ESSAYS

Identifying Early Modern Books: Challenges for Citation Practices in Book History and Early Modern Studies

By Meaghan Brown, Paige Morgan, Jessica Otis

November 2017

Introduction

Humanities students are currently taught to examine citations as a starting point for their own research. This so-called "citation mining," "bibliography mining," or "citation chaining" allows students to discover and analyze scholarly discourse. Such environmental surveys of the literature familiarize researchers not only with the current trends in methodology but also with the types of material used for evidence, the major research libraries and archives for a specific field, and the significant scholars working in those areas. Citation studies—a specific branch of the field of bibliometrics—brings rigor to this process and enables scholars to systematically analyze citation patterns. However, to study citations at a scale sufficient to produce new insights into humanities research practices requires digital methodologies and curated datasets of humanities citations. The humanities need and deserve the ability to conduct analytically rigorous bibliometric studies on their corpus of scholarly literature.

Creating bibliometric datasets for the humanities allows researchers to undertake sophisticated analyses: Statistically, what are the most commonly cited works of scholarship? Are there distinctive patterns in topic, training, or access to specific archives among scholars who cite each other, or who cite a specific canonical source text? Examining humanities citations at scale would allow us to gain a clearer understanding of the intellectual lineages and influences of scholars, as well as the impact of historical collectors and collections on research. Citation analysis allows us to trace the connections among humanists and to visualize communication patterns among multiple fields and disciplines.

This avenue of exploration into humanities datasets targets the cherished myth of the solitary genius, allowing scholars to gain greater insight into the intellectual debts among humanists and the scholarly structures that support their research. Developing these datasets offers greater opportunities for exploring the role of the library in scholarship, as both holding repository for sources under study and important node of scholarly communication.

Scholars working in the same physical space form both formal and informal communities in which they may share ideas, frustrations, and resources. Conversely, electronic databases and new modes of scholarly communication like blogs and Twitter have the potential to disrupt patterns of geographic association, providing access to scholarly events and resources at a distance. Repositories such as libraries and archives are currently sporadically represented in formal citations—typically when unique primary materials are cited—while the contributions of databases and social media are radically underreported.[1] These issues inspired the creation of the current study, which tested methods for identifying humanist citations accurately at scale, and provides recommendations for increasing capture of significant features of primary source citations in particular.

Unlike many bibliometric studies that focus exclusively on contemporary scholarship, we are explicitly interested in the challenges posed to humanities bibliometrics by the variety of primary sources that humanists study. The purpose of a primary source citation is, as with other citations, to allow another scholar to locate the source of your information and reproduce your findings. Primary sources are accessed by scholars in three forms: physical materials held in special collections, archives, libraries, and museums; facsimiles of these materials, accessed through online databases, microfilm, or print; and reprints accessed online or in print. The first two forms create challenges for citation. Only the final form is currently adequately addressed in available style guides, as modern reprints are accessed and subsequently cited in the same manner as modern monographs. In creating the three datasets described below and analyzing the citation practices they contain, we wanted to know how primary sources and modern scholarship are currently being cited, how well current methods of citation discovery recognized citations in humanities fields, and what could be done to improve citation discovery so that these processes may be studied more in the future.

History of Citation Practices in Early Modern Studies and Book History

The study of modern citation practices aims to show the form of scholarly communication as it develops. Bibliometric research and the methodology developed for it have typically focused on recent article publications in the sciences. This analysis has been more difficult to perform on humanities scholarship, due to the humanities' heavy reliance on single-author monographs and "older sources," by which most bibliometric studies mean secondary works more than fifteen years old. The monograph, while critical to many humanities fields, is currently a black box when it comes to studying citations, as the contents of in-copyright monographs are not commonly available as data for researchers. Humanities scholars must contend with a much broader range of sources than in the sciences, both in date and format. As T. N. Van Leeuwen notes, scientific knowledge ages within three to four years, while humanities books may influence a field for decades.[2] While studies of recent humanist scholarship are rare, humanist bibliometric datasets and analyses that take into account primary sources are even rarer.[3] Bibliometric studies in the humanities suffer from what Chris Sula and Matthew Miller have called a "sheer lack of data as compared with the sciences."[4]

Humanities citation data poses several challenges to the type of collection and analysis often performed on datasets from scientific disciplines. In the hard sciences, publication in journals, particularly in high-impact factor journals, is highly valued for academic advancement.[5] The scholarly reference aggregator Crossref has surpassed forty million metadata records for content as of 2010. Of these, 87% are from journals, while monographs and reference works amount to 5%, and the remainder are conference papers or other publication formats.[6] In the humanities, by contrast, the scholarly monograph is still the sine qua non of tenure-track life. By and large, humanities monographs within copyright are not available to researchers as digital texts in large numbers. Since bibliometric datasets rely heavily on the automatic processing of digital texts, the content of monographs, including their citations, is not yet as readily available for analysis and is explicitly excluded from key sources for bibliometric datasets.[7]

Since the services that create these datasets are not set up for humanities data, scholars interested in studying it must make their own dataset. Humanities data is available through journal aggregators such as JSTOR. One hopes that as such aggregators begin to include monographs, and as other digitization projects such as HathiTrust begin to make in-copyright monograph information available, monograph data will be invaluable to future studies. In the meantime, as humanists employ the same citation methods in both articles and monographs, this work can begin by developing methods for acquiring and analyzing humanities data from journals, which could then later be extrapolated to monographs.

Bibliometric studies focus on the sharing of ideas, and as such illuminate patterns and methods of communication among scholars. The sciences and humanities have marked differences not only in how and where collaboration and resource-sharing occur but also in the degree of formality attached to such communication and shared work, making scientific models for considering co-authorship and credit less applicable. Scientific research often employs large teams of researchers—into the thousands—each with degrees of credit determined by their position within author lists, while co-authorship in the humanities occurs in a much lower percentage of papers and in much smaller groups.[8] Scientific papers often explicitly denote collaborative activity through formal co-authorship, which enables easily automated methods of extracting relational data and reconstructing intellectual influence by analyzing the metadata of the article.[9] By contrast, humanities research assistance is often part of informal information exchanges and is thus included in less standardized formats within published works: in acknowledgement sections or as notes of appreciation in discursive footnotes.[10] In other words, while some bibliometric work in the sciences is made possible by the structure of repositories themselves, humanities bibliometric work is more reliant on the capture of data from the content of scholarly publications. As we discovered in this study, both the assistance and resources provided by libraries, archives, and repositories are even more likely to be coded obliquely—in notes of appreciation to specific and general librarians, in appendices, and in image captions. Because of its variability, humanities collaboration data requires different methods of discovery and aggregation, but should likewise be helpful in exploring networks of intellectual debt and tracing the development of ideas and fields of study.

Finally, humanities sources are far more variable than those used in the sciences, and are thus more challenging to automatically aggregate and analyze. Driven by the need to measure the impact of research funding, scientific bibliometric analysis often emphasizes the currency of research and focuses on recently collected and analyzed data. In contrast, the humanities articles we examined cited primary sources dating from the classical period to the twenty-first century that were as materially diverse as wax tablets, engraved prints, manuscript account books, and datasets.

