Erin M. Winter. Early English Books Online and the Text Creation Partnership: Applications, Innovations, and Alternatives. A Master's Paper for the M.S. in L.S degree. April, 2023. 51 pages. Advisor: Elliott Kuecker

Early English Books Online or EEBO is a database of more than 140,000 images of rare and early books used by scholars of history and literature for decades. EEBO is a longstanding microfilm product turned image database with established use patterns. EEBO-TCP, a set of 60,000 manually transcribed documents from EEBO, has enabled unprecedented opportunities for large-scale analysis of surviving English print. In this exploratory study, I examine the practices of teaching and research with EEBO and EEBO-TCP in American universities through findings derived from the qualitative coded insights of a focus group of professors and librarians. The focus group discussed teaching methodologies, alternative digital archives, and organizational techniques scholars and librarians have utilized in working with a EEBO and EEBO-TCP. This study is intended to provoke further research into the complex technical mediations underpinning digitization of early printed books and identify areas where academic libraries can facilitate the research process.

Headings:

Early printed books

Digital library use studies

Online databases

Study & teaching of database searching

Early English Books Online and the Text Creation Partnership: Applications, Innovations, and Alternatives

by
Erin M. Winter

A Master's paper submitted to the faculty
of the School of Information and Library Science
of the University of North Carolina at Chapel Hill
in partial fulfillment of the requirements
for the degree of Master of Science in
Information/Library Science.

Chapel Hill, North Carolina

April 2023

Approved by:

_____

Elliott Kuecker

Table of Contents

Introduction

Early English Books Online[1] or EEBO is a ProQuest database consisting of digitized images of 146,000 early books and full-text transcriptions of 60,000 early books first published between 1475 and 1700. The original books photographed for EEBO are dispersed among libraries across the United Kingdom and the United States. Page images in EEBO are digitized scans of microfilm reels originally sold to universities as the Early English Books collection by University Microfilms International. Access to the Early English Books microfilm collection was limited by the amount of time researchers could reserve with shared microfilm readers. At institutions willing to pay a high price, the release of EEBO in 1998 allowed for unprecedented access to scans of early books. EEBO quickly outsold its microfilm counterpart (Bilansky, 2017). Given the financial impracticality of traveling overseas to study surviving editions of rare books, EEBO has been a valuable resource for scholars of British history and literature at thousands of universities in the decades since its release (Gadd, 2009).

The established deficiencies of EEBO's microfilm images include low resolution, poor cropping, and missing or obscured pages, but EEBO is nonetheless a staple product at sufficiently large universities (Kichuk, 2007). Canonical literary figures of the period like Shakespeare, Milton, and Spenser have dedicated digital archives with modern scans,

---

[1] *Early English Books Online* (https://www.proquest.com/eebo/)

but many of the obscure works in EEBO – including plays, pamphlets, ballads, and sermons of historical value – are centuries out of print. For scholars without access to physical editions, EEBO is sometimes the only option for research into certain primary source documents.

EEBO Text Creation Partnership or EEBO-TCP[2] was a collaborative initiative between ProQuest and the University of Michigan that commissioned a selection of 60,000 of the microfilmed books in EEBO for manual transcription to produce highly structured TEI-XML representations of the printed text of each book. This initiative, after a five-year embargo imposed by ProQuest, gave researchers unprecedented access to the full text of tens of thousands of early books too fragile and too irregular to be processed with OCR technology (Mak, 2014). The TCP project is officially inactive. According to the official blog of the Text Creation Partnership, there are no plans at the time for the remaining 80,000 books in EEBO to be transcribed, or at least not from the microfilms held by ProQuest (2018).

The web interface of EEBO was relatively minimal until a recent redesign and rebranding effort by ProQuest, but it now supports search features more typical of a modern academic database (Froehlich, 2021). The free EEBO-TCP browser, however, has primitive search tools intolerant of English spelling variants[3]. Both products have relatively limited, out-of-date metadata on the provenance and subject of each book. The

---

[2] EEBO-TCP: Early English Books Online *Text Creation Partnership* (https://quod.lib.umich.edu/e/eebogroup/)

[3] For example, "vvork" and "work" are treated as different words and return distinct search results in the free EEBO-TCP frontend (https://quod.lib.umich.edu/) but not in EEBO (https://www.proquest.com/eebo) as of December 2022. Note that EEBO-TCP texts can be searched in the EEBO interface, but only by users at institutions that EEBO subscribe to EEBO.

aforementioned factors can hinder the process of teaching and research with

EEBO and EEBO-TCP.

It falls on professors and librarians to bridge the gaps in image quality, document

retrieval, and contextual metadata, whether through LibGuides, classroom instruction, or

dedicated workshops (Froehlich, 2019; Burke, 2021). Conversely, some institutions have

chosen to pursue alternatives to EEBO through digitization of specialized subcollections[4],

placing images in the hands of libraries rather than in the hands of vendors (Green, 2014).

For scholars of manuscripts, music, maps, and letters of the periods encompassed by

EEBO, the best digital resources for their work may be specialized subcollections, as

EEBO collects printed, published works rather than handwritten materials. However,

none of the free alternatives dedicated to books printed between 1475 to 1700 come close

to the size and scope of EEBO or EEBO-TCP.

In this study, I explore the practices of teaching and researching with EEBO and

EEBO-TCP at American universities. Though analysis of the findings of a small focus

group of professors and librarians, I seek to identify the supplementary layers of

organization, contextualization, and computational techniques researchers and librarians

have practiced while working with digitized microfilm images in an evolving community

of digital archives.

---

[4] See UCSB's *English Broadside Ballad Archive* (https://ebba.english.ucsb.edu/) for an example of a free,
digitized alternative resource limited to a small, specialized corpora.

Literature Review

The History and Intended Use of EEBO

The history of the Early English Books Online database has been thoroughly traced and interpreted in prior case studies by Gadd (2009), Mak (2014), and Gavin (2017). Scholars from the disciplines of literature, information science, book history, and bibliography have critiqued EEBO in the decades since its inception. As the complicated state of EEBO as it exists today can be traced to its origins and previous forms, I will summarize the history of EEBO here.

Decades before the technological infrastructure required for institutional repositories of digital books existed, a Michigan entrepreneur named Eugene Power started a business called University Microfilms in 1938. University Microfilms processed rare books from universities across the United States and abroad to create microfilm reels that could be sold to American universities that couldn't acquire the physical books (Gadd, 2009). Power's original intent was to create print-on-demand facsimiles of rare books to resell, but this initial business model was never widely deployed for the books in EEBO. Instead, motivated by the escalation of World War II in Europe, the American government gave Eugene Power financial backing to microfilm and preserve books in British libraries in 1940 (Gavin, 2017). It was these indexed microfilms that proved most interesting to scholars.

Eugene Power relied on the metadata scheme of the Short-Title Catalogue or STC to identify candidate books for microfilming. The STC is an extensive bibliography developed by the British Library that attempts to represent the printed record of what is now the United Kingdom and its former colonies between 1473 and 1700 (Mak, 2014). By repurposing the STC, Power garnered federal government and institutional support in 1940 for "microfilming select books, chosen for their likely research interest to American scholars" (Gavin, 2017, 84). American libraries didn't have collections of rare British books as extensive as their European counterparts, so microfilm reels of books from the STC made the collection marketable to American institutions.

The Early English Books collection was first sold as a large set of indexed microfilm reels, then as a collection of CDs, and finally as an online database subscription: Early English Books Online (Gadd, 2009). Without digitized texts and databases of scanned books, convenient access to primary sources would be a more common problem for American scholars and students of the period. An early survey-based study of EEBO users found that faculty and graduate students who had previously relied on the Early English Books microfilm collection "placed the convenience of working from any computer with Internet access at a high premium" (Lindquist & Wicht, 2007, 352). The survey also found academic librarians and other faculty members to be the most popular sources of research assistance when using EEBO.

