

2018

## Exploring the American Archivist: Corpus analysis tools and the professional literature

J. Gordon Daines III

*Brigham Young University - Utah*, [gordon\\_daines@byu.edu](mailto:gordon_daines@byu.edu)

Cory L. Nimer

*Brigham Young University - Utah*, [cory\\_nimer@byu.edu](mailto:cory_nimer@byu.edu)

Jacob R. Lee

[jacobrexlee@gmail.com](mailto:jacobrexlee@gmail.com)

Follow this and additional works at: <https://elischolar.library.yale.edu/jcas>

 Part of the [Archival Science Commons](#)

---

### Recommended Citation

Daines, J. Gordon III; Nimer, Cory L.; and Lee, Jacob R. (2018) "Exploring the American Archivist: Corpus analysis tools and the professional literature," *Journal of Contemporary Archival Studies*: Vol. 5 , Article 3.

Available at: <https://elischolar.library.yale.edu/jcas/vol5/iss1/3>

This Article is brought to you for free and open access by EliScholar – A Digital Platform for Scholarly Publishing at Yale. It has been accepted for inclusion in *Journal of Contemporary Archival Studies* by an authorized editor of EliScholar – A Digital Platform for Scholarly Publishing at Yale. For more information, please contact [elischolar@yale.edu](mailto:elischolar@yale.edu).

## EXPLORING *THE AMERICAN ARCHIVIST*

### Introduction

*The American Archivist* is the premiere archival journal in the United States. It is also the oldest. The topics covered in the journal reveal what the profession considered important and provides a window into the development of the American archival community. The volume of content covered in the journal's eight decades of existence presents a challenge to researchers interested in mining the journal for trends and themes in their efforts to construct an adequate understanding of the American archival community. Different researchers have taken a variety of approaches to meeting this challenge.

Periodic analyses of the literature of professional organizations is an important way of understanding the development and growth of professional communities. This article introduces corpus analysis and its associated tools as a productive new approach to exploring the archival professional literature. Corpus analysis is a methodology for performing text analysis that allows researchers to make comparisons between textual objects on a large scale. While corpus analysis has been used extensively in the humanities, it has not been used to examine the professional literature of archives, libraries, museums, or other related professions. This article uses a limited study to highlight how one of these instruments, Voyant Tools, could be used to examine the literature found in *The American Archivist* in order to better understand what the American archival community values and what issues have been important to them.

### Literature review

Since its founding in 1938, *The American Archivist* has served as a leading voice for the archival profession in the United States. Featuring articles, book reviews, and organizational news, the journal's impact on the American archival profession has been widespread and significant. An analysis of the journal's contents offers insights into the history and culture of the profession in the United States. Archivists have used several different approaches to assess the content of *The American Archivist*. These approaches have ranged from simple inventories of the types of articles in the journal to bibliometric studies of the content. Linguistic analysis methods and software provide a new means for exploring the journal's content and what it teaches us about the history and culture of the archival profession in the United States.

Published analyses of content from *The American Archivist* conducted within the archival profession have often been undertaken by the journal's editors in order to call for improvements to the literature. In 1952, Karl L. Trever suggested that the value of the journal was in providing "balanced content" that addressed the needs of a diverse group of archivists, manuscript curators, and administrators at a time when the archival profession was still developing its own identity.<sup>1</sup> While Trever believed that the journal was making significant contributions to the profession, he also recognized that the journal faced two significant challenges: (1) receiving enough content, and (2) making sure that that content was of practical use. One of the weaknesses of the article is

---

<sup>1</sup> Karl L. Trever, "The American Archivist: The Voice of a Profession," *American Archivist* 15, no. 2 (April 1952): 148.

that it does not include either a qualitative or quantitative analysis of article topics or other characteristics. This prevents us from clearly understanding what is meant by articles of a practical nature and precludes comparison with the types of content received at a later time period (or with other analyses of the journal's content). A later editorial analysis, produced by Richard J. Cox in 1994, includes a statistical analysis of the content in North American archival journals (including *The American Archivist*) during the period between 1970 and 1992. According to Cox, during this period only eighty-eight research articles were published, with the majority focused on either management issues (61.4 percent) or historical topics (31.8 percent). He lamented the fact that very few research articles were published in the journal and argued convincingly that "we as a field desperately need research."<sup>2</sup> He particularly emphasized the need for research that explores the *why* of archival practice and that helps develop archival theory.