Many humanities scholars are familiar with the challenge of identifying a "work," and those examining humanities bibliometric data will need to wrestle with this question as well. Is a canonical text in its third printing the same "work" as the first? What about scholarly editions or translations? These questions present challenges to the assessment of secondary sources as much as primary materials, since humanities monographs have far longer periods of influence than scientific works. Likewise, reference works that describe and catalog primary materials can cause challenges because the ways they identify and refer to primary sources are updated over time. For example, the Pollard and Redgrave *A Short–Title Catalogue of Books Printed in England, Scotland, & Ireland and of English Books Printed Abroad, 1475–1640* (STC) was introduced in 1926 and substantially revised in the 1970s and 1990s.[11] The STC was further remediated into the English Short Title Catalogue (ESTC) beginning in the 1980s, and citations that rely on the STC number to identify works must be considered in conjunction with which edition of the STC or ESTC was available

at the time of writing.12 A humanities-centric paradigm for bibliometric studies is needed—one that recognizes multiplicity (translations, new editions, remediations), acknowledges the influence of time on sources (both primary and scholarly), and captures the diversity of material types used to illuminate humanities scholarship.

Methodology

Acquiring Humanities Bibliometric Data

A bibliometric analysis that can yield useful insights hinges on the selection of an appropriate dataset. The dataset must be attuned to the specific questions the analysis is designed to illuminate, but it is also subject, inevitably, to practical considerations. Our datasets allowed us to survey recent practice in humanities citation rather than to provide a comprehensive analysis of specific fields. For this project, we focused on journal data acquired from JSTOR's Data for Research program, which has an established system for allowing researchers access to in-copyright journal articles.13 JSTOR contains electronic holdings for a substantial number of humanities journals: 333 history journals, 329 language and literature journals, and 15 bibliography journals, according to JSTOR's subject headings.14

To create a manageable sample for this study, we selected articles from three journals focusing on book history and three on early modern studies. These fields were selected for their interdisciplinarity, our familiarity with the subject matter and common reference works, and their likelihood of citing rare materials. We chose high-profile journals that provide a range of approaches within each field.15 While the resulting data may be particularly interesting to early modernists and book historians, the methodology should be replicable across a number of fields, and the basic insights relevant to a range of humanities disciplines. We requested two separate datasets from JSTOR. Our initial request was for JSTOR's automatically identified citations and keywords for six journals—*Book History*, *English Literary History* (*ELH*), *Studies in Bibliography*, *Huntington Library Quarterly*, the *Sixteenth Century Journal*, and *Renaissance Quarterly*—from 1985 to the present. Our subsequent request was for as many full-text articles from the same six journals as we could get. All of these journals follow the *Chicago Manual of Style*. We supplemented these two datasets with a third, much smaller dataset that we created ourselves.

Dataset 1: JSTOR-identified citations

JSTOR has created algorithms to locate citation data in Optical Character Recognition-scanned full-text articles, which allows them to create cross-links and promote other articles in their database via a reference tab on each article's page. We discuss both how JSTOR does this and the implications for the data, below. In our initial request, JSTOR's Data for Research (hereafter JSTOR) sent us automatically extracted citation data for five of our six requested journals. We were not able to acquire JSTOR's automatically extracted citation data from *Renaissance Quarterly*, due to a data retrieval error that sent us citation data from *Bibliothèque d'Humanisme et Renaissance* instead. We considered including *Bibliothèque d'Humanisme et Renaissance* as an alternative to *Renaissance Quarterly*, but the French journal has several practices at variance with the *Chicago Manual of Style* citation practice, and we decided that its inclusion would add too many new variables to the study. In particular, the use of first initials rather than full first names would have complicated efforts to identify authors using the Virtual International Authority File (VIAF; viaf.org). We were able to secure Renaissance Quarterly data for a subsequent dataset.

We requested 30,000 citations, JSTOR's largest request size, for each journal. The initial data pull of "text units" included full-length articles, as determined by JSTOR, as well as book reviews, society business, and other materials published in these journals. The second set of numbers, below in table 1, shows only the full-length articles.

Journal	Number of text units	Number of full-length articles

Book History	213	188
ELH	1,326	1,110
Huntington Library Quarterly	1,181	767
Sixteenth Century Journal	8,731	675
Studies in Bibliography	343	301

Individual factors in journal publication history affected the quantity of available articles: *Book History*, an annual founded in 1998, returned only 188 articles for 1998–2013, while *Studies in Bibliography*, published sometimes every two years rather than annually, contained 301 articles from 1985 to 2005/2006.16 *Sixteenth Century Studies*, a quarterly publication, contains far more book reviews than the other publications; when excluded, these brought its number of full-length articles down considerably. Each set began with works published in 1985 (with the exception of *Book History*), and the end date was determined by several factors—primarily, availability caused by the journal's own publication, JSTOR's holdings of it, and JSTOR's moving wall.17 While *Book History* has a shorter publication span than the other journals, it is among the most prestigious journals in its field available through JSTOR, and we deemed that its shorter publication span would not affect our analysis, given our interest in recent publications.

This dataset included bibliographic information for individual articles published in each of the five journals; i.e., that article's author, title, journal title, publication date, volume/issue, and page numbers, as well as citation data for each article. This citation data was encoded at the level of the individual reference. We converted the bibliographic information for each article (not including individual citation data) from XML to RDF triples, and reconciled author names against data in VIAF in order to be able to query the data using Unique Resource Identifiers (URIs) as opposed to strings of text. This data is both available in the project GitHub repository and to query at a SPARQL endpoint (http://dydra.com/paigecm/large-dataset-for-all-5-journals). We have created a range of precoded queries that can be run by users with no experience with SPARQL. Each query includes explanatory comments (indicated by the # sign at the beginning of the line) and many queries include a regular expression filter that allows further customization by users, such as searching for a particular author's name or a keyword.

Dataset 2: full-text articles

Our second dataset consisted of 7,500 full-text articles from all six journals in JSON format. In this subsequent data request, we aimed for as many articles as JSTOR would provide, in order to allow us to test a variety of citation-capturing methods. JSTOR's rolling blackout window and the lack of recent publications held in JSTOR for some of our target journals, such as *Studies in Bibliography*, began affecting most of our journals around 2009, resulting in a sharp decrease in available articles from 2009 to 2015, when this data was acquired. The full-text articles range in date from 1931 to 2015, with a median date of 1984 (see figure 1). The earliest publications in the dataset are from *Huntington Library Bulletin*.18 Twenty-four articles included in the dataset were from the spring of 2015. 58% of the full-text articles (4,369) were early modern studies, while 42% (3,131) were bibliographical.

