humanities researchers to think quantitatively or at scale. Then as now, scholars of both literature and theater tended to focus a given study on a limited number of texts or performances, sometimes even a single work or writer. On the other hand, there is a way in which the *Information Bank* could be seen in the 1980s as *behind* the times. The potentially new possibilities for inquiry that the database offered were simultaneously aligned with old-fashioned forms of theater history characterized by the search for factual information about past performances and by the urge to count them. In the early decades of the "cultural turn," applications of poststructuralist theories represented the leading edge of humanities research—and stood firmly against the arguably positivistic orientation of the *Information Bank*.

As this account suggests, the *Information Bank* struggled to define itself simultaneously as part of an existing intellectual tradition and as the next wave of scholarship, a balancing act that continues to challenge digital humanities practitioners. The archive furthermore indicates that Schneider and his team missed key opportunities to model the new kinds of inquiry that the *Information Bank* offered and to articulate these techniques' role within ongoing conversations in the field. The

17

climax of *Travels in Computerland* revolves around the last-minute rush to process queries for graduate student editors Leonard Leff and Muriel Friedman to use in their presentations at the MLA 1971 annual meeting in Chicago.^[13] These results, originally expected in late summer, were delivered in mid-December, about two weeks before the conference. This lag occurred because the technical side of the operation was focused on error-correction while the research side wanted only good-enough data to produce sample results as a proof of concept for fellow scholars [Schneider 1974, 143, 202, 212–13]. The final narrative chapter of *Travels* ends with an unplanned trans-Atlantic flight and an eleventh-hour triumph over programmatic errors, leading to the successful production of the output needed for the MLA seminar; unfortunately, help came too late, as Schneider is forced to admit in the memoir's postscript:

The reader may still be wondering how the seminar turned out at the Modern Language Association....Well, although Muriel brought her printout to the meeting, she did not treat it specifically, and Leonard, not having time to study his, left it at home and gave a theoretical discussion of the subject. Two months later I heard from a scholar who'd been there that he'd gotten the distinct impression that the project had fallen short of its goals. [Schneider 1974, 244]

The team was unable to demonstrate the kinds of results that could be obtained from querying the new database and the ways that those results could shed new light on current scholarly problems—such as, in Leff's case, the casting of Richard Brinsley Sheridan plays over time and the relationship of anti-Irish sentiment to the casting of controversial roles [Schneider 1974, 194–95]. Furthermore, this anecdote points to the misfit between the London Stage Project's outputs and the incentive structures within which its personnel operated. The graduate student editors who dedicated their time to the project were not able to translate that work into concrete findings that could be used for their dissertations, an issue that the scholarly community continues to work through today as we debate the best ways to articulate the value of data curation and digital tool development for hiring, promotion, and tenure—particularly when that labor leads to productive failures rather than peer-reviewed products like articles and books.

In another parallel to the present day, Schneider found himself up against other academics' impulse to reinvent the wheel. While offering *Information Bank* queries at cost to researchers, Schneider also offered his programs for purchase. Yet his 1980 essay indicates that there, too, interest fell short of expectations. He found that new humanities computing projects were unwilling to invest in prefabricated programs that could be lightly tailored to their purposes: "For some reason, almost everyone would rather write software from scratch than get it at a fraction of the cost ready-made, thoroughly tested and debugged....It does not seem to occur to scholars embarked for the first time on computer projects that what they want to do has ever been done before" [Schneider 1980, 33]. The unnecessary duplication of effort continues in the digital humanities community today, as granting agencies have historically tended to award projects that are building something new, rather than those that are drawing on or sustaining existing resources. This funding situation creates incentives to make new tools instead of adapting, maintaining, or updating ones that have already been made. The result is the disappearance of many projects that might otherwise form the foundation for subsequent work: Robin Camille Davis has found that nearly half of the projects presented at the 2005 Digital Humanities conference were no longer online a decade later [Davis 2015].

18

As the imperative to incentive sustainability and avoid duplication of effort has become more visible, sites like DHCommons and the Mellon-funded DiRT Directory have emerged to help people with similar interests find pre-existing tools as well as collaborators. The NEH's Office of Digital Humanities has taken the important step of extending its Advancement Grants to projects "revitalizing and/or recovering existing digital projects," rather than only incentivizing the creation of brand new ones. Likewise, the ACLS Digital Extension Grants are aimed at "enhancing established digital projects and extending their reach to new communities of users," rather than providing startup funding for nascent projects. In addition, many digital humanities research centers are beginning to take a tiered approach, differentiating between researchers whose projects can be accomplished using out-of-the-box solutions and those who truly need to build their projects from the ground up. These signs point to a growing commitment to raise the survival and adoption rates of digital humanities projects, ending the cycle of reinvention and reduplication.