Nonetheless, crucial features like full-text search had not been implemented at that time, and undergraduates sometimes struggled to use the keyword search tools EEBO had available at the time (Lindquist & Wicht, 2007). To search images with

complex, irregular type that cannot be recognized through optical character recognition or OCR tools is not a straightforward task. In fact, it is difficult enough that a panel of institutions commissioned human transcriptions of the documents of EEBO. Now that EEBO-TCP is complete and integrated into EEBO, it is possible for users to perform full-text searches on a special subset of 60,000 EEBO documents. In a 2013 survey of EEBO users, academic respondents described using "both images and full text" as their preferred method for reading books in EEBO (Siefring & Meyer, 14). For the remaining 80,000 documents without full-text transcriptions, a basic search will match spelling normalized keywords (documents with "VVilliam" as an author are matched to searches for "William") in title, author, description, and other metadata fields. These usability improvements in the modern EEBO interface make the database more approachable to new users.

The ability to search over full-text and bibliographical material in a web browser not only made traditional primary source research easier but also changed the methods by which scholars conducted research. A 2017 study found practitioners used EEBO and comparable primary source databases "to skim…to gain general background, to understand ideological and linguistic trends of the period" (Bilansky, 517). This is a drastic shift from the paradigm of the rare book reading room, where a scholar must pore over a rare book page by page in pursuit of a citation. Bilansky observes that "given the ubiquity of databases in literary scholarship, it makes sense that learning databases is…part of the acculturation of scholars and the learning of their trade" (520). Consciously or otherwise, scholars of the "EEBO generation" have developed research practices with the affordances of database search in mind. A 2008 panel of Early

Modernist scholars claims that digital archives like EEBO reveal "a more

complete discursive system, bringing broadsides, ballads, royal decrees, poetry, short

fiction, and prose narratives of all sorts into the critical purview" (Crowther et al., 6).

This increased interest in peripheral or underrepresented documents coincides with the

rise of new historicist criticism in literary scholarship, and, ultimately, the classroom.

      EEBO is intended for use as a primary source research tool for researchers,

professors, graduate students, and upper-level undergraduates. Nonetheless, EEBO and

EEBO-TCP have been successfully integrated into the graduate and undergraduate

classroom for decades. Crowther et al. argue that the scale of EEBO can "leave

instructors insecure in their own authority (not to mention technical prowess) in the

classroom…despite these risks in bringing EEBO to the classroom, the payoffs for

students can be remarkable" (2008, 9). A survey respondent in a 2013 study attested:

"[EEBO] was important for my undergraduate thesis…consulting the copy in the

Huntington Library as an undergraduate was not feasible!" (Siefring & Meyer, 20). For

undergraduates, who do not have the means to travel for research purposes, EEBO was

sometimes the only practical way to conduct research on books held in private libraries.

      A novel pedagogical exercise described by Burke has students compare a printed

Mary Wroth sonnet as available in EEBO to a manuscript version of the poem from the

LUNA archive at Folger Shakespeare Library; a word that is ambiguous in the EEBO

version is clear in the digitized authorial manuscript (2021, 142). This approach eases

students into bibliographic practice and the materiality of text through a vivid

demonstration. Professors and librarians have realized that digital archives like EEBO,

especially when supplemented by visits to physical archives, can be invaluable

tools for teaching the history of early printed books.

EEBO-TCP as a Quantitative Research Tool

The Text Creation Partnership, an independent nonprofit, started in 1998 as a

transcription collaboration among the University of Michigan, the University of Oxford,

and ProQuest, who maintained ownership of the original University Microfilms reels

when the company was bought out by Xerox (Gavin, 2017). The EEBO-TCP framework

can be distinguished from those of other digital libraries of transcribed books (Project

Gutenberg, Perseus) as follows: it is the result of a private-public partnership between a

private company and several universities, the painstaking manual transcriptions were

outsourced rather than volunteered, and every keyed file in EEBO-TCP directly

corresponds to a set of microfilm images in EEBO, a paid product. Like Project

Gutenberg and Perseus Digital Library, EEBO-TCP was ultimately intended to

democratize access to printed materials in the public domain.

According to a 2012 assessment of the project, initial funding was split between

ProQuest – which contributed 20% of the costs of the first phase of the project – and

more than one hundred partner libraires, which contributed the remaining 80% of the

costs (Welzenbach). Partner libraries had to pay a $50,000 introductory fee to opt into the

project. Early in the process, academic librarian and founding member Mark Sandler was

quick to credit the early success and interest in the project from other institutions to

ProQuest's participation in the endeavor (Sandler, 2004). However, ProQuest's

participation had an important stipulation: the transcribed documents were to be released

in two phases, with the institutions that backed the TCP (large research institutions that

could front the membership costs) getting early access to the second half of the

transcribed documents. Scholars from smaller institutions did not have full access to the

embargoed EEBO-TCP transcriptions until as recently as 2020.

Images in EEBO were transcribed into a dialect of XML called TEI, which

specifies that the XML files be organized into nested tags. TEI tags are designed to

mimic the structure of the printed page and to differentiate groupings of prose, verse,

annotations, and marginalia (Gavin 2019). TEI is a product of the Text Encoding

Initiative, a bibliographic consortium that produces sets of guidelines for digital encoding

of scholarly manuscripts. The TEI consortium was founded in 1987 (Mah et al., 1997)

before initiatives like EEBO-TCP were underway, and the guidelines are sometimes

ambivalent regarding the best option for representation of the complexities of

representing fragmented, damaged, or abbreviated documents. Therefore, TEI was

adapted to the needs of the editors of the TCP, and specialized practices for keying and

encoding EEBO images into XML emerged[5].

Having the TCP editors at the University of Michigan and Oxford scrutinize every

keyed page of the transcriptions produced by vendors at overseas firms was infeasible, so

TCP editors performed quality control by "proofreading a 5% sample of each book, letter

by letter… sending back those that do not meet the 99.995% accuracy specification back

to the keyers" (Welzenbach, 2012, 3). If the sample met standards, editors quickly

checked the tags and structure of the XML file against the divisions and pages within the

---

[5] Remnants of the detailed encoding instructions distributed among editors are preserved on the Text Creation Partnership website. XML encoding practices were debated, adapted, and decided by consensus. https://textcreationpartnership.org/docs/dox/instruct3.html

original book. Inevitably, mistakes in the XML file generation process slipped

in from both teams.

Scholars like Mah have argued for decades that the SGML or XML representation "transcription is thus, like editing, not only an act of interpretation, but an act of expressing interpretation usefully" (1997, 45). This applies even if the transcription process is dictated by detailed coding guidelines, as the complexities of medieval and renaissance orthography make transcription more than an act of translating words to a text document. The leaders of EEBO-TCP also made a crucial decision regarding document materiality when encoding and keying the EEBO images: they chose to exclude "all non-textual information and all handwritten material" from the EEBO-TCP XML files (Siefring & Willcox, 2012, 87). While focusing on textuality over materiality is a valid epistemological approach, as rare books often bear anachronistic ink and pencil markings from 20th-century libraries, there is no way to recover textual intent from microfilmed images of a book printed hundreds of years ago with complete confidence.

There are too many illegible letters, words, and even paragraphs for a transcriber to be completely confident about what was printed, much less what was *intended* to be printed, for all works in EEBO. Rather than have the transcribers, who were contracted offshore workers without domain knowledge, guess at missing or damaged characters, EEBO-TCP files make use of the <gap> XML tag. Transcribers were instructed to mark gaps in their transcriptions by guessing the extent, whether a character, word, or section of what was illegible (Mueller & Burns, 2016). A 2016 study by Mueller and Burns over a sample of 44,000 EEBO-TCP files estimates that "one out of 200 words is transcribed incompletely or not at all…The errors are not randomly distributed, but cluster heavily by

books and particular pages in those books" (155). That is, the average error rate

across the corpus is higher than the rate experienced when reading an arbitrarily selected

text because some microfilm images in EEBO are especially bad, and some books are

exceptionally damaged or illegible. Incomplete transcriptions should not dissuade

interested users from exploring EEBO-TCP, but they should be kept in mind when

computing linguistic analyses over the corpus.