Journal	Number of full-text articles
Book History	188
ELH	2,433
Huntington Library Quarterly	1,897
Renaissance Quarterly	909

Sixteenth Century Journal	1,027
Studies in Bibliography	1,046

We then ran text-string searches on the full-text JSON files using Python, to see if we could increase the rate of discovery for citation formats currently missed by JSTOR's automatic citation discovery software.
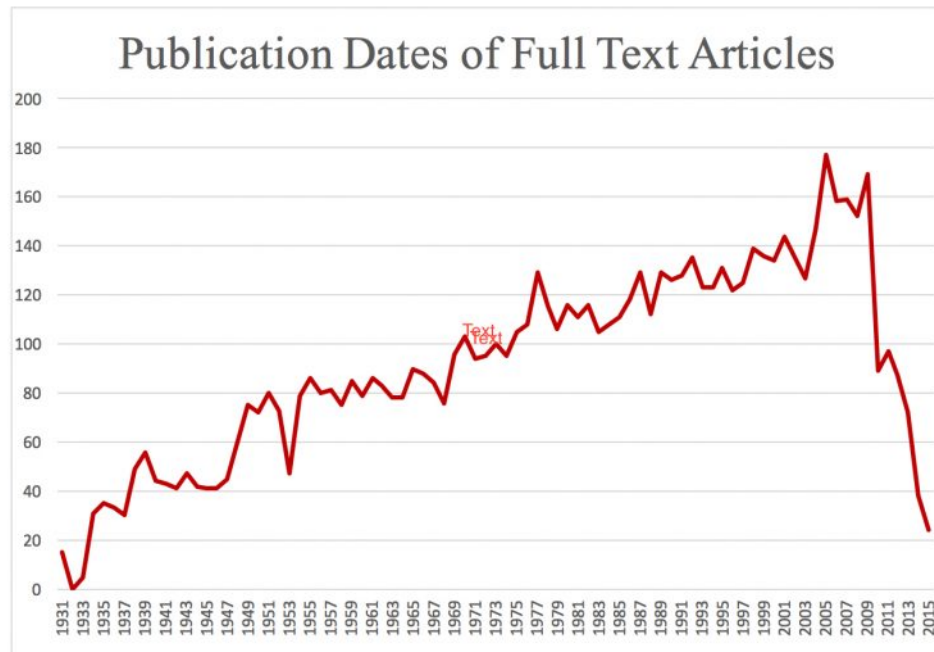

Figure 1. Dates of full-text articles.

Dataset 3: hand-encoded citations

Finally, we sampled fifty articles published between 1998 and 2013 in *Book History*. We hand curated the citations in each, encoding bibliographic elements such as author, publisher, date, place of publication, and elements connected to archival research such as shelfmarks and call numbers, collection names, and repositories. Our process resulted in data structured with finer granular detail suitable for more complex queries. Our structured data isolated fields, making it possible to develop database queries that allow us to ask questions such as, "Which authors cited archival materials housed in a particular repository?" or "How many of these articles include citations for materials with shelfmarks?" The enhanced data for *Book History* is available in the project GitHub repository and also available to query at a SPARQL endpoint (http://dydra.com/paigecm/book-history-enhanced-data.). This SPARQL endpoint also features prewritten and customizable queries. Our detailed methodology will be released as a white paper, titled "Identifying Early Modern Books: Adapting, Creating, and Processing Bibliometric Data for Research."19

We chose *Book History* for the hand-coded dataset for several reasons: as the journal for the Society of the History of Authorship, Reading, and Publishing, it's a leader in the interdisciplinary "book history" field, publishing more varied works than *Studies in Bibliography* and other more senior publications do. While *Book History* publishes a wide range of scholarship, it also has a significant portion of publications that fall within the early modern era, allowing our sample to have articles focused in both of our primary fields. Using the automatically culled citations from JSTOR as a starting point, project team members read the individual articles, transcribed citations missed by the automatic process, deleted items that were not citations, and lightly encoded the resulting citations list in XML tags. The XML tags were then processed to create a database for more sophisticated queries.

This process enabled us to analyze a combination of highly precise citation data from a small number of articles (dataset 3, hand-coded citations); messy and partially pre-processed citation data from a large number of articles (dataset 1, JSTOR-identified citations); and extremely messy citation data drawn from a further sample of full-text articles (dataset 2, full-text). Complete information about our process of extracting and analyzing this data, including programs and methods used, will be available in our forthcoming white paper.

Citations

Universal Style Guides

Programs for automatically recording citations recognize citations through pattern matching, page position, and font size. This can be seen in results of the automated citation-extraction algorithms employed by JSTOR. These algorithms run pattern-matching searches over sections of text that JSTOR identifies as possibly containing citation data. The search patterns are derived from the standardized citation patterns set forth in official, published style guidelines such as the *Chicago Manual of Style* and the *MLA Handbook*. The searches, therefore, rely on citations containing specific elements and only those elements—e.g., a certain number of characters followed by a comma, followed by quotation marks or italics, parentheses, and a final number. Semicolons are assumed to divide two citations. Modern secondary sources, such as articles, typically do not provide challenges to this pattern-based approach: e.g., Edwin Berck Dike, "Coleridge Marginalia in Henry Brooke's *The Fool of Quality*," *Huntington Library Bulletin* no. 2 (1931): 149–63. Disrupting the established pattern—for example, by separating elements with non-citation descriptive comments or having an element exceed typical length—can make the automatic discovery process miss elements. Works published prior to the nineteenth century, and citations of specific copies of modern works, often do not play by the rules.

We chose all six of our journals in part because they require their authors to follow the *Chicago Manual of Style* (*CMOS*) or a lightly modified house version, which minimized the variation that would be inherent in a comparison of journals using different style guides. In addition to the actual citation practices manifested in published articles, we also extensively reviewed the journals' published style guidelines in order to correctly attribute the patterns we uncovered, either to authorial practices or journal conventions. We are well aware that not all humanities journals employ the *CMOS—Digital Humanities Quarterly* uses Harvard citation style, an author-date format, in its online publications. Some variations we saw in the citations may be due to authors introducing elements from other styles, particularly when the *CMOS* is silent on the citation format for particular categories of materials.[20] For example, the 16th edition of the *CMOS*, the current edition at the time of writing this analysis, does not give guidance on how to cite material objects in its documentation section, although some guidance on "works of art" is given in the chapter on illustrations.[21] This arrangement presumes that all cited material objects are "works of art," and that all references to artworks are accompanied by a reproduction.

Few journals follow the *CMOS* verbatim. Most create their own variants of the style guide. At the time of our study, the most significant departures in our dataset were *Sixteenth Century Journal*'s refusal to allow the use of *ibid.* or the italicization of titles for books and journals, and *Renaissance Quarterly*'s omission of publisher names and last-access dates for online sources. Furthermore, these variant style guides are not always kept up to date when their canonical style guide changes. *Sixteenth Century Journal* "generally follows the *Chicago Manual of Style*, 15th edition, for grammar, punctuation, capitalization, hyphenation, and the like," but acknowledges the existence of the 16th edition and the *SBL Handbook of Style* as reference works for the citation of sources.[22]

Our hand-coded *Book History* dataset gave us a baseline for the accuracy of JSTOR's automatic citation-discovery algorithms. Even with some fuzzy matching to account for publishers' house styles, the algorithms only capture 89.1% of human-readable citations in our hand-coded article dataset, with individual article capture rates ranging from a dismal 24.7% to a complete 100%.23 The more significant an author's departure from the *CMOS*, the higher the likelihood that their citations would be missed.