19

In his efforts to market his tools, Schneider not only discovered that his humanities computing colleagues preferred to build their own software from the ground up; he also learned that his products were considered obsolete, making them increasingly difficult to market to those who might adopt or adapt them to other purposes. In response to this growing threat, Schneider was defiant. The Lawrence University archive houses a poignant 1976 letter in which Schneider responds to a potential funder's concerns about his technology being out of date:

20

Our system cannot become obsolete from advances in hardware, because our programs are written in BASIC and PL/1, assiduously supported by Digital Equipment Corporation and IBM, the two leading computer manufacturers: there is no chance that either will build computers that are incompatible with these programming languages, or that Lawrence would be so foolhardy as to buy computers incompatible with ten years programming work.^[14]

Such a statement may sound hubristic, but four decades later, digital humanists still tend to underestimate the speed with which our projects will become outdated. Digital preservation specialists continue to warn humanities researchers and digital content producers that the apparently durable file formats and access mechanisms of the present day are less stable than they may appear [Conway 2010] [Library of Congress 2013]. Schneider's story stands as a warning: if we wish for future projects to be able to build on our tools and datasets without having to start from scratch, then we must ensure that our outputs are designed to be accessible for years and decades to come.

Recovery Efforts and Directions Forward

While it is true that much of Schneider's work was lost, in recent years I have unearthed the project's partial remains and, with the help of many collaborators, begun recovering their functionality. Erin Dix, Archivist at Lawrence University, helped me retrieve not only the paper files from the project, but a set of 3.5" floppy discs labeled "LSP_data" containing plain-text ASCII files and metadata suggesting they were written in 1990.^[15] The data appears to represent a large majority of the performances from the original reference book, although some gaps have been identified; notably, several seasons from the 1730s and the 1780s are missing. It is unclear whether the data represents the edits made over the years by research assistants, or whether it represents the raw data as it arrived from Information Control Incorporated on magnetic tapes; it may also represent some intermediate stage.^[16] What is clear, however, is that it has *not* been run through the programs that parsed the data for querying and expanded the cast lists; equally clear is that it has been converted from EBCDIC to ASCII, with some unintended consequences. The underlying hexadecimal code has been shifted such that most of the performance dates are represented as special characters rather than numerals.^[17] Derek Miller has developed a script that corrects the hex; it is important to note, however, that the data itself cannot be forensically reverted, so programmatically corrected versions will always represent an approximation of the original.^[18]

The Lawrence University Archives preserved this data as well as the grant applications, correspondence, and press from the London Stage Project referenced throughout this essay; they did not, however, preserve Daland's code base, perhaps because it was thought to be archived at Harvard.^[19] Fortunately, Daland kept printouts of the programs and their documentation in his personal files, and he scanned and emailed these papers to me as a combination of PDF and TIF files. Although we experimented extensively with scan settings and image processing techniques to optimize the images for OCR, we ultimately found them to be resistant to character recognition and resorted to hand-transcription of sample sections of the code. Working with a mainframe computer at the University of Wisconsin-Madison, we tried to compile and run a hand-transcribed and -corrected version of STRUCTUR, the main parsing program. After exhaustive efforts to reconstruct and compile the programs in their original form, we determined that doing so would be impractical; it would force us to reproduce and deal with numerous constraints around memory, character sets, and encoding scheme conversion that need not be factors in a modern computing environment. As a result, I am currently collaborating with Todd Hugie (Director of Library Information Technology, Utah State University) to re-engineer the code base in Python, creating a new parser for the recovered flat-file data based on the principles represented in Daland's code. From there, we plan to transform the data into XML and JSON formats for preservation and sharing, then import the data into a relational database such as MySQL or MariaDB.^[20]