The EEBO-TCP transcriptions were placed under a five-year embargo before they

could enter the public domain. The first batch of 25,000 transcription files was released in

2010 to partner institutions, but the files weren't available to the public until 2015. The

second batch of 35,000 transcription files was released in 2020, so much of the earlier

research conducted using EEBO-TCP is based on less than half of the XML files now

available in the public domain. As the largest free corpus of transcribed early English

texts printed before 1700 available, EEBO-TCP occupies a singular position in the Early

Modern corpus linguistics (Basu, 2018). EEBO-TCP has obvious limitations in its

presentation of illustrated, damaged, or tabular texts for users without access to EEBO,

but there are dozens of projects dedicated to the evolution of the printed English language

made possible only by the scale and precision of EEBO-TCP.

Critically Evaluating EEBO

In Bonnie Mak's critical archaeology of EEBO, she observes that the EEBO

database in its current form elides both the composite nature of early books – which were

sometimes annotated, rebound before being preserved within archives – and the digital

layers of mediation separating the book as perceived in the interface from the physical

artifact (2014). She claims that "readers are thus encouraged to overlook as

inconsequential the material history of the scans that is graphically registered in

the digitizations themselves" (1519). This is because scans on EEBO, whether bitonal or

grayscale, are digitally compressed photographs of microfilmed photographs of books,

not photographs of a book. The intermediary stages of labor by librarians,

transcriptionists, bibliographers, and web programmers are concealed from the user.

Moreover, the process of converting microfilm reels to digital images is not as

straightforward as photographing a book. Each additional layer of processing in EEBO

has introduced opportunities for human error. As observed by Lesser in his descriptions

of flawed EEBO images, some of the *many* visual defects of the images in EEBO are

artifacts of the messy microfilm digitization process:

> The reel was run through a digitizing machine, and each of the page openings
> filmed on it was then stored as a digital file separately from its neighbors.
> Evidence of that process can occasionally be seen in EEBO, almost always (as
> with bibliographic clues generally) where it has gone wrong … the film was not
> properly advanced from one scan to the next … [or] the data became corrupted,
> resulting in digital static. (2019, 4)

For works deemed notable enough to be rescanned at their respective archives, their

EEBO facsimiles will inevitably be superseded by copies without defects from

microfilming or the microfilm digitization process.

Researchers and librarians have also criticized the dependence of public research

of centuries-old documents on the whims of a commercial, for-profit database provider

like ProQuest. These warnings came to pass when the Renaissance Society of America,

an independent scholarly organization, had its institutional subscription to EEBO revoked

by ProQuest (Overholt, 2015). An extract from the email sent to society members by the

RSA Executive Committee is reproduced below:

> The basis for the cancellation is that our members make such heavy use
> of the subscription, this is reducing ProQuest's potential revenue from library-
> based subscriptions. We are the only scholarly society that has a subscription to
> EEBO, and ProQuest is not willing to add more… (Overholt, 2015)

Though the Renaissance Society of America's access to EEBO was quickly restored in the wake of public outcry on Twitter from the scholarly community, this incident demonstrates the need for alternative methods for access of the primary source documents in EEBO. Independent researchers or scholars from institutions unable to afford subscriptions would have been the most impacted by this decision had ProQuest gone through with their threat, with their ability to conduct primary source research hanging in the balance. Moreover, larger institutions with current access have also explored ways to build resilience to the whims of private companies, who cannot be trusted to always behave in the best interest of scholars.

Critically Evaluating EEBO-TCP

The most salient critique of EEBO-TCP is its separation from the digitized microfilm images of EEBO. Though all EEBO-TCP transcriptions are now in the public domain and readily available for download, the images from which they were transcribed are exclusive to scholars at institutions with an EEBO subscription. In 2012, James Cummings praised the potential for scholarly editing and editorial transparency in encoding medieval and classical manuscripts in as digital editions in TEI format (2012, 69). Cummings, however, envisioned digital editions where page by page links were embedded in the SGML/XML files to connect transcriptions to their corresponding page images. EEBO-TCP takes advantage of this feature to inject EEBO URLs into EEBO-

TCP page sections, but the embedded links point to paywalled image files[6] that require an institutional login. This restrictive access is contrary to the vision of open digital editions envisioned by early digital humanities scholars (Miller, 2007). Even a project backed by multiple R1 universities cannot be trusted to continue in perpetuity without stable funding.

The EEBO-TCP project gradually exhausted its funding, but the rigor and effectiveness of its processing pipeline could serve as lessons for future generations of scholars and archivists. The digital humanist Shawn Martin proposed that unlike many scholarly projects, the TCP was organized in a "distributed model in cooperation with over one hundred libraries…TCP works with scholarly projects throughout the world in which the projects use TCP as a base" (2014, 48). The typical scholarly project is organized vertically, with a concentrated group of faculty, staff and students taking ownership of the project management, distribution, storage, and infrastructure. Under Martin's distributed model, smaller scholarly projects would rely on the TCP and its infrastructure as a foundation. Martin calls for "libraries to band together to improve a corpus of items… rather than spending millions of dollars to build the same basic level of infrastructure over and over again for different scholarly projects" (2014, 57). The question is, can these smaller archives be sustained without restrictions to access or impositions of embargos?

---

[6] Tested December 4th, 2022.

Alternatives to EEBO and EEBO-TCP

EEBO may be the biggest database dedicated to early English books, but a universities and public libraries have begun to release scans of their rare books in the public domain to the Internet Archive. Modern scans created on equipment designed for digitizing rare books almost always outperform the bitonal images on EEBO, even to the eyes of a layperson. In a usability study comparing editions of books available on both EEBO and the Internet Archive, EEBO had a more thorough collection of antique books than the Internet Archive. However, subjects found EEBO's page images were of inferior legibility and quality (Brightenburg, 2016). This is to be expected given the differences in sensitivity and resolution of a contemporary book scanner as compared to digitized microfilm, but it is worth acknowledging how the differences in quality impact research and instruction.

ProQuest overhauled the EEBO interface in 2020, rendering some of Brightenburg's figures obsolete (Froehlich, 2021). The small studies of EEBO and EEBO-TCP usage by undergraduates and faculty members conducted by academic libraries were published before the changes to the EEBO interface and the public release of EEBO-TCP Phase II (Siefring & Meyer, 2013; Brightenburg, 2016). Therefore, I feel there is a significant gap in library science research of the roles and behaviors inherent in research and instruction with EEBO and EEBO-TCP as it stands today. The TISDR benchmarking study conducted in 2013 found that, sampling participants who heavily used EEBO, less than half knew what EEBO-TCP was, much less where to access it (Siefring & Meyer, 2013, 16). Textual extracts of the EEBO-TCP transcripts are

integrated into the EEBO interface underneath page images, so EEBO's core

users now benefit from EEBO-TCP whether or not they know about the project by that

name.

From what I can tell, no formal studies of EEBO and EEBO-TCP have been

conducted in the decade since the above studies were published, as the Brightenburg

study was an experiment rather than a community survey. While this case study cannot

replicate the scope of a large-scale scale research initiative, I believe it surfaces

generalizable interests and concerns in a growing ecosystem of digitization projects of

early printed materials. I would also like to evaluate, in ways that prior studies could not

have anticipated, stakeholder awareness of the availability of alternative EEBO-TCP

corpora like EarlyPrint, an active joint project between Northwestern University and

Washington University in St. Louis which "aims to create a deduplicated digital library of

most English books published before 1700" (Williams, 2022) just as EEBO-TCP did. As

these alternatives weren't available at the time the studies mentioned above were

conducted, I feel there is a sufficient gap in the literature to justify this study.