Independent researchers have also published on the breadth and limitations of the journal's content. In a 1992 study, Mary Sue Stephenson analyzed the characteristics of 390 articles published in *The American Archivist* over a twenty-year period, compiling data on the primary topic of each article and the demographics of their authors. Based on this analysis, Stephenson found that "the primary subject area receiving the highest level of coverage was the General Literature category, followed by Use of Archives and Historical Manuscripts, and Repositories."<sup>3</sup> Stephenson's article is purely descriptive and focuses primarily on who was writing for the journal and what types of content were being published during this period. While this article gives us important insights into the journal's content, it makes no arguments about how those insights help us better understand the archival profession. This article also uses a different methodology than other analyses of the journal's content, which precludes deeper comparisons. More recently, Wakefield Harper conducted a comparative study of archival journal content based on selected volumes from four decades, which included a topical analysis of individual articles. His results indicated a greater diversity of content, with a "marked increase in the attention paid to functional areas of archival science."<sup>4</sup> However, Harper's article uses yet a different methodological approach for his analysis.<sup>5</sup>

While these studies provide some points of comparison for our analysis, in many cases their scope either goes beyond *The American Archivist* or is focused on a limited period. In the case of bibliometric analyses, while such studies are common in library science literature, Paul Conway and William E. Landis have noted the potential for subjectivity in topic assignment and have

---

<sup>2</sup> Richard J. Cox, "An Analysis of Archival Research, 1970–92, and the Role and Function of the *American Archivist*," *American Archivist* 57, no. 2 (Spring 1994): 287. Cox also published an analysis of the wider American archival literature, which included a statistical summary of topics addressed (see Richard J. Cox, "American Archival Literature: Expanding Horizons and Continuing Needs, 1901–1987," *American Archivist* 50, no. 3 (Summer 1987): 306–23).

<sup>3</sup> Mary Sue Stephenson, "*The American Archivist*, 1971 to 1990: A Demographic Analysis of the Articles," *American Archivist* 55, no. 4 (Fall 1992): 538–61.

<sup>4</sup> Wakefield Harper, "A Content Analysis of Archival Journal Literature" (MA thesis, University of North Carolina at Chapel Hill, 2010), 29.

<sup>5</sup> Another recent article, Jessica Wagner Webster's "'Filling the Gaps': Oral Histories and Underdocumented Populations in *The American Archivist*, 1938–2011," *American Archivist* 79, no. 2 (Fall-Winter 2016): 254–82), uses a bibliometric approach to explore a single topic rather than the corpus as a whole.

questioned the reliability of such studies.<sup>6</sup> Their point is a valid one and highlights the need to explore new approaches to content analysis. These new approaches should allow us to explore the full text of the journal itself for professional and cultural trends.

Corpus analysis, and its related fields of corpus linguistics or “text mining,” provide an alternative to bibliometric methods that focuses on the use of the data provided by the words in a body of text. This “big data” approach uses statistics to analyze a representative body of text, looking at word frequency and collocation of terms as well as the topical content of the corpus.<sup>7</sup> As with other “big data” approaches, corpus analysis relies on the idea that patterns and ideas will emerge from the data as it is examined and analyzed. Software applications for conducting this research have become increasingly accessible for linguists and digital humanities scholars, including web-based products such as Voyant Tools.<sup>8</sup> These applications allow scholars, journalists, and other researchers to explore corpora, and to develop compelling visualizations of linguistic data.<sup>9</sup>

Tools for corpus analysis have been used to identify trends in a broad range of historical texts. Large corpora such as the Corpus of Early English Correspondence (CEEC), the Corpus of Historical American English (COHA), or the Google Books project have been used for studies in social and linguistic history,<sup>10</sup> while other researchers have developed their own subject-specific corpora to examine specific themes. Articles such as Dermot O’Reilly and Mike Reed’s study of professionalization in public service policy statements, or Brian S. Budgell, Alice Kwong, and Neil Millar’s diachronic study of the *Journal of the Canadian Chiropractic Association* suggest possibilities for examining a body of professional literature.<sup>11</sup> Both of these articles argue that a

---

<sup>6</sup> Paul Conway and William E. Landis, “Open-Access Publishing and the Transformation of the *American Archivist* Online,” *American Archivist* 74, no. 2 (Fall-Winter 2011): 492. Examples of library bibliometric studies include K. P. Singh and Harish Chander, “Publication Trends in Library and Information Science: A Bibliometric Analysis of *Library Management* Journal,” *Library Management* 35, no. 3 (2014): 134–49; Ming-yueh Tsay and Zhu-yee Shu, “Journal Bibliometric Analysis: A Case Study on the *Journal of Documentation*,” *Journal of Documentation* 67, no. 5 (2011): 806–22; and Akhtar Hussain, Nishat Fatima, and Devendra Kumar, “Bibliometric Analysis of the ‘Electronic Library’ Journal (2000–2010),” *Webology* 8, no. 1 (June 2011), accessed January 20, 2017, <http://www.webology.org/2011/v8n1/a87.html>.