Our work to restore the *London Stage Information Bank* responds to several needs in the scholarly community, including a fundamental need for a database of the performance records in *The London Stage*. The reference books remain one of the most frequently consulted resources in eighteenth-century studies, and while they are available in searchable form through HathiTrust, keyword search remains an unsatisfactory method of querying the rich, relational information contained in those pages. Scholars continue to perform hand-counts of performances of interest, as Elaine McGirr does in her monograph on playwright, actor, and theater manager Colley Cibber [McGirr 2016], and as the Cambridge Ben Jonson project did in constructing their searchable "Performance Archive" of stagings of Jonson plays from the seventeenth century to the present.^[21] In 2016, publisher Adam Matthew released a new primary source collection, *Eighteenth Century Drama: Censorship, Society and the Stage*, which is based on the Huntington Library's Larpent Collection and includes a

21

22

23

searchable database based on *The London Stage*. This release generated significant interest in the eighteenth-century and theater studies communities, but the limitations quickly became apparent: the collection carries a high subscription fee and only institutions, not individual scholars, are permitted to subscribe, making the collection inaccessible to all but members of the wealthiest institutions. Furthermore, the database itself is designed to permit queries only along specific parameters such as title or date, rather than any kind of exploratory statistical analysis of the full dataset. For these reasons, an open-access, open-source database remains desirable to many researchers in these fields.

As I and my collaborators undertake to re-engineer and revitalize the *London Stage Information Bank*, we are mindful of the obvious lessons in preservation and sustainability offered by the history sketched out above; any new version of the database will need to adhere to current best practices for sustainability, developed in consultation with librarians and archivists who have expertise in this area. However, we are also aware that one of the best ways to ensure preservation is to attract and maintain a large and engaged user base. In order to avoid the disconnect explored above between the original tool's affordances and the dispositions and concerns of its target users, the new *London Stage* database would need to be adapted in several key ways to meet the needs of today's humanities scholars: it would need to acknowledge its genealogy by building on the past iterations; it would need to accommodate and even highlight the ambiguity and messiness of the data; and it would need to contribute to current efforts to develop data ontologies that make sense for theater studies. The rest of this essay considers how these features would enable the database to speak to current questions and debates in humanities scholarship.

Beyond offering a useful research tool, a *London Stage* database built on the foundations of the *Information Bank* would contribute to recent efforts to expose the epistemological foundations of the 1970s- and 1980s-era humanities computing projects that continue to underpin many large-scale databases used today.^[22] Users who wish only to investigate specific questions about the theater of the time could do so, but the resource would also accommodate those hoping to examine more closely the layered history of the system itself. We envision an interface that allows users to download and view our reengineered parsing program alongside the original version; likewise, we imagine allowing users to download the flat-file database in its raw and parsed forms, as well as versions of the same data stored in XML and JSON formats. Such an interface would offer a unique opportunity to look underneath the hood of a 1970s-era humanities computing project, and to compare the underlying technological and ontological structures of Schneider's database to those of its updated counterpart. This comparison could offer a window onto the architecture and assumptions about the nature of humanities data that formed the invisible foundation for humanities computing work in the era—assumptions that eventually made their way into many of the digital resources we rely on today, often without full knowledge of their provenance or architecture. In this respect, our project responds to Bonnie Mak and Ian Gadd's calls for what Mak calls an "archaeological approach" to collections like *Early English Books Online* that incorporate and reproduce the biases inherent in cataloging, preservation, and digitization efforts that go back decades [Mak 2014] [Gadd 2009] [Gadd 2015].^[23]

In the case of the *London Stage Information Bank*, we can already begin to suggest the orientation towards humanities data that it represents, based on the query results published as the *Index to the London Stage*. Domain-specialist users of the *Index* have long recognized in its entries an orientation towards historical data that is ill suited to the ambiguity surrounding much of this material. For instance, a recent case study of eighteenth-century London playbills shows that the many theatrical adaptations of Aphra Behn's *Oroonoko*, which often go by the same name in *The London Stage*, actually represent a variety of responses to and interpretations of the Oroonoko legend [Vareschi and Burkert 2016]. This kind of uncertainty in the historical record, a reminder of our mediated access to the past, is not a concern built into the architecture of the *Information Bank*. Instead, Schneider's database takes an equivalent string of characters (play title, personal name, theater location, etc.) to signify a self-same historical entity.