Research Questions and Key Terms

This exploratory study was intended to capture trends in contemporary usage patterns of EEBO and EEBO-TCP that may not have been observed in previous literature. I examine with particular attention the continuously evolving layer of supplementary tools, interfaces, and bibliographical data that has developed to enhance a relatively static database product, as well as alternatives to EEBO and EEBO-TCP that have emerged in the past decade. The following research questions informed focus groups discussion questions and moderation decisions:

1.   How are EEBO and EEBO-TCP used as research and teaching tools at American universities? What additional materials are used to supplement database images when teaching with EEBO?

2.   Are institutions exploring recently developed alternatives to EEBO and EEBO-TCP for accessing Early Modern English texts? How viable are the alternatives for exploring and analyzing early English books that fall outside of the literary canon?

3.   What are the perceived barriers to instruction with EEBO and EEBO-TCP? Are there barriers of usability? Are there financial barriers to access?

Given the proliferation of acronyms pertaining to the Early English Books Online project and its derivatives, I have defined the following key terms below.

- Early English Books Online (EEBO) – An online database product of tens of thousands of Early Modern English texts including transcriptions, original microfilm images, and metadata from the ESTC catalog.

- ESTC – A monumental early 20th-century effort in bibliography to catalog all extant printed English texts from 1550 to 1700. Only relevant so far as it is the source of valuable metadata in EEBO and EEBO-TCP and that it was used to select books for scanning.

- Early English Books Online Text Creation Project (EEBO-TCP) – A separate initiative, released in stages, of transcriptions in TEI-XML format of 60,000 selected works from EEBO.

- Text Encoding Initiative (TEI) – A bibliographic consortium that produces sets of guidelines for SGML and XML document annotation style and structure. The TEI-P3 iteration guidelines were used to produce EEBO-TCP.

Methodology

This study sought to explore the use of EEBO and EEBO-TCP as research and pedagogical tools within American universities, the techniques and contexts – both historical and emerging – of their uses as teaching tools, and instructor awareness of alternative digital resources for access to Early Modern English texts. In response to recent initiatives towards open access research material in academic libraries, I also assessed participant attitudes towards free digital archives of early printed material – which are generally narrower in scope than databases like EEBO – in focus group discussions and survey questions.

The focus group approach, which was previously used by Bodleian libraries to "benchmark" EEBO and EEBO-TCP (Siefring & Meyer, 2013), allowed me to collect detailed information about how EEBO and EEBO-TCP are used as research and instruction tools, as well as how they are perceived by the academic community. I recruited five participants, who, due to scheduling conflicts, were divided into two micro focus groups that were conducted remotely over Zoom.

In a semi-structured interview format, both focus groups were given items from the same pool of questions, but I moderated with openness towards new topics that emerged organically during discussion. Both focus group sessions were automatically transcribed, and I performed qualitative coding on the edited versions of the transcriptions. Participant familiarity with EEBO, EEBO-TCP, and comparable

alternatives to EEBO was assessed through a short Qualtrics survey. The

survey also allowed me to quickly assess which attributes of a database of early printed

books – namely image quality, latency, open access, and metadata – were most important

to the experience of research or teaching with digital archives.

Positionality / Researcher Role

I previously collaborated as a graduate student assistant with two professors on

unrelated projects involving the study of EEBO-TCP using quantitative applications and

natural language processing tools. I was inspired by this work to explore applications of

EEBO and EEBO-TCP at academic institutions in a broader context. I have more

familiarity with EEBO and EEBO-TCP than the average library science student

specifically because of my prior work in corpus linguistics, so I was careful to recruit

participants from outside of the small circle of librarians and professors with whom I had

formerly affiliated. This was done to ensure that study findings could generalize to EEBO

and EEBO-TCP as they are generally used by academic institutions, rather than in the

context of highly specialized quantitative applications. Finally, I have no affiliation with

ProQuest or allegiances to EEBO as a product, nor is this study intended to inform or

endorse any commercial product.

Research Participants

Participant recruitment was selective, given the complexity of the task being

researched. Between five and eight participants are recommended for focus groups of

experts, so three literature professors and two liaison librarians were ultimately recruited

to the study. This smallness can be just by the high level of participant expertise in the

topic, as smaller focus groups are preferred in studies of the behavior of experienced

users (Krueger & Casey, 2015, 198). A single five-person session would have been preferred, but, when no consensus time was found, micro groups were formed such that at least one librarian and at least one professor was present in each group. The sessions of three and two participants respectively were held five days apart.

EEBO is a database product with a very specific target audience: university staff, faculty, and students conducting research in the history, language, and context of English printed materials at research institutions willing to pay the steep costs of access. The focus group sampling frame was a small, disproportionately educated community with a bent towards professors in literature, history, or linguistics departments, and academic librarians supporting those departments. In gathering this focus group, I was conscious of the bias inherent in this sample.

Given the exploratory intent of this study, I relied on recommendations from colleagues to find professors and librarians within the narrow target demographic to which my research questions were applicable. Only professors and librarians who have used EEBO or EEBO-TCP were eligible for the study. Initial participants were recruited through email and encouraged to give out the names of academic colleagues performing similar research.

Data Collection Methods

This study utilized two data collection methods: a focus group session and a short Qualtrics survey delivered during the participant recruitment period. Participants that did not complete the survey before the focus group session were sent a reminder email, and the survey was completed by all participants. To preserve anonymity, participant email addresses were not stored in Qualtrics results. The short survey established

demographics, years of experience and relative familiarity with EEBO and its

alternatives, and primary EEBO/EEBO-TCP use cases for all participants. Given the

semi-structured nature of the focus group session, this survey was necessary to

contextualize focus group responses.

I moderated the focus group discussions, making sure all participants have

opportunities to share insights, and creating just enough structure that discussion touched

on both EEBO and EEBO-TCP. I took advantage of UNC's institutional Zoom

subscription to record the call to cloud storage before archiving it on SharePoint. After

downloading the videos to my local machine, I used the automatic captions generated by

Zoom as a starting point for the construction of transcriptions in Atlas.ti.

Data Analysis Methods

The focus group sessions used Zoom's automatic captioning transcription feature

to generate approximate captions. These rough captions served as a template for manual

transcription correction and formatting. I used the video and transcript linking feature in

Atlas.ti to correct the focus group transcripts against recorded footage. Transcription

correction was necessary because Zoom's voice recognition features struggled to

interpret acronyms like "EEBO" and "EEBO-TCP" correctly. Once the transcripts were

cleaned and anonymized, I qualitatively coded focus group transcripts in Atlas.ti using a

grounded theory approach in order to identify key themes, and then I organized these

themes into subheadings for further discussion.

Though I guided the topics pursued in the focus groups to address underlying

research questions, the codes are a grounded product of the transcripts and not of my

preconceived knowledge about EEBO and EEBO-TCP. After the transcript codes were

finalized and tabulated, I used the transcripts to generate pertinent pull quotes

from the focus group sessions as example instances of each code for the final paper.

Given the small survey sample, I have avoided using or presenting conventional

quantitative analysis methods for survey results. Instead, I have treated the survey results

as an additional set of qualitative data to supplement and contextualize the focus group

transcript findings.

Findings

The focus group participants consisted of three professors and two academic librarians. Each participant worked at one of four large American universities with institutional access to EEBO. Participants collectively reported performing research and instruction tasks with EEBO materials spanning the wide chronological breadth of the historical documents represented in the database – 1475 to 1700 – in the supplementary survey. Two participants reported working primarily with works from the Early Modern period when using EEBO, another with the works of Restoration writers like John Dryden, and one with philosophers like Bacon and Locke. This breadth of interests was sufficient to unearth some of the limitations of EEBO in terms of its offerings to scholars of different historical periods.