<sup>7</sup> Viktor Mayer-Schönberger and Kenneth Cukier, *Big Data: A Revolution That Will Transform How We Live, Work, and Think* (Boston: Houghton Mifflin Harcourt, 2012); Michael McCarthy and Anne O’Keefe, “Historical Perspective: What Are Corpora and How Have They Evolved?” In *The Routledge Handbook of Corpus Linguistics*, ed. Anne O’Keefe and Michael McCarthy (London and New York: Routledge, 2010): 3–13. Additional resources on corpus analysis include R. Harald Baayen, *Word Frequency Distributions* (Dordrecht, Netherlands: Kluwer Academic Publishers, 2001), and John Sinclair, *Corpus, Concordance, Collocation* (Oxford: Oxford University Press, 1991).

<sup>8</sup> Voyant Tools, “Voyant: See Through Your Text,” accessed February 2, 2017, <http://voyant-tools.org/>.

<sup>9</sup> Stéfan Sinclair and Geoffrey Rockwell, “Text Analysis and Visualization: Making Meaning Count,” in *A New Companion to Digital Humanities*, ed. Susan Schreibman, Ray Siemens, and John Unsworth (Malden, MA: Wiley/Blackwell, 2016): 274–90; and Giorgio Uboldi and Giorgio Caviglia, “Information Visualizations and Interfaces in the Humanities,” in *New Challenges for Data Design*, ed. David Bihanic (London: Springer, 2015), 207–18.

<sup>10</sup> Kate Tranter, “Corpus Analysis of Historical Documents,” in *The Encyclopedia of Applied Linguistics*, ed. Carol A. Chapelle (Hoboken, NJ: John Wiley and Sons, 2014); Jean-Baptiste Michel et al., “Quantitative Analysis of Culture Using Millions of Digitized Books,” *Science* 331, no. 6014 (January 2011): 176–82.

<sup>11</sup> Dermot O’Reilly and Mike Reed, “The Grit in the Oyster: Professionalism, Managerialism, and Leaderism as Discourses of UK Public Services Modernization,” *Organization Studies* 32, no. 8 (2011): 1079–1101; and Brian S.

profession's literature has important lessons to teach us about the process of professionalization. The concept of professionalization is an important one for American archivists and there is much to learn from the profession's literature—particularly *The American Archivist*.

There are a number of different tools available for performing corpus analysis. Voyant Tools is a relatively new product and there are few examples of research using Voyant Tools in the academic literature. Among these articles are Lorie A. VanChena's "Reading German Culture, 1789–1918," which looks at the contextual use of the terms "love" and "death" in German literature, and Jeffrey Drouin's "Close- and Distant-Reading Modernism: Network Analysis, Text Mining, and Teaching the *Little Review*," which examined word frequency and use in a pre-World War I literary journal.<sup>12</sup> In each of these cases, the authors were able to use a statistical analysis to obtain new insight into a particular culture, to prompt new questions, and even to permit reassessment of certain scholarly claims. Voyant Tools offers similar possibilities for an analysis of *The American Archivist*.

## Methodology

It was our hope that applying the tools of textual analysis to *The American Archivist* would provide similar insights into the development of the archival profession in the United States. We had to complete several steps to prepare the corpus for analysis. This included obtaining and normalizing a copy of the journal text, importing the cleaned files into Voyant Tools, and selecting terms for analysis. The system then provides a series of textual analysis and visualization tools for exploring the content.

### *Preparing content for analysis*

Preparing the content of *The American Archivist* for textual analysis was a labor-intensive activity. We first obtained PDF copies of every issue of the journal from 1938 to 2015 from the Society of American Archivists' publications staff. We then examined the content of each issue to determine whether to include the entire available text, or if only portions (such as articles and case studies) should be included. After some discussion, we decided to include all the content from each issue, as our focus was not limited to the content of research articles but extended to overall trends in professional communication within the community.<sup>13</sup> Prior to analysis, we also

---

Budgell, Alice Kwong, and Neil Millar, "A Diachronic Study of the Language of Chiropractic," *Journal of the Canadian Chiropractic Association* 57, no. 1 (2013): 49–55.