The *Information Bank* also inherited many of the problems built into its source material—such as, for instance, the lack of reliable premiere dates for plays that debuted before the leading theaters began running daily newspaper advertisements around 1705. Prior to that time, most performance dates in *The London Stage* are largely conjectures based on publication dates and references in published editions of plays [Milhous and Hume 1974]. This kind of guessing results in entries like one for a presumed February 1697 performance of *Timoleon*: "It is not certain what company produced this play, if it was acted; and it may not have been staged" [Avery et al. 1960–1968, 473]. Schneider's database has no way of representing this kind of ambiguity, which is central to how scholars today approach theater history and culture; McGirr, for example, devotes several pages of her book on Cibber to a discussion of the limitations of the available data about performances of his plays as represented in *The London Stage* [McGirr 2016, 10–13]. The *Information Bank's* approach is therefore out of step with today's theories about quantitative inquiry in the humanities, which aim to acknowledge the provenance,

limitations, and fuzziness of the data. While 1970s-era computing may have required researchers to sacrifice complexity in order to conserve memory and storage resources, today we have the capacity to encode more information about the provenance and limitations of data, and many digital humanities projects are actively invested in finding ways to register and visualize uncertainty.

The acknowledgment of ambiguity in the data takes on a particular importance for theater studies, which, as Caplan notes, takes as its objects "incomplete records of performance events rather than the live event itself." For Caplan, databases of theatrical records "tackle a recurring and significant challenge in our field—the ephemerality of our medium and the dispersal of theatrical ephemera that may shed light on a performance event" [Caplan 2015, 356–57]. One prominent example is AusStage, an ambitious database of "programs, ticket stubs, newspaper clippings and so on" that seeks to document all dramatic performances across Australia from 2001 to the present, as well as some additional historical performance events. [24] As Miller points out, projects like AusStage "force us to define performance's ontology," articulating foundational premises such as "the relationship between performances and works, sex and gender identities, and how our contemporary vocabulary for performance translates that of previous eras" [Miller forthcoming]. Building on his experience as a research coordinator for AusStage, Jonathan Bollen has recently published an essay surveying the convergences and divergences of data models for theater history; the comparison raises critical questions about such fundamental issues as the basic unit of inquiry (A single evening's performance? An event lasting several evenings but having a common title and cast? A production that might span multiple locations and years or even decades of runs?). As more theater databases are developed, each must decide whether to articulate its own unique ontology or to strive for interoperability with other databases by adopting shared structures and assumptions about the objects of study [Bollen 2016]. Our revitalization of the *London Stage Information Bank* would likewise need to engage in such critical self-reflection, deciding where to maintain the existing data model and where to refine it, as a contributor to ongoing efforts to develop data models tailored to theater studies.

Ultimately, eighteenth-century literary, theatrical, and historical research would be best served by a new *London Stage* database that finds ways to foreground the situated, captured nature of the data and the layered history of its transmissions and remediations; this is the direction we hope to take the project in the future. At the same time, however, it is important to realize that even our recognition of these gaps in the data is mediated by the ways scholars have accessed and thought about these archival records in the past. As Johanna Drucker and others have reminded us, the history of the data's collection and transmission is an intrinsic part of the object of study; when we query a database of theater records, we actually query a database of composite objects that encode the history of how performance has been seen, recorded, preserved, cataloged and studied. [25] If we aim to acknowledge the messiness that lies behind the apparent solidity of digital or numerical representation, a necessary step is to make visible the projects that have helped to reveal or occlude that messiness over the years, and to unpack the ways they worked to do so. Only then can we design, make, and analyze in ways that resist reproducing unchecked assumptions that went into the collection and curation of existing cultural archives and their past digital remediations.

The reengineered *Information Bank* will not only embody a cautionary tale about digital preservation and provide a revitalized, open-access resource for exploring data about eighteenth-century theatrical performances; just as importantly, it will model and enable a critical approach to the architecture of that data. In doing so, it will align with the interests and research dispositions of today's humanities scholars, who often seek to harness the possibilities of quantitative methods while maintaining a critical stance towards the digital condition. In reflecting these current concerns, the new resource will aim to reach a wider and more invested audience than Schneider's *Information Bank* was able to. Like an eighteenth-century adaptation of an Elizabethan play recorded in the pages of *The London Stage*, this database project performs an act of recovery and revival that is simultaneously an act of re-imagination—one that pushes against the temptation to view the current moment as new and unprecedented, and instead invites its audience to a more layered understanding of its place in larger cultural and historical processes.