EEBO as Ubiquitous in the Early Modern Research Community

While at least some experience with EEBO was an eligibility criterion for participation in the focus group, and therefore every participant had experience with EEBO, all participants reported long-standing familiarity with the product. Younger scholars tend to have little or no experience with the microfilm originals that made up Early English Books, but knowledge of – at minimum – Early English Books Online seems ubiquitous among English professors of the period and English liaison librarians alike. In the survey results, focus group participants reported an average of 11 years of experience with EEBO and a minimum of 5 years. In response to a hypothetical question

about teaching users well-versed with the period but not with databases, an instruction librarian explained: "Sometimes I work with senior faculty, but I think EEBO's been around so long that anybody interested in the Early Modern period has already been using it." For scholars of historical periods that fall between 1475 to 1700, EEBO is a research fixture.

Referencing another Early English Books product, a librarian mentioned seeing a handful of books printed from Early English Books microfilm images from the brief print-on-demand phase of University Microfilms International in a university library. One English professor had experience with the Early English Books microfilms themselves. "In the early 2000s, I realized if I went back to the original microfilm, I could get a much clearer and crisper picture of the page," they explained. Before publishing a dissertation that referenced books in EEBO, they had spent time comparing key primary source passages from EEBO images against the originals at a rare book archive to make sure their assertions about the printed text were correct.

Attitudes towards the substitution of digital facsimiles for printed materials in academic writing have shifted in the intervening decades. A 2012 survey of EEBO-TCP users – a technically-savvy population of EEBO users comfortable with XML files by necessity – found that nearly 97% of the respondents encouraged their students to use online resources like EEBO and EEBO-TCP, but 34% instructed their students to cite print copies of digitized materials when writing papers and journal articles (Blaney & Siefring, 2017). That is, the modern academic community seems confident enough in some digital facsimiles of early printed books to conduct research with them without

checking against physical copies, but the citation of print copies of rare books is still preferred by some publishers.

With regards to usage of alternative or complementary digital archives mentioned in the survey, EEBO-TCP was the most popular, with two professors and a librarian reporting having conducted research with it before. Between focus group discussions and survey comments, two professors and one librarian reported having used the LUNA archive of the Folger Shakespeare Library recently. The only alternative that was as familiar to participants as EEBO, in terms of project recognition and long-standing familiarity, was The Internet Archive.[7] The Internet Archive is an unusual case in that it is a general purpose digital archive of public domain materials that happens to have some digitized early books from the period, but the image quality, metadata, and selection of these books depend on the donor institution (Brightenburg, 2016). This irregularity with regards to early printed books could explain why only one professor reported using The Internet Archive as a research tool despite its widespread awareness among participants.

Image Quality Problems with EEBO

Image clarity in digital book facsimiles is especially important when distinguishing between ambiguous characters in degraded type, and scholars have written volumes justifying interpretations of textual cruxes created by ambiguous words at key junctures in Shakespeare plays. For several books on EEBO, image quality problems created by dated image compression standards and botched microfilm processing make them inferior in legibility to the microfilm originals to this day (Lesser, 2019). ProQuest

---

[7] The Internet Archive, https://archive.org/

does not distinguish between bitonal (dated from 1998 to 2011) and grayscale

(dated from 2012 to the present) microfilm scans in its public-facing EEBO metadata[8], so

I cannot estimate how many of images of the more than 140,000 books are of inferior

quality. A professor of English explained the difficulties poor image quality creates when

teaching EEBO to novices of the study of early printed materials:

> You can't always tell if something is a woodcut or an engraving. The muddiness
> of the microfilm is really bad, so they're more of a starting point. I would never
> recommend students to go on to use them as a primary source for bibliographical
> research.

For research into the decorative elements of books like woodcuts and engravings, EEBO

is understood to be insufficient, as these problems were documented in the early studies

of Kichuk (2007) and Gadd (2009). However, another English professor asserted that

they were confident in their ability to read the textual contents of EEBO images. For

scholars using EEBO for contextual research, text – whether from an EEBO image or

transcribed EEBO-TCP surrogate – is often enough.

To compensate for EEBO's shortcomings in image quality, the professors and

librarians of the focus group had long-standing routines for getting access to better

quality digital images when EEBO was insufficient. A librarian suggested "requesting a

version with another [modern] scan through interlibrary loan…there are a lot of ways to

try to get a better digital reproduction if that's necessary." Given that all participants

worked at institutions in or near cities with access to dedicated rare book libraries, two

professors recommending sending students to request a comparable edition at a nearby

---

[8] See the "Updated Title List" download link under the "Yearly Title Releases" section of the EEBO
Content LibGuide, as of April 2023. https://proquest.libguides.com/eebopqp/content

physical archive, whether in-person or via an interlibrary loan request.

However, as one librarian explained, for many users, the image quality on EEBO (and

supplementary transcriptions where available) is sufficient "especially if you're looking

more for the text and content than the visual aspect."

EEBO and EEBO-TCP Workflows

No two participants had identical research or instruction workflows when working

with EEBO and EEBO-TCP. For professors conducting research or collecting primary

source materials, simplicity and robustness in an interface was preferred to the "shopping

cart" model practiced by EEBO-TCP – which was entirely unused by the participants

who used EEBO-TCP – or the "Save to My Research" feature in EEBO, which was

ignored by researchers in favor of saving images to local disk or cloud storage.

As much as ProQuest has tried to integrate database search and document

selection across its hundreds of products into a single interface, all three professors

reported seeking out documents from other digital archives or even personal photographs

of items in special collections to supplement their work. It is unlikely a single parent

product would ever serve all their research needs.

Research of early books and historical documents is a process of collation, so

document organization was extrinsic to EEBO and ProQuest for all participants. One

English professor explained the difference between how they organize digital images

from EEBO and cell phone images taken at physical archives when conducting research:

> For EEBO specifically, I download PDFs and keep them in a folder with other
> research documents. The only way I've hit on for organizing research photos
> taken in person is to use the call number as a folder, then upload them from my
> phone to Dropbox. Then, I organize the files by research project, or sometimes by
> holding institution, if they're not related to a project.

Another professor, under a simpler scheme, organized their primary source

images in Word documents by the date they were either photographed or retrieved from

the internet. On the most technical end of the spectrum, a professor specializing in

quantitative methods circumvented the EEBO-TCP interface entirely by downloading the

60,000 XML files in bulk, saving them to disk, and using R to load documents into

memory as needed.

The librarians, however, did not report compiling subcollections of EEBO or

EEBO-TCP documents because they tended to support rather than conduct long-term

research projects. Instead, they were more likely to give organizational suggestions to

undergraduate students learning primary source research tools like EEBO. One librarian

mentioned teaching students "how to create the My Research Account [on ProQuest]" but

also how to export EEBO PDFs and citations for use in citation management tools like

Zotero.

Teaching with EEBO and EEBO-TCP In and Out of Libraries

With regards to instructing students or faculty with EEBO and comparable digital

archives goes, four survey participants reported teaching with EEBO, three with EEBO-

TCP, and two with general purpose digitized book archives like The Internet Archive and

HathiTrust. For teaching undergraduates, the professors reported using archives with

better scans like Folger's LUNA or Penn's Colenda to demonstrate example primary

source documents. However, EEBO is very much in scope for graduate students of the

Early Modern Period.

An English professor described teaching a rare books workshop for graduate

students by "bringing out the physical STC books when I teach so I can talk about the

history of how that catalogue developed…the metadata about reel position, and how that connects to the original microfilm." Teaching EEBO alongside its complicated bibliographic history recalls Gadd's argument that the "canny user of EEBO needs to be conscious of…the origins of its bibliographic data" (2009, 687). Given that the idiosyncrasies of EEBO can be traced to the microfilm era and the STC, learning to use EEBO proficiently can serve as a lesson in the history of bibliography, information science, and digitization.