Local Variation in Practice

Academic citations in the humanities generally require a kind of linear dependency; a human reading an article should encounter a full citation at the first mention of a source, then short-form references to the source if it is mentioned again. Some of these short-form references are standardized by style

guidelines; however, authors sometimes make additional abbreviations of long titles or the names of libraries to reduce redundancy, streamline the repeated citation of a single item, or manage references to a number of related primary sources (see figures 2 and 3).24 These abbreviation practices can appear to reduce redundancy in highly familiar texts (and for most authors, nothing is more familiar than the article one has been writing for months), but for human and machine readers unfamiliar with the subject matter, such abbreviations can disrupt the ability to extract relevant data.

MALYNES AND MISSELDEN   15

CCC   Gerard de Malynes, *The Center of the Circle of Commerce* (London:
      W. Jones, 1623); *STC* 17221.
ET   Thomas Mun, *England's Treasure by Forraign Trade. Or, The Bal-
      ance of our Forraign Trade is the Rule of our Treasure* (London:
      J. G. for Thomas Clark, 1664 [*STC-Wing M3073*]), reprint (Fair-
      field, N.J.: Augustus M. Kelley, 1986).
EV   Gerard de Malynes, *Englands View, in the Unmasking of Two Para-
      doxes: With a Repplication unto the answer of Maister John Bodine*
      (London: R. Field, 1603); *STC* 17225.
FT   Edward Misselden, *Free Trade. Or, the Meanes to Make Trade
      Flourish*, 2d ed. (London: J. Legatt for S. Waterson, 1622); *STC*
      17987.
LM   Gerard de Malynes, *Consuetudo, vel Lex Mercatoria, or the Ancient
      Law-Merchant* (London: A. Islip, 1622); *STC* 17222.
MFT   Gerard de Malynes, *The Maintenance of Free Trade* (London:
      J. L.[egatt] for W. Sheffard, 1622); *STC* 17226.
SG   Gerard de Malynes, *Saint George for England Allegorically De-
      scribed* (London: R. Field for W. Tymme, 1601); *STC* 17226a.

Figure 2. Abbreviations employed for primary sources throughout the article, explained in a concluding table.

Notes

1. John Sandsbury, *Ilium in Italiam*, 1608, Bodleian Library (hereafter BodL) MS Eng Poet fol. 10.

Figure 3. Endnotes showing the use of shelfmarks.

Authors sometimes rely on context to supply elements of a formal citation, such as eliminating an author's name when her work is discussed in the sentence, or a publisher's name in a list of his published works. The absence of a standard citation element sometimes ironically indicates the importance of that element to the subject under discussion.

Conversely, adding additional information through discursive footnotes can be highly legible to humans, but it is not machine-readable because it breaks the pattern recognition required for the automatic extraction of citations: e.g., "Judith Williamson discusses these gaps in so-called hermeneutic ads in chap. 3 of *Decoding Advertisements* (Boston: Boyars, 1978)."25 In this example, the author's name, Judith Williamson, was excluded from the captured citation along with the other discursive footnote matter. Many of the citations in our final dataset lack author information, not because it was absent from the article but because it was distanced from the formal citation itself by analysis, exegesis, or commentary.

Location, Location, Location

JSTOR and individual researchers such as Sula and Miller use a combination of font size and location on the PDF page to identify sections of text on which these pattern-matching algorithms should be run. This methodology can be particularly problematic for articles that cite a large number of primary sources. In bibliographical journals such as *Studies in Bibliography*, where copy-specific information about material objects is often of vital importance to the argument, appendices that compare the features of many copies may appear as tables or lists. Because

TABLE 1.2. *Pericles* (1619)

	tp	R	S	T	V	X	Y	Z	2A	2B
Copy										
C2	2	2	23	2	2	23	6.1	2	1\14	2
L	2	2	23	2	1\14	7.1	6.1	1\14	23	2
O	23	2	2	2	2	7.1	6.1	2	1\14	23
Y2	2	2	2	2	2	7.1	6.1	2	1\14	2
F(1)	2	2	23	2	1\14	23	6.1	2	1\14	2
F(2)	23	2	2	2	2	7.1	23	2	6.1	23
F(3)	2	2	23	2	2	7.1	6.1	2	1\14	2
F(8)	2	2	23	2	2	7.1	6.1	2	1\14	2
F(9)	23	23	2	2	2	7.1	6.1	2	1\14	23
F(10)	23	2	2	2	2	7.1	6.1	2	1\14	2
F(11)	--	2	23	2	2	7.1	23	2	1\14	23
F(12)	--	2	23	2	2	7.1	23	2	1\14	?
HD	23	2	23	2	2	7.1	6.1	2	1\14	23
Y	2	2	2	2	2	7.1	6.1	2	6.1	--
NY(1)	2	2	2	2	2	7.1	6.1	2	2	2
NY(2)	23	23	2	2	2	7.1	6.1	2	1\14	23
N	2	2	2	2	2	7.1	6.1	2	1\14	2
HN	2	2	23	2	2	7.1	6.1	2	1\14	2
L18	2	23	23	2	2	7.1	6.1	2	2	2
E	2	23	2	2	2	7.1	23	2	1\14	2
E2	6.1	2	23	2	2	7.1	6.1	2	1\14	--

Figure 4. Table showing the sequence of watermarks in copies of William Shakespeare's *Pericles*, employing STC location codes to indicate repository.

they are printed in the same font size as the body of the article, these citations are invisible to automated citation discovery and cause additional complications for human researchers interested in library references (see figure 4).26 Such tables and lists need to be able to provide location information in a concise and clear method, often resulting in libraries being referenced by code systems (e.g., "DFo" for the Folger Shakespeare Library in Washington, DC), rather than by their names.27 Researchers interested in identifying copies examined by bibliographers not only have to contend with finding and recording the numerous citations to specific works but also with translating the codes for libraries into recognizable repository names.

Image captions are similarly missed by the current discovery process because the citations they contain occur within the body of the text and, depending on journal, may be displayed in a font size larger than that of the footnotes. In some cases, a work may be cited only in an image caption possibly because the author sees a second full citation in footnotes as redundant. While legible to humans, this form of citation would be missed completely by anyone scanning footnotes for a full list of consulted works or searching using automatic means. Because of *CMOS* recommendations for credit lines in illustrations, image captions are also more likely to contain references to libraries and other repositories than footnotes, even when citing the same material. The addition of a process for capturing image captions would substantially improve the representation of primary sources in automatically generated reference lists.

A partial and labor-intensive solution to this problem can be implemented through text-string searches, for phrases such as "by permission of" or "courtesy of," or through location and font-size matching in the cases of journals that set image captions at a smaller font size than article text.

Primary Source Challenges

Most authors adhere to style guidelines when citing modern secondary works in their foot- and endnotes. The biggest challenge for humanities bibliometrics comes from the citation of primary sources. While style guides usually provide advice for citing manuscript sources, they tend to treat printed primary sources as if they were modern printed works. In practice, however, we found that many scholars emend or add to the citation format for printed books when citing rare materials in order to indicate to their reader that they consulted a specific item. Indeed, with the prevalence of digital-facsimile resources, some scholars see citing the holding library and shelfmark as a means of clarifying that they inspected the physical book. The more a citation varies from the norm for a printed book, however, the more likely it will be missed in whole or in part by the automatic citation discovery process as it currently stands.