<sup>12</sup> Lorie A. VanChena, "Reading German Culture, 1789–1918," presented at "Distant Readings/Descriptive Turns: Topologies of German Culture in the Long Nineteenth Century," 21st St. Louis Symposium on German Literature & Culture, Washington University in St. Louis, March 29–31, 2012; Jeffrey Drouin, "Close- and Distant-Reading Modernism: Network Analysis, Text Mining, and Teaching the *Little Review*," *The Journal of Modern Periodical Studies* 5, no. 1 (2014): 126–27.

<sup>13</sup> While we felt that the decision to use the full text of the issues provided a greater overall view of the journal's impact, this may prevent direct comparisons to previous bibliometric studies of *The American Archivist*'s content noted above.

needed to clean up and normalize the OCR (optical character recognition) text in the PDF files, including the removal of hyphenation and character errors.<sup>14</sup>

Having cleaned the data, the authors turned to data analysis. The intent of textual analysis is to identify patterns and trends in a textual corpus. The analysis of word frequency requires a large data sample because smaller data samples open the door for greater uncertainty in findings. The larger the sample size, the more information and context is provided, which reduces uncertainty. Linguist R. Harald Baayen has suggested that an appropriate size for a text corpus is close to twenty million words.<sup>15</sup> Fortunately, the word count for *The American Archivist* corpus (as defined above) is almost eighteen million words, which allows us to be confident in our results.

### *Voyant Tools and data analysis*

The cleaned text of *The American Archivist* was loaded into Voyant Tools, an open-source, web-based textual analysis program. Developed by Stéfan Sinclair and Geoffrey Rockwell, Voyant Tools is a user-friendly, entry-level tool described by its creators as being “for humanists who wish to spend more time exploring the corpus than learning complicated statistical and analytical software.”<sup>16</sup> Voyant Tools can be used for reading and analyzing large text corpora, which can be uploaded either by cutting and pasting the text directly or by selecting an electronic file (e.g., DOCX, PDF) from a local drive. Once the corpus has been loaded, Voyant Tools provides a variety of ways to analyze texts (see fig. 1). Default tools include the Cirrus, Reader, Trends, and Summary tools, though a range of additional tools are also available.<sup>17</sup> Among these tools, we were particularly interested in those that could be used to examine term frequency and use, which would allow for comparison of term use over time. The document terms grid in the Trends tools provided raw statistical usage data, with a count of each word’s frequency throughout each volume and time period. We also used the Collocates tool, which identifies terms with either identical or inverse frequency trends in order to understand connections between them. The Links tool highlights the strength of connections between selected terms in a more visual display. For example, it demonstrated the close relationship between the term “state” in *The American Archivist* and the terms “archives” and “records” based on statistical measures.

---

<sup>14</sup> The use of hyphens (-) in the journal was a significant issue that prevented Voyant Tools from clearly identifying words. The research assistant working on the project initially attempted to fix this problem manually, but later copied the text of the PDF files into Microsoft Word and replaced the hyphenated words with their actual counterparts. For example, “archiv-al” became “archival.” At the same time, characters such as symbols, single letters, and numbers were removed either by hand or programmatically. While we had planned to also remove common words to help facilitate the analysis, we found that Voyant Tools included a “stop-word” function that allows the software to ignore certain words. This removed a variety of numbers and symbols from the analysis, and we were able to further refine this functionality by editing the exclusion list. Before utilizing the adjusted “stop-word” function, the most frequent words in the text corpus were “the” (1,187,966), “of” (804,282), and “and” (556,514). After activating the “stop-word” function, the most frequent words in the corpus were the more predictable terms “records” (122,090), “archives” (117,743), and “archival” (65,550).

<sup>15</sup> Baayen, “Word Frequency Distributions.”

<sup>16</sup> Stéfan Sinclair and Geoffrey Rockwell, “Teaching Computer-Assisted Text Analysis: Approaches to Learning New Methodologies,” in *Digital Humanities Pedagogy: Practices, Principles, and Politics*, ed. Brett D. Hirsch (Cambridge: Open Book Publishers, 2012): 259. For an outsider review, see Megan E. Welsh, “Review of Voyant Tools,” *Collaborative Librarianship* 6 (2017): 96–97.

<sup>17</sup> For a full list of tools and associated documentation, see Voyant Tools, “Tools Index,” Voyant Tools, accessed March 3, 2017, <http://docs.voyant-tools.org/tools/>.