As a paid database product, access to EEBO is typically locked to either the IP address of a subscribing institution or locked to the SSO credentials of students and faculty of a subscribing institution. The two librarians reported that they had not experienced problems of access to EEBO because they primarily worked with faculty and students *within* an institution that subscribed to EEBO. An English professor who has led rare book classes as a visiting instructor explained that when teaching workshops at other academic institutions, EEBO access is provided to visiting scholars through onsite access or through temporary credentials given out by the host library. None of the survey participants reported difficulties with EEBO access when collaborating with faculty and graduate students at other universities.

However, the EEBO paywall can become problematic when teaching populations outside typical academic institutions. One librarian mentioned a slight caveat, "maybe some high school students [would have difficulty], but if they're on campus, they have access to our resources." When working with populations who never have EEBO access, teaching with materials featured in EEBO requires workarounds. A professor of English

who conducts outreach with local prisons described their solution when

working with incarcerated people:

> I've done archival work while teaching in prisons. They don't have *any* internet access…I've chosen a text, printed five or six pages, made photocopies, and taken that in. My workaround has been taking in the actual books [from local archives] and letting them see the original book instead as much as possible.

Though this is an atypical context, it is worth noting the limitations of paywalled digital

archives when working with the broader community.

Non-Archivists and Digital Archives in the Post-Pandemic Age

One of the most surprising themes that emerged across the focus groups was the

way that the COVID-19 pandemic changed the relationship between academics,

librarians, and digitized primary sources. On one hand, researchers were increasingly

dependent on digital facsimiles for *any* primary source access, but, on the other hand,

archives were more willing to scan documents on demand for affiliates. One humanities

librarian explained that how their process changed after the pandemic:

> Because of that shift with special collections, I feel like we in libraries got more comfortable with pointing people towards digitized material in lieu of bringing people into archives…pointing students towards digital material like Adam Matthew products when teaching rather than saying go to special collections.

The pandemic encouraged librarians to become more confident in assessing and

recommending digital resources like EEBO. Meanwhile, a librarian who routinely

instructed in tandem with staff at special collections in primary source research

instruction observed that "the same students that are in EEBO, working with primary

texts, are also going to be visiting special collections." Student proficiency in researching

with both physical and digital archives was actively encouraged by course

objectives in the affiliated academic department.

Paradoxically, the closure of physical library spaces at universities may have

strengthened the relationship between some professors and institutional archives. Forced

to confront the limitations of private databases like EEBO and Hathi Trust, professors

accustomed to physical primary sources found themselves relying on and recommending

open access digital materials. One professor described the relationship between digital

and physical archives as complementary:

> Once students start looking at EEBO and [Folger's] LUNA, they want to go to the libraries…The [city] is paper-packed with opportunities…Students go to the libraries and say, 'I'll be going to rare book libraries for the rest of my life.' EEBO can be a gateway drug to understanding how much they gain by going into those spaces.

In this use case, EEBO can be used in combination with bibliographic tools like the

ESTC and USTC, library catalogs, and finding aids to ease undergraduates into the

process of accessing and using materials in rare book libraries. It is much faster to

determine if a given book is of interest for physical examination by looking over a digital

copy than to request the book, schedule an appointment, and realize the book is not what

was desired. Digital archives can unlock physical archives for audiences intimidated by

textual finding aids intended for librarians and archivists.

Ambivalence Towards Adoption of New EEBO Features

Though the EEBO interface was recently overhauled from its former "Chadwyck-

Healey" branding to a ProQuest interface with a unified basic search function (Froehlich

2021, 115) none of the professors remarked on the recent interface changes as significant

to their research work in focus group discussions. If changes were observed – there were

references to the new "big search bar" – they were treated as aesthetic considerations at best. Because the experienced researchers already had specific documents in mind when using EEBO, changes to the search interface were trivial as long as the same documents were retrieved for a given query. This expertise with EEBO and its underlying structure was reflected in survey results: professors and librarians alike reported they could "find a specific document from EEBO given author, title, publisher, and publication date metadata" and that they could teach others to do the same with a high degree of confidence.

Both a librarian and a professor observed that faceted search had been recently implemented on EEBO, which allows users to restrict their query to certain date ranges. This has obvious benefits for scholars looking for documents with keywords from within narrow historical time frames like the English Civil War. Surprisingly, EEBO did not have a quick way for users to narrow their search queries from the results page until the 2020 update, according to an official tutorial[9] explaining the difference between the ProQuest and Chadwyck-Healey versions.

As regards reading material in the EEBO book interface, a professor observed that page navigation within books had been improved after the overhaul and that the (EEBO-TCP) transcriptions were now aligned with page images where applicable, which "made it easier to negotiate between the text and the images." In older EEBO interfaces, transcriptions and EEBO images were stored and displayed separately.

---

[9] See this presentation from the ProQuest official LibGuides for documentation of the interface changes. https://proquest.libguides.com/ld.php?content_id=56714724

However, the recent improvements to EEBO's documentation, educational resources, and cataloging materials are undeniable, and these were remarked upon librarians. EEBO now allows librarians to download MARC records of the items in EEBO, allowing for seamless integration of EEBO materials into library catalogs. As an instruction librarian explained, "I like to show people how to search the catalog that will take them to EEBO. The interchangeability of searching EEBO or the catalog records makes a difference for my teaching." For universities with hundreds of database subscriptions, unified catalogs can greatly simplify the research process for users with a specific text in mind.

The prioritization of product fundamentals over analytical features was corroborated by survey findings, where the image quality of digital facsimiles was, on average, considered the most significant factor to the ultimate quality of database products. Moreover, those interested in quantitative research tend to have a personal set of preferred programming languages, tools, and visualization platforms anyway. A professor who works with quantitative methods urged simplicity in interface design, as complicated interfaces "become dated more quickly" than streamlined ones.

Imagining and Marketing Alternative Databases

The following table lists key attributes of databases and digital archives of early printed books introduced during focus group discussions or preemptively explored in the supplementary survey. This list is by no means complete; I learned about several other products while compiling my literature review, but I have here limited myself to those mentioned in study transcripts.

As a database dedicated to the digitization of books rather than manuscripts, EEBO is lacking in unpublished documents like letters, bureaucratic records, or privately circulated materials. For example, scholars interested in early women's writing – which tended to stay in manuscript form – are likely better served by specialty archives when searching for primary sources.

Name	Active	Size	Documents	Contents	Period	Access	Ownership
Early English Books Online	Yes	146,000+ digitized books, millions of images	Printed books	Images, some transcriptions	1475-1700	Paid	Clarivate
EEBO-TCP	No	60,000 digitized books	Printed books	TEI XML transcriptions	1475-1700	Open	Clarivate, several universities
LUNA: Folger Digital Image Collection	Yes	135,000+ images, 300+ fully digitized books	Books, maps, art, manuscripts, clothing	Images, select full-text transcriptions	1550-1700+	Open	Folger Shakespeare Library
Perdita Manuscripts	Yes	230 documents, fully transcribed & searchable	Women's manuscript writing in Britain	Images, transcriptions	1500-1700	Paid	Adam Matthew
Penn in Hand, Colenda	Yes, merged in Colenda	6000+ manuscript images	Manuscripts in Penn's Rare Book Library	Images	800-1900+	Open	University of Pennsylvania
English Broadside Ballad Archive	Yes	9000+ ballads	Broadside ballads	Images, transcriptions, XML, recordings	1600-1700	Open	University of California, Santa Barbara

Figure 1 A compiled table of digital archives and database products mentioned in focus group discussions

Folger's LUNA has a narrow scope – limited to the collections of the Folger Shakespeare Library – but it was praised by professors especially for the high resolution of its images by the focus group participants who conduct or support research of the Early Modern period. One professor mentioned that, when scrutinizing minute aspects of a page, it was sometimes easier to make out fine details on a LUNA image than the original book. In fact, two professors and one librarian mentioned LUNA as an archive they have recommended their students use.