Citations that point to copy-specific information in a printed work, such as provenance or marginalia, are more likely to include library information. While borrowed from manuscript practice, these added features appear in the reverse order of the official *CMOS* style, where manuscript citations list the smallest to the largest unit—item, author, papers, collection, repository, and occasionally conclude with the location of the repository: e.g., "Minutes of the Committee for Improving the Condition of the Free Blacks, Pennsylvania Society for the Abolition of Slavery, 1790–1803, Papers of the Pennsylvania Society for the Abolition of Slavery, Historical Society of Pennsylvania, Philadelphia."28 In the case of printed books, the standard print format comes first, followed by copy-specific information; e.g., William Cecil, *The Copie of a Letter Sent Ovt of England to Don Bernardin Mendoza* (London: Jacqueline Vautrollier for Richard Field, 1588), A2^r, Bodleian Library, Douce M 202.29 Because the pattern-matching software has already recognized a full citation at the page reference (A2^r), it would miss the library name and shelfmark, in this case the Bodleian Library, and "Douce M 202," a shelfmark from the collection of Francis Douce.

Item-specific information is interesting to other researchers who would like to examine those specific items, of course, but the ability to track specific items in citations would offer additional benefits. Access to information about citation frequency and content could help librarians develop collections that respond to researcher interests, identify gaps in discovery tools, or orient programs around frequently consulted items. Such bibliometric studies may be particularly helpful in illuminating the influence of fellowship programs. At the very least, consistent citations of specific holdings could help researchers find the libraries that are most relevant to their interests and thus increase traffic to those libraries. Scholars interested in discovering citations of works held in specific libraries will need to do more experimentation to determine what text-string searches work best. While in our study a search of the full-text dataset for the term "library" turned up too much noise (there were too many articles *about* libraries in the book history journals), we were able to identify cited libraries by searching for "shelfmark" and the related term "classmark."30 We also had some success searching for the names of individual librarians and institutions.

The additions and changes scholars make to a standard *CMOS* printed-works citation appear to reflect the different levels of citation granularity demanded by modern scholarly research methods, although the ad hoc nature of the changes make their application inconsistent. In the vast majority of cases, the citation of copy-specific information necessitated the inclusion of a call number. In our dataset of fifty *Book History* citations, we found only one example of an annotated book cited with a library name but without a shelfmark.31

A Snapshot of the Text: Photographic Facsimiles as Stand-Ins for Printed Books

In recent years, libraries and archives have worked extensively to increase engagement with their holdings through digitization, while also making major purchases in digital facsimile databases to provide access to material digitized by for-profit companies such as ProQuest and Gale-Cengage. But even recognizing, as G. Thomas Tanselle does in his seminal 1989 article "Reproduction and Scholarship," that reproduction "can never be thought to obviate examination of originals," the rates for the citation of major microfilm series and digital facsimile databases are far lower than we expected.32 4,369 or 58% of the full-text articles we examined were early modern in focus, yet the primary facsimile repository database for early modern studies, Early English Books Online (EEBO), and its parent microfilm series, Early English Books (EEB), *together* appeared in only 46 articles. Numbers for other common online early modern databases are even smaller in our dataset: the Eighteenth Century Collections Online (ECCO), appears in twelve articles, while British History Online (BHO), a digital collection of primary sources relating to Great Britain, is cited by eleven.33

EEBO: An Invisible Database

Given the significance of EEBO to early modern studies, we were particularly interested in the "culture of non-citation" that appeared to erase databases from the academic literature.34 We considered that the time span for publication may be affecting these numbers, as EEBO first came online in 1998. However 1,909 of the articles were published after the year 2000. Date was also a poor predictor of whether the microfilm or the online database would be cited. The count for the Early English Books microfilm series includes a 2009 *ELH: English Literary History* article on broadsides that cites the microfilm twenty-five times.35In some cases, researchers may not have institutional access to online databases, but their libraries may have, or may be able to borrow, the microfilm reels. Conversely, information for online or print sources may be included in the citations for works actually accessed via microfilm, because scholars and editors believe microfilm is becoming more difficult to access. In these instances, microfilm reel numbers often appear in conjunction with the citation of another format, such as a print citation with repository information or an online database: e.g., the citation of the State Papers microfilm reel alongside a State Papers Online URL.

The scarcity of explicit database citations in early modern and bibliographical studies does not mean that scholars do not *use* databases in these fields. In the calendar year 2015, scholars and staff at the Folger Shakespeare Library looked at 24,515 page images in EEBO in 6,089 distinct sessions, and they downloaded 1,181 PDFs.36 This number does not include scholars who may have logged into EEBO through a proxy service to their home libraries. So where is EEBO in their citations?

A possible answer may be found in the nearly 9,000 Short Title and Wing catalogue numbers (typically labeled "STC" and "Wing"), which supplement the publication information of early modern books. Such edition-level identification numbers can be used to disambiguate between two similarly titled editions, but we suspect that these references are most commonly included to serve as unique identifiers to quickly and easily find a rare work in EEBO and other databases. The Early English Books microfilm series specifically drew its coverage mandate from the STC, and some tutorials for using the microfilm series direct users to use the STC number to locate specific works in the microfilm index.37 Although limited by its national focus, the STC represents a broad enough range of early modern printed books that many other databases allow for searching by STC number.

Another possible reason for the absence of references to EEBO is deliberate omission. In an examination of history masters theses at a midsized public university, Graham Sherriff notes that many of his sources contained what he characterizes as "incorrect citations," with problems including "incorrect or missing dates, duplicates, and the citation of a text's original publication rather than that of the specific edition consulted."38 Sherriff was particularly concerned

with scholars using a medieval or early modern publication date when they had, in fact, consulted a modern scholarly or facsimile edition. The significant absence of citations for EEBO and other facsimile databases suggests that, rather than being an error, this citation pattern represents a fundamental disjunction between scholarly practice and its textual representation. Further evidence in support of this theory comes from informal conversations and blog posts, as well as a 2013 user survey for the EEBO Text Creation Partnership, in which 34% of respondents professed to have cited the print version of texts they accessed online, with an additional 25% admitted to teaching their students to follow the same method.39 This practice is so widespread that some scholars grow frustrated when databases do not make available all the information necessary to convert a digital citation into a print one.40

Responding to the availability of facsimiles through microfilm readers, long before digital databases, Tanselle lamented that "everyone knows (though many people act as if they do not know) that every form of reproduction can lie, by providing a range of possibilities for interpretation that is different from the one offered by the original."41 The citation pattern suggests that the "act" Tanselle identified—the practice of using facsimiles and physical books interchangeably—has compounded with what Jonathan Blaney, project editor for British History Online, characterizes as the "strong presumption in parts of the humanities to prefer print wherever possible" to suppress citation references to facsimiles.42 The print citation, with higher social cachet and a formula recognized by all major citation manuals, is used instead. This practice complicates the citation process, if the purpose of citations is to direct readers to your actual sources, leading to a situation where, as Blaney says, "the reader . . . might not know that there is a version literally at their fingertips."43