Acknowledgments

I wish to thank Mark Vareschi for planting the seed of this project by pointing me to Schneider and Daland's 1971 *Computers and the Humanities* essay; Erin Dix at the Lawrence University Archives for her assistance locating and navigating the records of the *London Stage Information Bank*; members of Schneider's original team, including Cindy Serikaku and Nick Schneider, for helping to reconstruct the history of the project, and especially Will Daland for his involvement with the recovery effort and his careful review of the accuracy of my findings; Derek Miller for writing a script to

repair underlying hexadecimal errors in the recovered data; Todd Hugie for his ongoing effort to reengineer Will's code base; Brad Pasanek for giving me the opportunity to share this work in its early stages at the 2015 MLA annual meeting; Brianna Marshall and the Research Data Services community at the University of Wisconsin-Madison, as well as Bronwen Maseman and her graduate students, for their feedback on this work; Susan Barribeau and Betty Rozum for connecting me with the right people to keep the project going at key moments; Dorothea Salo, Cal Lee, Kam Woods, and Carl Stahmer for patiently answering my media forensics questions; Steven Dast for his help with image processing for OCR; Jack Keel for his assistance compiling and debugging PL/1 code; and Steel Wagstaff, Irene Zimmerman, Angela Moore-Swofford, and Angelina Zaytsev for their help making *The London Stage* available to the public through HathiTrust. This work has been partially supported by funding from the Department of English at Utah State University.

Notes

[1] *Theatre Journal* has recognized and begun attempting to bridge this divide. In 2016, the journal published two special issues on "Digital 'Issues': Rethinking Media in/and/as Performance" (68.3, Sept. 2016) and "Theatre, the Digital and the Analysis and Documentation of Performance" (68.4, Dec. 2016).

[2] Against the view of software as immaterial, Kirschenbaum emphasizes the "material circumstances that leave material (read: forensic) traces—in corporate archives, on whiteboards and legal pads, in countless iterations of alpha versions and beta versions and patches and upgrades" [Kirschenbaum 2008, 15]. These are precisely the kinds of forensic traces I bring to bear on my study of the *London Stage Information Bank* and my efforts to recover its layers of history and inscription.

[3] Jonathan Bollen describes vividly the persistence of data models even as the data is transformed for use in other projects: "datasets have the fluidity of plastic; they flow when prodded, pushed, or pressed, when the heat is turned up and when placed under stress. In the vocabulary of project management, datasets are dumped, scanned, stretched, squeezed, shoved, massaged, cleansed, and washed—in plainer terms, refactored and reformatted—and when all else fails, rekeyed... Inevitably, to migrate data is to manipulate: exports reveal imperfections, imports throw exceptions. And in the process of manipulation, one feels the resistance of the data model. Like a habit to be retrained or plastic's retention of its form, a dataset retains the memory of its model even as it is transformed" [Bollen 2016, 619]. Bollen draws here on his experience coordinating research for AusStage (see Note 24), an ambitious theatrical database that absorbed data from several precursors developed in the 1970s and 1980s.

[4] In collaboration with Southern Illinois University Press and HathiTrust, I was able to negotiate the release of Google's scans of *The London Stage* reference books to the public. You can now search, view, and download all eleven books at <https://catalog.hathitrust.org/Record/000200105>, as well as the *Index* at <https://catalog.hathitrust.org/Record/000299859>. See Burkert 2014 for more information.

[5] The compilers and editors of the original reference series who served as the Advisory Board of the *London Stage Information Bank* were William Van Lennep, Emmett L. Avery, Arthur H. Scouten, George Winchester Stone, Jr., and Charles Beecher Hogan. Additional Advisory Board Members included Allardyce Nicoll, Sybil Rosenfeld, Cecil Price, Philip Highfill, Kalman Burnim, Carl Stratman, John Robinson, and William Armstrong.

[6] The graduate student markup editors were Leonard Leff, Marcia Heinemann, Muriel Friedman, and Mark Auburn. Additional non-specialist markup editors included Devon Schneider, Ben Schneider III, and Dorothy Church.

[7] This workflow was reviewed for accuracy by Will Daland in 2015.