Figure 1. Voyant Tools analysis dashboard view.

To better understand the trends identified by the software, Voyant Tools also produces a range of visualizations. These visualizations allow researchers an alternative mode of interacting with statistical data generated by the tools, while suggesting further avenues of exploration. The Trends tool, for example, can graph word frequencies as a simple line graph of word use over time rather than being limited to providing numerical outputs. These graphs can display both the raw or relative frequencies of terms, or term stems, across all of the documents in a corpus or across segments in particular documents, depending on the options the user selects. Relative frequencies show the number of occurrences per normalized count of one million terms in a given year, while raw frequencies are the absolute number of occurrences of a term per year. Due to the quantity of words across each year and volume, it proved more accurate to display relative frequencies on the y-axis.<sup>18</sup> Graphs may also be produced using the relative frequency of multiple terms to allow easy comparison of their use. These graphs allowed us to begin to detect patterns within those word frequencies, and the visual nature of the output of the tool is easy to understand. For example, we were able to look at the relative use of terms commonly associated in the archival literature such as “access,” “preservation,” and “reference” to look at trends and connections between the use of these terms in *The American Archivist* (see fig. 2).

<sup>18</sup> When using relative frequencies, it is crucial to note the placement of the decimal in order to prevent a misreading of the results that could potentially skew interpretation of the data. Some of the graphs produced relative frequencies that were incremented by 0.0001 and others were incremented by 0.05. Depending on the actual distance between points, one seemingly volatile graph could actually be more level, and vice versa.

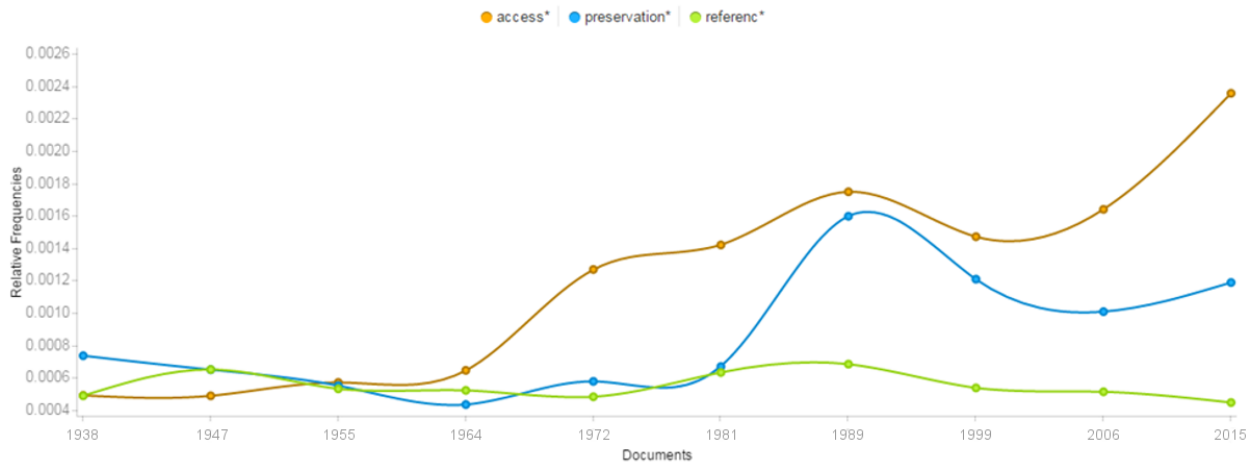


Figure 2. Relative frequency comparison of the terms access (“access\*”), preservation (“preservation\*”), and reference (“referenc\*”) between volumes 1 and 78. Smoothing of trend line applied by reduction in number of x-axis segments.

These relative frequency graphs also allowed a clear representation of the relationships between archival transmission standards and their associated content standards (see fig. 3) in *The American Archivist*’s content, or the shifting understanding of archival descriptive tools within the literature (see fig. 4).