Best Practices for Human Comprehension and Machine Computation

We have compiled a few suggestions for best practices to increase human legibility for citations as well as best practices to promote greater citation capture using automatic citation discovery. In both cases, the most effective means of communicating citations is consistency. We strongly recommend that authors follow the given style guidelines—whether *Chicago Manual of Style* or some other set of guidelines—to ensure the comprehensibility and regularity of their citations. It is particularly important to follow style guidelines related to abbreviations; *CMOS* recommends authors keep abbreviations to a minimum, introduce them at first use, and—when they are numerous—document them thoroughly in a single place in the article or book.44

For authors:

- To ensure human comprehension and accurately represent the scholarly work being performed, authors should cite the work they have consulted.
- When citing primary sources consulted at an archive, include the library name and shelfmark.
  - Using *CMOS* to show consultation of a printed work in an archive:
    - Christopher Marlowe, *The Famous Tragedy of the Rich Ievv of Malta* (London: John Beale for Nicholas Vavasour, 1633). Folger Shakespeare Library, STC 17412 copy 2.
    - William Cecil [Richard Leigh, pseud], *La Copie d'vne Lettre Enuoyée d'Angleterre à dom Bernardin de Mendoze* ([London: Jacqueline Vautrollier for Richard Field], 1588). Bodleian Library, Vet. A1 e.118.
- When citing works consulted through digital facsimile or transcription, database information, especially durable URLs whenever possible, should be included. At a minimum, a short form of the database name may signal the method of consultation.
  - Using *CMOS* to show consultation of a digital facsimile:
    - John Fletcher, *The Faithfvll Shepheardess* (London, [Edward Allde] for R[ichard] Bonian and H[enry] Walley, [1610?]). EEBO.
  - Using *CMOS* to show consultation of a digital transcript:

- John Fletcher, *The Faithfull Shepheardess* (London, [Edward Allde] for R[ichard] Bonian and H[enry] Walley, [1610?]. EEBO-TCP A00962.
- All works referenced should be fully cited in foot- or endnotes, even if the work is used solely as an illustration and has a caption.
- Authors should avoid, whenever possible, editorializing within the citation itself. A discursive footnote may precede the citation, but the order of the citation elements should not be interrupted by commentary.

For editors and the creators of style guides:

Clearly, in some cases, the minimum citation recommended for books is inadequate. For the *Chicago Manual of Style* and other standard style guides, advice for citing manuscripts may serve as a guide for citing rare printed materials, but we would strongly suggest that style guide editors consider creating standard citations for printed materials in special collections and archives. The archival turn in the humanities has meant that copy-specific studies—such as those involving marginalia, material history, and provenance studies—are becoming increasingly important to our understanding of human culture. Standardizing the additions made to the typical print citation and creating a more clear and full citation form for rare printed materials would aid significantly in creating reliable and consistent references to this material.

Furthermore, it is impossible to predict what future objects may become a locus of scholarship and require citation outside the given style guidelines. We suggest that style guides include a general citation format for material objects—for realia as well as artwork. How should an author cite Edgar Allan Poe's desk, a gravestone, or a rubber ducky, all items held in various repositories?45 How do we cite such items when they are not held in repositories but are instead found in historical homes, graveyards, or on store shelves? We also recommend that editors and the creators of style guides provide guidance to authors who need to create new types of citations when current style guidelines fail to cover some "edge case" of scholarship—similar to the Modern Language Association's newest attempt to create one generic method of citation— but in *addition* to current style guidelines, not in *lieu* of specific style guidance.

For data publishers:

We believe that people who create, publish, and/or maintain this kind of data should not only be aware of potential uses for it but also of the tools and practices that scholars use to work with it. Awareness of the detailed practices associated with bibliometric research will make it more likely that publishers will provide data in useful configurations. Doing so may increase the discoverability and use of their products.

For data miners:

See our whitepaper on our process, "Identifying Early Modern Books: Adapting, Creating, and Processing Bibliometric Data for Research."46

Conclusion

The goal of citations in general, and of our recommendations, is to support transparency and aid in the critical evaluation of scholarly arguments. Building and analyzing reliable datasets of humanities citations is difficult because of the variety of forms humanities works may take; however, it is not impossible, and the benefits of analyzing such datasets are enormous. Humanities citation analysis enables us to understand current research practices at a global level, beyond the scope of individual scholars' carefully cultivated fields of expertise. This global approach has the potential to identify vital sources that form important connections between disciplinary fields. At a more localized level, recognizing the contributions that special collections archives and facsimile databases make to our research not

only helps future scholars find the material necessary to confirm or contest our conclusions but also allows libraries to evaluate the effectiveness of pricey subscriptions, funds for visiting scholars, and collecting practices and priorities. To get a better picture of where the humanities are going, we need to have a better dataset of where we've been in the digital repository, the archive, and the library.47

0.

1. Judith Siefring and Eric T. Meyer, *Sustaining the EEBO-TCP Corpus in Transition: Report on the TIDSR Benchmarking Study* (London: JISC, 2013), 17, http://ssrn.com/abstract=2236202 (http://ssrn.com/abstract=2236202). [↵]

2. Thed N. Van Leeuwen, "Bibliometric Research Evaluations, Web of Science and the Social Sciences and Humanities: A Problematic Relationship?" *Bibliometrie—Praxis & Forschung* 2 (2013): 1–2, http://www.bibliometrie-pf.de/article/view/173 (http://www.bibliometrie-pf.de/article/view/173). [↵]

3. Jordi Ardanuy, "Sixty Years of Citation Analysis Studies in the Humanities (1951-2010)," *Journal of the Association for Information Science and Technology* 64, no. 8 (2013): 1751–55, doi:10.1002/asi.22835 (http://dx.doi.org/10.1002/asi.22835). For an example of proposed humanities bibliometrics techniques limited to secondary source analysis, see A. J. M. Linmans, "Why with Bibliometrics the Humanities Does Not Need to Be the Weakest Link: Indicators for Research Evaluation Based on Citations, Library Holdings, and Productivity Measures," *Scientometrics* 83, no. 2 (2010): 337–54, doi:10.1007/s11192-009-0088-9 (http://dx.doi.org/10.1007/s11192-009-0088-9). [↵]

4. Chris Sula and Matthew Miller, "Citations, Contexts, and Humanistic Discourse: Toward Automatic Extraction and Classification," *Literary and Linguistic Computing* 29, no. 3 (2014): 452, doi:10.1093/llc/fqu019 (http://dx.doi.org/10.1093/llc/fqu019). [↵]

5. Impact factor is the number of citations of a journal divided by the number of articles published in the journal. Impact factor only measures citations in other journal articles and, while used for tenure and promotion in some regions, is problematic for humanists. [↵]

6. Anonymous, "40 Million Crossref DOIs Preserve the Record of Scholarship," *News Release* (newsfeed), February 5, 2010, https://www.crossref.org/news/2010-02-05-40-million-crossref-dois-preserve-the-record-of-scholarship/ (https://www.crossref.org/news/2010-02-05-40-million-crossref-dois-preserve-the-record-of-scholarship/). [↵]