[8] The so-called "ladder system" of abbreviating casts lists was developed by the editors to conserve ink and paper, but it remains a perpetual frustration to users of the reference books. An illustrative example is offered by an entry for Tuesday, June 25, 1708. *The Constant Couple* was performed at the Theatre Royal – Drury Lane with a cast "As at Queen's, 20 Oct. 1707, but Clincher Sr – Pinkethman; Lady Lurewell – Mrs Knight; Parly – Mrs Moor" [Avery et al. 1960–1968, 172]. The researcher turns back 17 pages to the entry for that date and finds a fuller cast list: "Sir Harry – Wilks; Col. Standard – Mills; Smugler – Johnson; Vizard – Husband; Clincher Sr – Bowen; Clincher Jr – Bullock; Dicky – Norris; Lady Lurewell – Mrs Oldfield; Lady Darling – Mrs Powell; Angelica – Mrs Bradshaw". Combining the information from these two entries, it is possible to determine that the cast list for the June 25 performance was as follows: Sir Harry – Wilks; Col. Standard – Mills; Smugler – Johnson; Vizard – Husband; Clincher Sr – Pinkethman; Clincher Jr – Bullock; Dicky – Norris; Lady Lurewell – Mrs Knight; Lady Darling – Mrs Powell; Angelica – Mrs Bradshaw; Parly – Mrs Moor. This illustration involves only two entries for the sake of simplicity. However, in many instances, a researcher must follow the trail back through numerous dates and many actor-role substitutions in order to determine the cast list for a performance of interest.

[9] "Final report on Phase 2 (1972-1975)," retrieved from Lawrence University Archives. The sixteen editors whose names I have been able to gather are Catherine Boggs, Catherine Steiner, Marc Weinberger, Joseph Jacobs, Ruth Steiner, Connie Hansen, Sarah Larsen, Laurie Johnson, Sue Kock, Peter Pretkel, Lynn Seifert, Louise Freiberg, Elizabeth O'Brien, Jan Surkamp, Mark Burrows, and Kathy Rosner. Accounts of the project, including Schneider's books as well as archival documents, vary as to whether the total number of student editors was seventeen or eighteen.

[10] Correspondence retrieved from Lawrence University Archives.

[11] Private correspondence with curator Susan Pyzynski in May 2014 indicated that the tapes may have been lent to a faculty member at Harvard; according to a February 2015 correspondence between Derek Miller and curator Micah Hoggatt, the tapes were lent to the IT department.

[12] The full list of funders includes the National Endowment for the Humanities, the American Council of Learned Societies, the American Philosophical Society, the Andrew Mellon Foundation, the United States Steel Foundation, the Billy Rose Foundation, Lawrence University, and individual gifts from Mrs. John A. Logan, Charles Beecher Hogan, Faith Bradford, Dr. and Mrs. J. Merrill Knapp Jr., and an anonymous Friend of Lawrence University.

[13] The session in question is listed in the MLA 1971 program as Seminar 28, "The Future and Expansion of *The London Stage 1660-1800*: Computerized Information Bank," led by George Winchester Stone, Jr. on Tuesday, December 28. *PMLA* 86.6 (Nov. 1971): 1131.

[14] Correspondence retrieved from Lawrence University Archives.

[15] No documentation of the specific provenance of these discs accompanied them, nor did a search through the IT department or Archive logs turn up evidence of the chain of transmission. It is unclear when or how the data was migrated forward from magnetic tapes to floppy.

[16] Daland explained that the student editors on the project would have corrected the raw source data, after it had been run through the correcting program (ICISCAN) but before it had been run through the parsing program (STRUCTR) and the cast-list-expansion program (LADDR). He compared this to correcting the master of a tape, rather than the copies. For that reason, it remains entirely possible that the data files recovered from the floppy disks at Lawrence contain the data that continued to be corrected throughout the 1970s and was queried to produce the *Index to the London Stage* [Schneider 1979]. (Phone correspondence with Daland, April 2015).

[17] Forensic investigation of disk images of the floppies (using BitCurator) was inconclusive but suggested the hexadecimal errors were not introduced through the process of accessing or copying the data from the floppy disks. I consulted with Carl Stahmer for an outside opinion, and he concurred that the damage most likely happened during the EBCDIC to ASCII conversion that produced the data on the disks, rather than during the process of accessing the disks at a later date (correspondence with Carl Stahmer, April 2016).

[18] Correspondence with Derek Miller, July 2015.