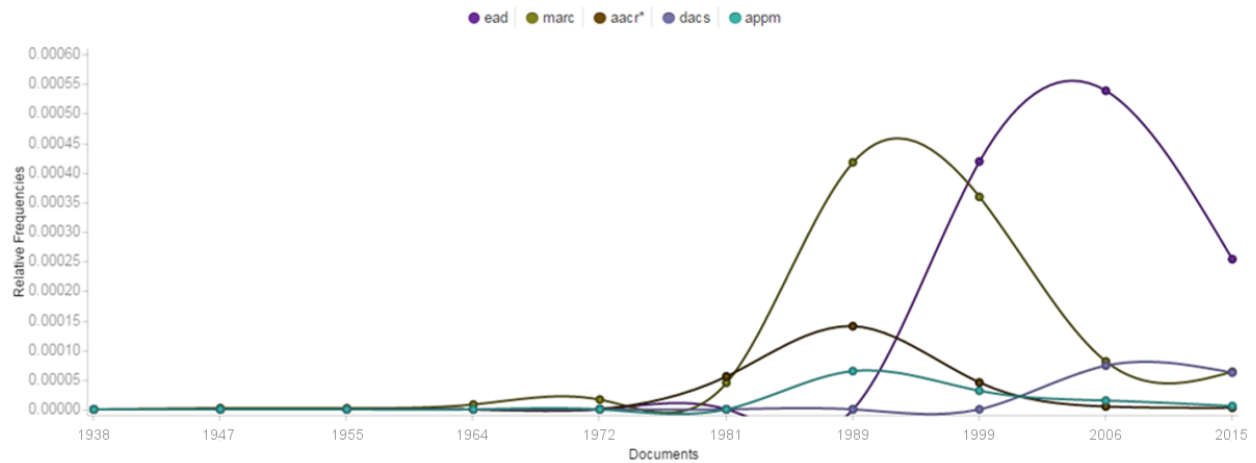


Figure 3. Relative frequency of the terms EAD, MARC, DACS, AACR (“aacr\*”), and APPM between volumes 1 and 78. Smoothing of trend line applied by reduction in number of x-axis segments.



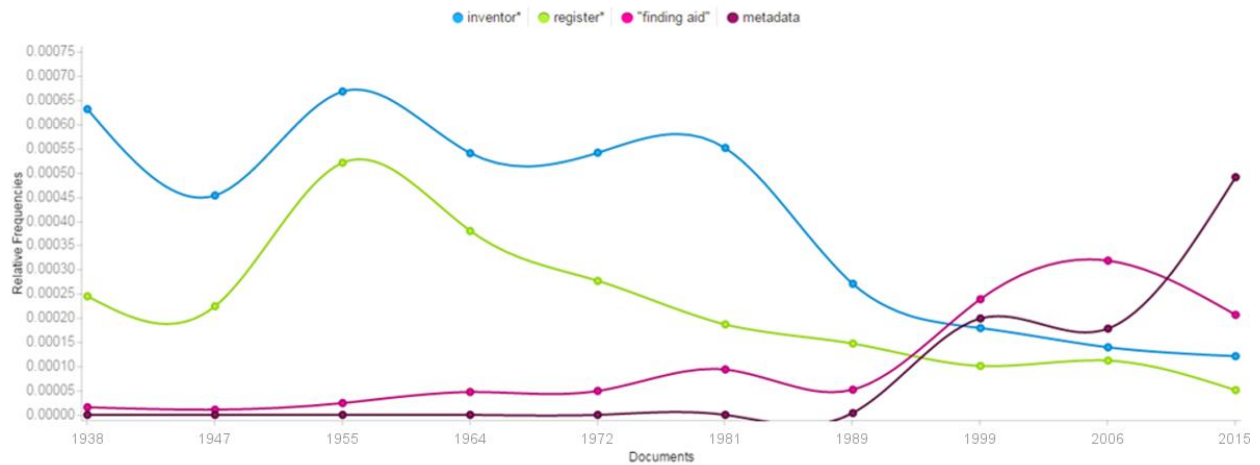


Figure 4. Relative frequency of the terms inventory (“inventor\*”), register (“register\*”), finding aid (“finding aid”), and metadata between volumes 1 and 78. Smoothing of trend line applied by reduction in number of x-axis segments.

The Bubblelines tool provides a different type of visualization of term usage within the corpus. Selected terms or term stems are shown as circles on a line, with their display location based on where they occur within the body of text (see fig. 5). When terms appear together, the size of the circle is larger in the display. By using multiple terms, researchers are able identify areas of collocation and potential relationships. Using the Bubblelines tool to search for terms with expected connections, such as “description” and “standards” (see fig. 5), allows researchers to identify periods of convergence between these concepts in the text of *The American Archivist*.