7. Linmans, "Why with Bibliometrics," 338. [↵]

8. Lisa Spiro, "Collaborative Authorship in the Humanities," *Digital Scholarship in the Humanities* (blog), April 21, 2009, https://digitalscholarship.wordpress.com/2009/04/21/collaborative-authorship-in-the-humanities/ (https://digitalscholarship.wordpress.com/2009/04/21/collaborative-authorship-in-the-humanities/). [↵]

9. Amy Brand, Liz Allen, Micah Altman, Marjorie Hlava, and Jo Scott, "Beyond Authorship: Attribution, Contribution, Collaboration, and Credit," *Learned Publishing* 28, no. 2 (2015): 151–55, doi:10.1087/20150211 (http://dx.doi.org/10.1087/20150211). [↵]

10. One co-author's favorite such acknowledgment in a discursive footnote reads, "Herbert McCabe made this observation to me late in an evening of wine, whiskey and argument." Diarmaid MacCulloch, *The Reformation: A History* (New York, NY: Penguin Books, 2003), 780. [↵]

11. A. W. Pollard, G. R. Redgrave, W. A. Jackson, F. S. Ferguson, and Katharine F. Pantzer, eds., *A Short–Title Catalogue of Books Printed in England, Scotland, & Ireland and of English Books Printed Abroad 1475–1640* (STC), 2nd ed., 3 vols. (London: Bibliographical Society, 1976–91). [↵]

12. English Short Title Catalogue, British Library website, accessed October 9, 2016, http://estc.bl.uk/ (http://estc.bl.uk/). For more on the history of the English Short Title Catalogue, see "English Short Title Catalogue – History," British Library website, accessed October 9, 2016, http://www.bl.uk/reshelp/findhelprestype/catblhold/estchistory/estchistory.html (http://www.bl.uk/reshelp/findhelprestype/catblhold/estchistory/estchistory.html). [↵]

13. "Data for Research," JSTOR, accessed October 9, 2016, http://about.jstor.org/service/data-for-research (http://about.jstor.org/service/data-for-research). [↵]

14. "Browse By Subject," JSTOR, accessed October 9, 2016, http://www.jstor.org/subjects (http://www.jstor.org/subjects). See especially the subpages http://www.jstor.org/subject/history (http://www.jstor.org/subject/history); http://www.jstor.org/subject/literature (http://www.jstor.org/subject/literature); and http://www.jstor.org/subject/bibliog (http://www.jstor.org/subject/bibliog). These are not mutually exclusive lists. [↵]

15. At the time of data collection, *Book History, Studies in Bibliography,* and *Huntington Library Quarterly* were the three most prestigious, academically focused bibliography journals indexed by JSTOR. JSTOR's Bibliography category includes a number of professionally focused library journals, and journals with more regional and format-focuses, such as *American Periodicals*. At the time of data collection, *Papers of the Bibliographical Society of America* was not yet available. [↵]

16. "History," The Society for the History of Authorship, Reading and Publishing website, accessed October 9, 2016, http://www.sharpweb.org/main/purpose-history/ (http://www.sharpweb.org/main/purpose-history/). *Studies in Bibliography*, the Bibliographical Society of the University of Virginia website, accessed October 9, 2016, http://bsuva.org/wordpress/studies-in-bibliography/ (http://bsuva.org/wordpress/studies-in-bibliography/). [↵]

17. JSTOR's moving wall is a rolling blackout period that determines when an article becomes available: "Each journal has a "moving wall," defined as a time lag between the most current issue published and the content available on JSTOR. Most archival journals have moving walls of between three and five years, but publishers may elect walls anywhere from zero to 10 years. Content is added to collections annually, which means that the cost per page for the JSTOR archive collections decreases every year" ("What's in JSTOR: Journals," JSTOR, https://about.jstor.org/whats-in-jstor/journals/ (https://about.jstor.org/whats-in-jstor/journals/)). [↵]

18. For example, Edwin Berck Dike, "Coleridge Marginalia in Henry Brooke's *The Fool of Quality*," *Huntington Library Bulletin* no. 2 (1931): 149–63. Both *Huntington Library Quarterly* and *Renaissance Quarterly* changed names over the period of study, from *Huntington Library Bulletin* and *Renaissance News* respectively. [↵]

19. This white paper will document the process and specific steps taken to transform very simply structured XML data into data suitable for querying in a nonrelational database. Paige Morgan, Meaghan Brown, and Jessica Otis, "Identifying Early Modern Books: Adapting, Creating, and Processing Bibliometric Data for Research," (forthcoming). [↵]

20. Heather Froehlich, when blogging about citing online databases, described her citation practices as having "been cobbled together from normative citation practices with input from the collection creators." Heather Froehlich, "Suggested Ways of Citing Digitized Early Modern Texts," *Heather Froehlich* (blog), August 6, 2015, https://hfroehli.ch/2015/08/06/suggested-ways-of-citing/ (https://hfroehli.ch/2015/08/06/suggested-ways-of-citing/). [↵]

21. No part of the chapter on illustrations lays out a list of required citation elements, instead mentioning appropriate orders for elements if included: e.g., "When a caption provides the dimensions of an original work of art, these follow the work's medium and are listed in order of height, width, and (if applicable) depth. This information need appear only if relevant to the text, unless the rights holder requests that it be included." "Including Original Dimensions in Captions," 3.27, *The Chicago Manual of Style,* 16th ed. (Chicago: University of Chicago Press, 2010), http://www.chicagomanualofstyle.org/16/ch03/ch03_sec027.html (http://www.chicagomanualofstyle.org/16/ch03/ch03_sec027.html) [↵]

22. "Sixteenth Century Journal Author Instructions," *The Sixteenth Century Journal,* accessed October 9, 2016, https://www.escj.org/pdf/SCJguidelines%202013.pdf (https://www.escj.org/pdf/SCJguidelines%202013.pdf). [↵]

23. These numbers are based on a subset of thirty-six articles in the fifty-article hand-coded dataset, as some recent articles did not yet have JSTOR-generated citation data available and were hand-transcribed to provide a more varied date range. [↵]

24. Andrea Finkelstein, "Gerard de Malynes and Edward Misselden: The Learned Library of the Seventeenth-Century Merchant," *Book History* 3 (2000): 15, http://www.jstor.org/stable/30227309 (http://www.jstor.org/stable/30227309); Jeffrey Todd Knight, "'Furnished' for Action: Renaissance Books as

Furniture." *Book History* 12 (2009): 66, http://www.jstor.org/stable/40930539 (http://www.jstor.org/stable/40930539). [↵]

25. This example taken from Erin A. Smith, "How the Other Half Read: Advertising, Working-Class Readers, and Pulp Magazines," *Book History* 3 (2000): 229, http://www.jstor.org/stable/30227317 (http://www.jstor.org/stable/30227317). [↵]

26. R. Carter Hailey, "The Shakespearian Pavier Quartos Revisited," *Studies in Bibliography* 57 (2005/2006): 180, http://www.jstor.org/stable/40372110 (http://www.jstor.org/stable/40372110). [↵]