[19] They did, however, keep copies of SITAR and SORTSIT, programs developed after Daland left the project that enabled the data to be continuously updated and sorted (correspondence with Erin Dix, March 2016).

[20] All of the data and programs recovered from Lawrence, along with their documentation, have been deposited with permission in MINDS@UW, the secure supported repository of the University of Wisconsin-Madison. In addition, the scans Daland produced of the printed code base were deposited in the same location, along with all relevant documentation. The collection can be accessed at <https://minds.wisconsin.edu/handle/1793/71768>. This represents an attempt to preserve the project's outputs against the kind of total loss that they nearly suffered once before.

[21] The Performance Archive, based in part on hand counts from *The London Stage*, is available at <http://universitypublishingonline.org/cambridge/benjanson/reference/k/browse/performance>. It is part of the interactive online companion to *The Cambridge Edition of the Works of Ben Jonson Online* (2015).

[22] For evidence of the growing understanding of the need to understand the formative moments of digital humanities, when many projects were begun that exert an invisible but powerful influence on digital humanities work today, see the recent special issue of *Digital Humanities Quarterly* on "Hidden Histories: Computing and the Humanities c. 1965-1985." The editors of that issue, Julianne Nyhan and Andrew Flinn, recently published a related book titled *Computation and the Humanities: Towards an Oral History of Digital Humanities* [Nyhan and Flinn 2016].

[23] For a case study of the dangers of failing to recognize these biases in the process leading to a dataset's creation, see Pechenick, Danforth, and Dodds on the limitations of the Google Books Corpus [Pechenick et al. 2015].

[24] AusStage, "Database Design" and "Data Scope," available at <https://www.ausstage.edu.au/pages/learn/about/database-design.html> and <https://www.ausstage.edu.au/pages/learn/about/database-design.html>.

[25] Here I refer to Johanna Drucker's influential characterization of humanities data as "capta," a term that "acknowledges the situated, partial, and constitutive character of knowledge production, the recognition that knowledge is constructed, taken, not simply given as a natural representation of a pre-existing fact" [Drucker 2011]. Christof Schöch rejects the idea that a new term is needed, but aligns with Drucker in his definition of data within humanities inquiry: "a digital, selectively constructed, machine-actionable abstraction representing some aspects of a given object of humanistic inquiry." This definition importantly draws attention to the added "layer of mediation" created by transforming cultural artifacts into discrete units of digital information [Schöch 2013]. Along similar lines, see Gitelman and Jackson [Gitelman and Jackson 2013].