Adam Matthew, a vendor that specializes in digital manuscript archives, came up briefly in both focus group sessions. One professor mentioned them as a vendor with interesting new databases, while another mentioned the Adam Matthew product Perdita as a good resource for early women's manuscript writing. Compared to the massive scale of EEBO, Perdita is limited to a couple hundred manuscripts. However, these smaller collections can allow for consistency and individual attention.

Conversely, Penn in Hand, now absorbed into the Colenda Digital Repository, is a free manuscript archive of digitized public domain items from the Kislak Center at the University of Pennsylvania. One professor of English reported, "I'm teaching [a recipe] from Penn in Hand today…as 15 minutes of something fun." Penn in Hand represents both the promise and challenges of small, localized archives: the eclectic set of items – ranging from Vedic manuscripts to Early Modern recipes – makes it tricky to recommend to scholarly communities without prior familiarity with the holdings at the Kislak Center.

The English Broadside Ballad Archive[10] or EBBA surfaced in focus group discussions as a prototypical "genre-based" digital archive. A professor explained that "though there are some ballads in EEBO, the assortment is haphazard and split among the [Thomason] tract collections." EBBA, on the other hand, is dedicated to the "mounting of all surviving early ballads printed in English, with priority given to the black-letter broadsides of the seventeenth century" (Fumerton et al., 2012) and benefits from being designed for the display, taxonomy, and search of broadside ballads exclusively. EBBA is not a new archive, but rather an older digital archive that has been redesigned, iterated upon, and continuously grown over the past two decades.

I closed both focus group discussions with questions on community knowledge of new archives: how do you as a professor or librarian find small open-source archives that, by nature, aren't directly marketed to libraries by a vendor? As a contributor to digital humanities projects, I knew from experience that ongoing digital projects without both outside engagement and internal financial support tend to dwindle in usage then fall into disrepair. The answer wasn't straightforward. Professors tended to learn about new archives from other professors, whether in-person, through social media, or when mentioned in blog posts. One professor said they "look for libraries that have interesting [digital] holdings…I have a dozen that I use regularly in teaching. I like gathering them over time as I teach."

---

[10] *English Broadside Ballad Archive.* https://ebba.english.ucsb.edu/

Regarding the development and marketing of small archives *within* a particular university community, one professor encouraged researchers developing new archives to rely on library or institutional infrastructure whenever possible:

> Go the library, find out what [others] do for hosting content, because, odds are, you're going to get something that already does 95% of what you want it to do. It will require less maintenance because it's already supported at your institution…Advertise a little bit and make sure it's indexed.

Under this shared repository or hosted archive model, academic libraries are better equipped to recommend, maintain, and troubleshoot digital archives cultivated on campus.

With regards to marketing new digital archives, a librarian suggested integration of instructional materials like "lesson plans, examples of the material in action, or a published work by a scholar who has used this archive" into the archive website. The Women Writers Project[11], Folger Shakespeare Library[12], and EEBO[13] (though these were added recently) provide materials like LibGuides, resource lists, and lesson plans for use by librarians and faculty. Premade lesson plans can facilitate instruction with new databases by librarians or professors, while example papers give researchers trust in the scholarly value of the archive material.

Fledgling digital archives could benefit from tapping into networks of academic users to spread awareness. One librarian noted the persistent requests from vendors advertising paid products to universities crowd out other communications about new

---

[11] Women Writer's Project, https://www.wwp.northeastern.edu/
[12] Digital Resources at the Folger, https://folgerpedia.folger.edu/List_of_digital_resources_at_the_Folger
[13] LibGuides: Early English Books Online on the ProQuest Platform https://proquest.libguides.com/eebopqp

databases and archives. Like the professor participants, librarians trusted their

peers and colleagues over targeted advertisement. As an instruction librarian explained,

"the benefit of library conferences and academic conferences is that scholars come to me

and say this is a resource. Then I'll add it to our lists, especially if it's freely available."

Impact, Limitations, and Conclusions

This study is intended to provoke further, more structured research within literature departments, history departments, and academic libraries on the patterns of document retrieval, metadata supplementation, and use of alternatives when teaching and researching with EEBO/EEBO-TCP. For example, department liaisons at other academic libraries could organize diary studies that track usage patterns with EEBO and EEBO-TCP by librarians or faculty on a weekly basis over the course of a semester. Or, librarians could send out a survey of EEBO usage targeting the portions of the student body most likely to use EEBO: graduate students and advanced undergraduates in the humanities.

It is unlikely that scholars at institutions with expensive EEBO subscriptions could adequately represent the concerns of independent scholars or scholars at institutions that cannot afford expensive database subscriptions. This study leaves gaps for similarly nascent research on the needs, behaviors, and search mechanisms practiced by scholars of the period who do not have the means to access EEBO.

Through this study, I have attempted to elaborate the labor, whether in digitization, retrieval, or corpus organization, of research and teaching with EEBO and EEBO-TCP by professors and librarians. EEBO and EEBO-TCP have undeniably enabled and transformed research for medievalist and renaissance scholars, students, and librarians, but it would be naïve to assert that EEBO represents the best and only model for digital libraries of early books. I believe the academic library is an ideal site of critical

remediation between scholars and digitized rare books, perhaps even the place

to accumulate and organize higher quality, more efficient, better annotated, digital

versions of the books that have created and contextualized the English-language literary

landscape.

References

Basu, A. (2018). Form and Computation: A Case Study. In D. Currell & I. Issa (Eds.), *Digital Milton* (pp. 111–128). Springer International Publishing. https://doi.org/10.1007/978-3-319-90478-8_5

Bilansky, A. (2017). Search, reading, and the rise of database. *Digital Scholarship in the Humanities*, *32*(3), 511–527. https://doi.org/10.1093/llc/fqw023

Blaney, J., & Siefring, J. (2017). A Culture of non-citation: Assessing the digital impact of British History Online and the Early English Books Online Text Creation Partnership. Digital Humanities Quarterly, 11(1). https://www.proquest.com/docview/2555194938/abstract/9814B0594817413CPQ/1

Brightenburg, C. (2016). The digitization of early English books: A database comparison of Internet Archive and Early English Books Online. *Journal of Electronic Resources Librarianship*, *28*(1), 1–8.

Burke, V. E. (2021). Unconventional Experiments: Teaching Early Modern Women Writers. *Criticism*, *63*(1–2), 141–149. https://www.proquest.com/scholarly-journals/unconventional-experiments-teaching-early-modern/docview/2547639299/se-2

Crowther, S., Jordan, E., Wernimont, J., & Nunn, H. (2008). New Scholarship, New Pedagogies: Views from the "EEBO Generation." *Early Modern Literary Studies*, *14*(2). http://purl.oclc.org/emls/14-2/crjowenu.html

Cummings, J. (2012). The materiality of markup and the Text Encoding Initiative. In *Digitizing medieval and early modern material culture* (pp. 49–81). Iter.

Early English Books Online—Early English Books Online—ProQuest. Retrieved September 28, 2022, from https://www.proquest.com/eebo/

EarlyPrint. EarlyPrint. Retrieved December 1, 2022, from https://earlyprint.org/

Froehlich, H. (2021). Early English Books Online. *The Papers of the Bibliographical Society of America*, *115*(1), 114–117. https://doi.org/10.1086/712791

Froehlich, H. (2019). Using EEBO-TCP in Research & Teaching. In L. Rodrigues & E. Pappas (Eds.), *#DLFTeach Toolkit: Lesson Plans for Digital Library Instruction*. Digital Library Federation Digital Library Pedagogy Working Group. https://doi.org/10.21428/65a6243c.c4ffecb3

Fumerton, P., Stahmer, C., McAbee, K., & Browne, M. P. (2012). Vexed Impressions. In *Digitizing Medieval and Early Modern Culture* (pp. 257–285). Iter.