27. Two common repository code systems are employed in the journals we studied: the Machine-Readable Cataloguing (or MARC) code list for organizations, and the library codes used in the STC. The Library of Congress hosts a MARC code translator, and we have transcribed the STC location codes for easier access. "Search the MARC Code List for Organizations Database," MARC Code List for Organizations, accessed October 21, 2016, https://www.loc.gov/marc/organizations/org-search.php (https://www.loc.gov/marc/organizations/org-search.php); The full transcription of the STC Library codes with cross references to ESTC library codes encoded in TEI, as well as an HTML version, can be found at our GitHub, https://github.com/jmotis/IdEMB (https://github.com/jmotis/IdEMB). [↵]

28. See this and additional examples in CMOS (sixteenth edition) 14.240. [↵]

29. Shelfmarks, also sometimes called "pressmarks" or "casemarks," originally designated the physical location of a book within a library. Today the term is largely synonymous with call numbers, unique identifiers attached to an item irrespective of its physical location within a library, especially as items move into rare book libraries while retaining shelfmarks assigned by previous owners. "Classmarks," in contrast, assign identifiers based on the content of the work. The Library of Congress call number system and the Dewey Decimal system are both classmark systems. [↵]

30. These are terms that rarely appear outside of shelfmark citations, even in bibliographical studies. Although relatively few scholars use the term "shelfmark" when including them, searches for these specialized terms provided a low-recall but high-precision return for identifying the libraries that contributed to the scholarship: in other words, they didn't pull back everything, but nearly everything pulled back was what we wanted. [↵]

31. We speculate that the author made this choice because the book was part of a circulating collection at a university library, rather than part of a special collections library. The library did have only one copy of that work, and when we inquired, we were able to confirm that the circulating copy did have the marginalia on the pages mentioned, so citing the repository only was sufficient for us to find the quoted material. [↵]

32. G. Thomas Tanselle, "Reproductions and Scholarship," Studies in Bibliography 42 (1989): 37, http://xtf.lib.virginia.edu/xtf/view?docId=StudiesInBiblio/uvaBook/tei/sibv042.xml;chunk.id=vol042.02;toc.depth=1;toc.id=vol042.02;brand=default (http://xtf.lib.virginia.edu/xtf/view?docId=StudiesInBiblio/uvaBook/tei/sibv042.xml;chunk.id=vol042.02;toc.depth=1;toc.id=vol042.02;brand=default). [↵]

33. We searched for both full names and standard abbreviations in compiling these numbers. [↵]

34. EEBO has recently received additional attention for this issue, see Jonathan Blaney and Judith Siefring, "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, no. 1 (2017): http://www.digitalhumanities.org/dhq/vol/11/1/000282/000282.html (http://www.digitalhumanities.org/dhq/vol/11/1/000282/000282.html). [↵]

35. Eric Nebeker, "Broadside Ballads, Miscellanies, and the Lyric in Print," *ELH* 76, No. 4 (Winter, 2009): 989–1013, http://www.jstor.org/stable/27742970 (http://www.jstor.org/stable/27742970). [↵]

36. Erin Blake, email to Meaghan Brown, July 6, 2016. [↵]

37. Ironically, the advice for using the EEB microfilm and EEBO is, itself, widely distributed without attribution. We located this information at "Early English Books – Microforms," Truman State University, accessed October 9, 2016, http://library.truman.edu/microforms/early_english_books.asp (http://library.truman.edu/microforms/early_english_books.asp). The same information can be found at "Early

English Books on Microfilm," Queens College Library, accessed October 9, 2016, http://qcpages.qc.cuny.edu/~nagnew/eeb.htm (http://qcpages.qc.cuny.edu/~nagnew/eeb.htm). [↵]

38. Graham Sherriff, "Information Use in History Research: A Citation Analysis of Master's Level Theses," *Libraries and the Academy* 10, no. 2 (2010): 169, doi:10.1353/pla.0.0092 (http://dx.doi.org/10.1353/pla.0.0092). [↵]

39. Samuli Kaislaniemi, "How Should You Cite a Book Viewed in EEBO?" *Copious but Not Compendius* (blog), February 27, 2014, http://blogs.helsinki.fi/kaislani/2014/02/27/how-to-cite-eebo/ (http://blogs.helsinki.fi/kaislani/2014/02/27/how-to-cite-eebo/); Froehlich, "Suggested Ways of Citing"; Siefring and Meyer, *Sustaining the EEBO-TCP Corpus*, 17. [↵]

40. Blaney and Siefring, "A Culture of Non-Citation," par. 13–15. [↵]

41. Tanselle, "Reproductions and Scholarship," 33. [↵]

42. Jonathan Blaney, "The Problem of Citation in the Digital Humanities," in *Proceedings of the Digital Humanities Congress 2012*, eds. Clare Mills, Michael Pidd, and Esther Ward. Studies in the Digital Humanities. Sheffield: HRI Online Publications, 2014, https://www.hrionline.ac.uk/openbook/chapter/dhc2012-blaney (http://www.hrionline.ac.uk/openbook/chapter/dhc2012-blaney). [↵]

43. Blaney, "The Problem of Citation in the Digital Humanities." [↵]

44. *The Chicago Manual of Style*, 10.3, http://www.chicagomanualofstyle.org/16/ch10/ch10_sec003.html (http://www.chicagomanualofstyle.org/16/ch10/ch10_sec003.html). [↵]

45. The following citations are modeled on our advice for following manuscript patterns, but we fully admit the following desk citation would benefit from size information, in the manner of an art object. Anonymous, Edgar Allen Poe's writing desk, circa 1835–1837. Personal Effects Collection, Harry Ransom Center, University of Texas at Austin, http://norman.hrc.utexas.edu/poedc/details.cfm?id=7 (http://norman.hrc.utexas.edu/poedc/details.cfm?id=7); Gravestone of Lee Howard Dobbins, 1853. Oberlin College Library Special Collections; *Shakespeare: The Original Collectible Bath Duck* (San Raphael, CA: Celebriducks, c1999). Folger Shakespeare Library, ART 253594 (realia) (B1b). [↵]

46. Morgan, Brown, and Otis, "Identifying Early Modern Books: Adapting, Creating, and Processing Bibliometric Data for Research," (forthcoming). [↵]

47. This work was made possible by generous funding from the Andrew W. Mellon Foundation (https://mellon.org) and the Council on Library and Information Resources (https://www.clir.org). [↵]

Meaghan Brown

Digital Production Editor – Folger Shakespeare Library

Paige Morgan

Digital Humanities Librarian – University of Miami

Jessica Otis

Digital Humanities Specialist, University Libraries – Carnegie Mellon University

FOOTNOTES

LICENSE

SHARE

🐦(https://twitter.com/intent/tweet?text=Identifying Early Modern Books: Challenges for Citation Practices in Book History and Early Modern Studies on https://www.archivejournal.net/essays/identifying-early-modern-books/)
𝐟(http://www.facebook.com/sharer/sharer.php?u=https://www.archivejournal.net/essays/identifying-early-modern-books/)
𝓰+(https://plus.google.com/share?url=https://www.archivejournal.net/essays/identifying-early-modern-books/)

Archive Journal (https://www.archivejournal.net/)