Works Cited

- Avery et al. 1960–1968** Avery, E. L., Hogan, C. B., Van Lennep, W., Scouten, A. H., and Stone, G. W. (eds), *The London Stage, 1660-1800: A Calendar of Plays, Entertainments and Afterpieces, Together with Casts, Box-Receipts and Contemporary Comment. Compiled from the Playbills, Newspapers and Theatrical Diaries of the Period*. Southern Illinois University Press, Carbondale (1960-68). Available at <https://catalog.hathitrust.org/Record/000200105>.
- Bay-Cheng 2012** Bay-Cheng, S. "Pixelated Memories: Theatre History and Digital Historiography," *American Society for Theatre Research* plenary, Nashville, Tennessee, November 2012. Available at https://www.academia.edu/2131876/Pixelated_Memories_Theatre_History_and_Digital_Historiography
- Bloom et al. 2013** Bloom, G., Bosman, A., and West, W. N. "Ophelia's Intertheatricality: Or, How Performance Is History," *Theatre Journal* 65.2 (2013): 165-82.
- Bollen 2016** Bollen, J. "Data Models for Theatre Research: People, Places, and Performance," *Theatre Journal* 68.4 (Dec. 2016): 615-632.
- Burkert 2014** Burkert, M. "Digital Resource Announcement: The London Stage, 1660-1800." Theater History blog, Association for Theater in Higher Education (19 September 2014). Available at <http://www.athe.org/blogpost/1090217/197683/Digital-Resource-Announcement-The-London-Stage-1660-1800>.
- Caplan 2015** Caplan, D. "Notes from the Frontier: Digital Scholarship and the Future of Theatre Studies," *Theatre Journal* 67.2 (May 2015): 347-59.
- Conway 2010** Conway, P. "Preservation in the Age of Google: Digitization, Digital Preservation, and Dilemmas," *The Library Quarterly* 80.1 (2010): 61-79.
- Davis 2015** Davis, R. C. "Taking Care of Digital Efforts: A Multiplanar View of Project Afterlives," *Modern Language Association*, Vancouver, British Columbia, January 2015. Available at: <http://www.robincamille.com/presentations/mla2015>
- Drucker 2011** Drucker, J. "Humanities Approaches to Graphical Display," *Digital Humanities Quarterly*, 5.1 (2011). Available at: <http://www.digitalhumanities.org/dhq/vol/5/1/000091/000091.html>
- Gadd 2009** Gadd, I. "The Use and Abuse of Early English Books Online," *Literature Compass* 6.3 (2009): 680–92,
- Gadd 2015** Gadd, I. "History of Early English Books Online," *Folgerpedia* (27 March 2015). Available at http://folgerpedia.folger.edu/History_of_Early_English_Books_Online.
- Gitelman and Jackson 2013** Gitelman, L. and Jackson, V. "Introduction" to L. Gitelman (ed), *"Raw Data" Is an Oxymoron*. MIT Press, Cambridge, Mass. (2013).
- Kirschenbaum 2008** Kirschenbaum, M. *Mechanisms: New Media and the Forensic Imagination*. MIT Press, Cambridge, Mass. (2008).
- Library of Congress 2013** Library of Congress. "Sustainability of Digital Formats: Planning for Library of Congress Collections." Washington, D.C. (2013). Available at: <http://www.digitalpreservation.gov/formats/sustain/sustain.shtml>.
- Mak 2014** Mak, B. "Archaeology of a Digitization," *Journal of the Association for Information Science and Technology*, 65.8 (2014): 1515-26.
- McGirr 2016** McGirr, E. *Partial Histories: A Reappraisal of Colley Cibber*. Palgrave Macmillan, London (2016).
- Milhous and Hume 1974** Milhous, J. and Hume, R. "Dating Play Premieres from Publication Data, 1660-1700," *Harvard Library Bulletin* 22 (1974): 374-405.
- Miller forthcoming** Miller, D. "Database and Performance." In N. Leonhardt (ed), *The Routledge Companion to Digital Humanities in Theatre and Performance* (forthcoming).
- Nyhan and Flinn 2016** Nyhan, J. and Flinn, A. *Computation and the Humanities: Towards an Oral History of Digital Humanities*. Springer (2016).
- Parikka 2012** Parikka, J. *What is Media Archaeology?* Polity Press, Cambridge, UK (2012).
- Pechenick et al. 2015** Pechenick, E. A., Danforth, C. M., and Dodds, P. S., "Characterizing the Google Books Corpus: Strong Limits to Inferences of Socio-Cultural and Linguistic Evolution," *PLoS ONE* 10.10 (2015). Available at <http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0137041>.
- Schneider 1971** Schneider, B. R. "The Production of Machine-Readable Text: Some of the Variables," *Computers and the Humanities* 6.1 (1971): 39-47.
- Schneider 1974** Schneider, B. R. *Travels in Computerland; or, Incompatibilities and Interfaces: A Full and True Account of the Implementation of the London Stage Information Bank*. Addison-Wesley, Reading, Mass. (1974).

- Schneider 1979** Schneider, B. R. (ed), *The Index to The London Stage*. Southern Illinois University Press, Carbondale (1979). Available at <https://catalog.hathitrust.org/Record/000299859>.
- Schneider 1980** Schneider, B. R. "The London Stage Project: Its Status and Future." In J. Raben and G. Marks (eds), *Data Bases in the Humanities and Social Sciences*. Ed. North Holland Publishing Company, Amsterdam (1980) 31–34.
- Schneider 1984** Schneider, B. R. *My Personal Computer and Other Family Crises; or, Ahab and Alice in Microland*. Macmillan, New York (1984).
- Schneider and Daland 1971** Schneider, B. R. and Daland, W. "The 'London Stage' Information Bank," *Computers and the Humanities*, 5.4 (1971): 209-14.
- Schöch 2013** Schöch, C. "Big? Smart? Clean? Messy? Data in the Humanities," *Journal of Digital Humanities*, 2.3 (2013).
- Vareschi and Burkert 2016** Vareschi, M. and Burkert, M. "Archives, Numbers, Meaning: The Eighteenth-Century Playbill at Scale," *Theatre Journal* 68.4 (2016): 597-613.