Gadd, I. (2009). The Use and Misuse of Early English Books Online. *Literature Compass*, *6*(3), 680–692. https://doi.org/10.1111/j.1741-4113.2009.00632.x

Gavin, M. (2017). How to Think about EEBO. *Textual Cultures*, *11*(1/2), 70–105.

Green, H. E. (2014). Facilitating Communities of Practice in Digital Humanities: Librarian Collaborations for Research and Training in Text Encoding. *The Library Quarterly: Information, Community, Policy*, *84*(2), 219–234. https://doi.org/10.1086/675332

Krueger, R., & Casey, Mary Anne. (2015). *Focus Groups* (5th edition). SAGE Publications. https://us.sagepub.com/en-us/nam/focus-groups/book243860

Lesser, Z. (2019). Xeroxing the Renaissance: The Material Text of Early Modern Studies. *Shakespeare Quarterly*, *70*(1), 3–31. https://doi.org/10.1093/sq/quz001

Lindquist, T., & Wicht, H. (2007). Pleas'd By a Newe Inuention?: Assessing the

   Impact of Early English Books Online on Teaching and Research at the University of

   Colorado at Boulder. *The Journal of Academic Librarianship*, *33*(3), 347–360.

   https://doi.org/10.1016/j.acalib.2007.01.011

Mah, C., Flanders, J., & Lavagnino, J. (1997). Some Problems of TEI Markup and Early

   Printed Books. *Computers and the Humanities*, *31*(1), 31–46.

   https://doi.org/10.1023/A:1000464519769

Mak, B. (2014). Archaeology of a digitization. *Journal of the Association for Information

   Science and Technology*, *65*(8), 1515–1526. https://doi.org/10.1002/asi.23061

Martin, S. (2014). Providing a Base for E-Editing: The Text Creation Partnership Project. *New

   Ways of Looking at Old Texts*. 47–59.

Miller, D. L. (2007). Building a Spenser archive one scan at a time. *Duke University Libraries*,

   20(2/3), 14-19. https://scholarcommons.sc.edu/cgi/viewcontent.cgi?article=1263

Mueller, M., & Burns, P. (2016). Collaborative Curation and Exploration of the EEBO-TCP

   Corpus. In *Early Modern Studies After the Digital Turn* (pp. 145–165). Iter Press.

Overholt, J. (2015). Together, we can FrEEBO. *Medium*.

   https://medium.com/@john_overholt/together-we-can-freebo-b33d39618f8

Sandler, M. (2004). Academic and Commercial Roles in Building "The Digital Library."

   Collection Management, *28*(1–2), 107–119. https://doi.org/10.1300/J105v28n01_09

Siefring, J., & Meyer, E. T. (2013). *Sustaining the EEBO-TCP Corpus in Transition: Report

   on the TIDSR Benchmarking Study* (SSRN Scholarly Paper No. 2236202).

   https://doi.org/10.2139/ssrn.2236202

Siefring, J., & Willcox, P. (2012). More than was Dreamt of in Our Philosophy. In

Digitizing Medieval and Early Modern Material Culture (pp. 83–111). Iter Press.

The Text Creation Partnership. (2018). About EEBO-TCP. *Text Creation Partnership*.

Retrieved December 4, 2022, from https://textcreationpartnership.org/tcp-texts/eebo-tcp-early-english-books-online/

Welzenbach, R. (2012). *Transcribed by hand, owned by libraries, made for everyone: EEBO-TCP in 2012*. http://deepblue.lib.umich.edu/handle/2027.42/94307

Williams, R. (2022). *Preparing 17th-Century Texts for the EarlyPrint Digitization Project –

LIBRARIES | Blog*. Retrieved December 5, 2022, from

https://sites.northwestern.edu/northwesternlibrary/2022/05/12/preparing-17th-century-texts-for-the-earlyprint-digitization-project/

Appendix A. Supplementary Survey

EEBO and EEBO-TCP: Entrance Survey

1. Do you consider yourself primarily a ...
    a. Librarian
    b. Professor
2. Which of the following digital collections are you **familiar** with?
    - Early English Books Online: Text Creation Partnership (EEBO-TCP)
    - EarlyPrint (earlyprint.org)
    - The Internet Archive
    - HathiTrust
    - Eighteenth Century Collections Online (ECCO)
3. Is there a digital collection for the 15th to 18th centuries important to your work not listed above? If so, enter it here?
4. When using EEBO or a comparable alternative, documents from which **authors** do you access most frequently?
5. When using EEBO or a comparable alternative, documents from which **historical periods** do you access most frequently?

6. Which of the following digital collections have you used for **research** purposes?

- Early English Books Online: Text Creation Partnership (EEBO-TCP)
- EarlyPrint (earlyprint.org)
- The Internet Archive
- HathiTrust
- Eighteenth Century Collections Online (ECCO)

7. Which of the following digital collections have you used for **teaching** purposes?

- Early English Books Online: Text Creation Partnership (EEBO-TCP)
- EarlyPrint (earlyprint.org)
- The Internet Archive
- HathiTrust
- Eighteenth Century Collections Online (ECCO)

8. For how long have you known about each of the following digital corpora? (in years, rounding up) Early English Books Online (EEBO)

- Early English Books Online: Text Creation Partnership (EEBO-TCP)
- EarlyPrint (earlyprint.org)
- The Internet Archive
- HathiTrust
- Eighteenth Century Collections Online (ECCO)

9. For how long have used each of the following digital corpora? (in years, rounding up)

- Early English Books Online: Text Creation Partnership (EEBO-TCP)
- EarlyPrint (earlyprint.org)
- The Internet Archive
- HathiTrust
- Eighteenth Century Collections Online (ECCO)

10. Please evaluate the following statements about your use of EEBO (on a Likert scale from "Strongly Disagree" to "Strongly Agree")

- I feel confident that I can find a specific document from EEBO given author, title, publisher, and publication date metadata.
- EEBO enables me to access identify trends across multiple documents.
- The EEBO interface slows down my workflow.
- I could teach others to use EEBO.
- For documents that exist both in EEBO and in other collections, I prefer to use alternative sources.
- Web page latency is an issue for me when using EEBO.

11. What are some of the challenges (or perceived challenges) of teaching others to use EEBO? Are they specific to undergraduates? Professors?

12. For retrieving or using documents available both on EEBO and alternative sources, which alternative sources do you prefer? Why?

13. Do you have any additional comments about your EEBO usage patterns?

14. Please evaluate the following statements about your use of EEBO (on a Likert scale from "Strongly Disagree" to "Strongly Agree")

- I feel confident that I can find a specific document from EEBO-TCP given author, title, publisher, and publication date metadata.
- EEBO-TCP enables me to access identify trends across multiple documents.
- I prefer to use a text editor or terminal to search EEBO-TCP XML files to the web interface.
- I could teach others to use EEBO-TCP.

- For documents that exist both in EEBO-TCP and in other collections, I prefer to use alternative sources.
- Web page latency is an issue for me when using EEBO-TCP.

15. Do you have any additional comments about your EEBO-TCP usage patterns?

16. Evaluate the importance of the following qualities of a hypothetical digital archive of Early Modern texts to your day to work. (on a Likert scale from "Strongly Disagree" to "Strongly Agree")

- Image quality
- Open access, available off-campus and without an institutional subscription The EEBO interface slows down my workflow.
- Support for free text search with normalized/modernized spellings
- For documents that exist both in EEBO and in other collections, I prefer to use alternative sources.
- Usability
- Metadata quantity and quality
- Curated corpora by theme, topic, or period
- Latency

17. Any last comments? Thank you for taking the time to complete this survey.