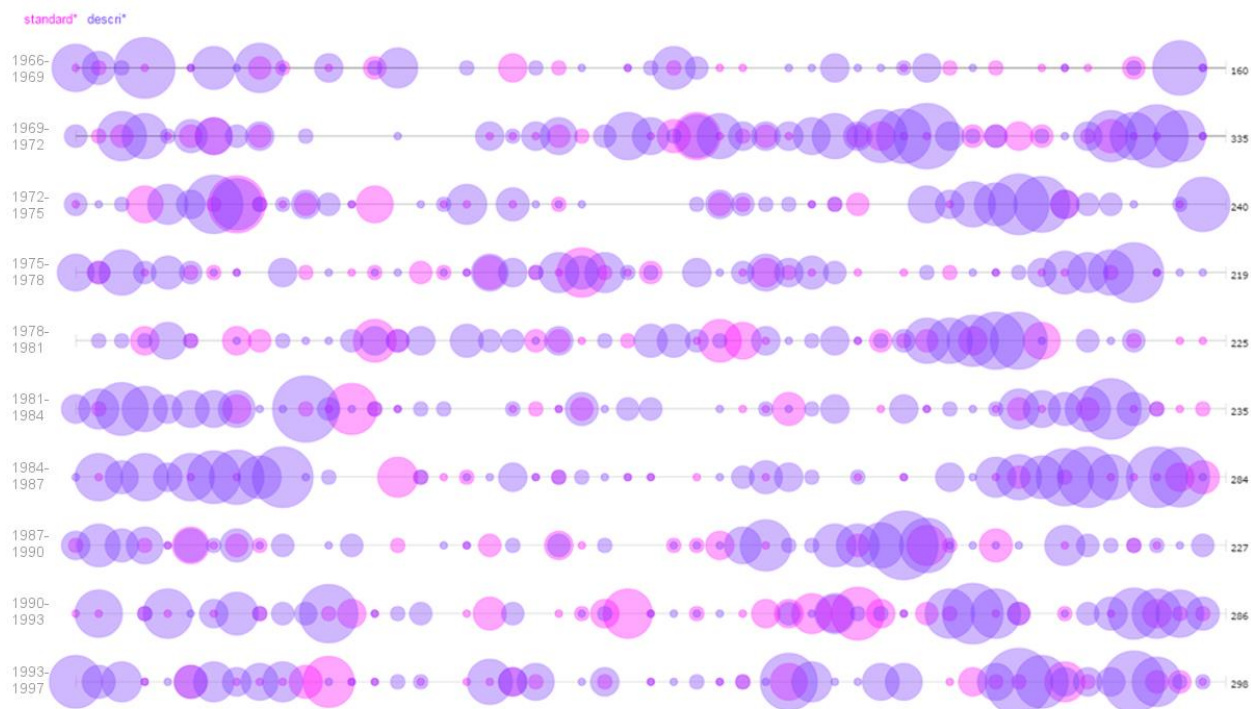


Figure 5. Bubblelines view of occurrence and collocation of terms for description (“descri\*”) and standards (“standard\*”) between approximately volumes 31 and 60 (1968–1997).

### Example: Archival theory<sup>19</sup>

In order to further explore the possibilities for corpus analysis, we decided to perform a limited study that explores the potential of these tools. The issue that we used to perform this exploration was the topic of archival theory as it appears in *The American Archivist*. Discussions of theory have been featured prominently in recent years in the journal.<sup>20</sup> The role of theory has also been prominently featured in many published histories of the American archives community, beginning with Richard C. Berner’s *Archival Theory and Practice in the United States: A*

<sup>19</sup> Several different concepts are represented by specific terms as part of this limited study. They include theory, archival theory, profession, research, standard, and certification. We have not attempted to define these terms and their larger concepts as we were only interested in the frequency with which they appeared in the text of the journal. We recognize that many of these concepts have contested meanings and that other terms have been used to describe the same sorts of activities referenced by the selected terms. A more in-depth study would need to define the concepts and broaden the search parameters in order to uncover all the ways that a specific concept was represented in the journal. Our point is to show the potential power of corpus analysis and to encourage its use by archivists.

<sup>20</sup> For example, see Paige Hohmann, “On Impartiality and Interrelatedness: Reactions to Jenkinsonian Appraisal in the Twentieth Century,” *American Archivist* 79, no. 1 (Spring-Summer 2016): 14–25; Alex H. Poole, “Archival Divides and Foreign Countries? Historians, Archivists, Information-Seeking, and Technology: Retrospect and Prospect,” *American Archivist* 78, no. 2 (Fall-Winter 2015): 375–433; Brian Shetler, “Archives and Recordkeeping: Theory into Practice,” *American Archivist* 78, no. 2 (Fall-Winter 2015): 581–84; and Scott Cline, “Archival Ideals and the Pursuit of a Moderate Disposition,” *American Archivist* 77, no. 2 (Fall-Winter 2014): 444–58.

*Historical Analysis* (1983), but also seen in more recent books.<sup>21</sup> This understanding of the history of archives is also included in official Society of American Archivists' publications, such as James M. O'Toole and Richard J. Cox's *Understanding Archives & Manuscripts*.<sup>22</sup> In these works, the authors have sought to treat the development of archival theory from a historical perspective and shed light on how archivists perceive the concept of theory and theorists.

According to many modern narratives, archival theory developed through the contributions of American and European archival professionals. According to O'Toole and Cox, archivists in the United States such as Margaret Cross Norton, Oliver Wendell Holmes, and Theodore Schellenberg created "a body of archival theory" based on the earlier work of Samuel Muller, J. A. Feith, and Robert Fruin in the Netherlands and Sir Hilary Jenkinson of Great Britain.<sup>23</sup> While this list of early archival theorists varies somewhat in available historical treatments, these seven authors are generally referenced as the most influential.

However, not everyone has accepted this view of the historical centrality of archival theory. Trevor Livelton has suggested that even some of the theorists identified by the literature were not particularly accepting of the idea.<sup>24</sup> Beginning in the early 1980s and carrying through the early 1990s the pages of *The American Archivist* provided a venue to air conflicting viewpoints on the issue.

Frank G. Burke initiated the discussion in a 1981 article in which he argued that it was vital that archivists pursue the development of archival theory. He explicitly tied archival theory to notions of professionalization and advocated for a formal definition of archival theory.<sup>25</sup> Harold T. Pinkett augmented the discussion by attempting to link American archival theory to European antecedents and began the process of defining what fell within the purview of archival theory.<sup>26</sup> Lester J. Cappon joined the discussion, asking "what, then, is there to theorize about?" He approached the question from the standpoint that the essential aspects of archival theory were already fully developed.<sup>27</sup>

John W. Roberts took issue with the whole concept of archival theory in a pair of articles in the late 1980s and early 1990s. He argued that "[t]here are two strains to what passes as archival theory: one strain is archival but not theoretical, and deals with the practical, how-to, nitty-gritty of archival work . . . The other is theoretical but not archival, and is concerned with

---

<sup>21</sup> Richard C. Berner, *Archival Theory and Practice in the United States: A Historical Analysis* (Seattle: University of Washington Press, 1983). More recent histories include John Ridener, *From Polders to Postmodernism: A Concise History of Archival Theory* (Sacramento: Library Juice Press, LLC, 2009); and Francis X. Blouin, Jr. and William G. Rosenberg, *Processing the Past: Contesting Authority in History and the Archives* (Oxford: Oxford University Press, 2012).

<sup>22</sup> James M. O'Toole and Richard J. Cox, *Understanding Archives & Manuscripts* (Chicago: Society of American Archivists, 2006).

<sup>23</sup> *Ibid.*, 65–66.

<sup>24</sup> Trevor Livelton, *Archival Theory, Records, and the Public* (Lanham, MD: Society of American Archivists and Scarecrow Press, 1996), 15.

<sup>25</sup> Frank G. Burke, "The Future Course of Archival Theory in the United States," *American Archivist* 44, no. 1 (Winter 1981): 40–46.

<sup>26</sup> Harold T. Pinkett, "American Archival Theory: The State of the Art," *American Archivist* 44, no. 3 (Summer 1981): 217.

<sup>27</sup> Lester Cappon, "What, Then, Is There to Theorize About?," *American Archivist* 45, no. 1 (Winter 1982): 19–25.

historiography.”<sup>28</sup> Roberts returned to his major point in another article, writing that discussion about “[a]rchival theory misses the whole point about what is important in the field, as it inevitably concentrates on what archivists do rather than on what they know.”<sup>29</sup> For Roberts, archival theory was a red herring distracting archivists from doing their practical work. Several archivists responded to Roberts, both in *The American Archivist* and other journals, and argued for the importance of archival theory.<sup>30</sup> Archival theory is now widely recognized as an important part of what makes an American archivist a professional.

### Findings

In light of these arguments, we determined to explore how archival theory has been treated across the publication history of *The American Archivist*. While historians have suggested that modern archival theory was the result of a long period of incremental development, looking at trends in *The American Archivist* text corpus suggests a different view of our professional past. With our aim of better understanding how corpus analysis tools could inform our understanding of the archival profession, we decided to do a limited examination of the use of the term “theory” in *The American Archivist*. Looking at the word stem “theor\*” in Voyant Tools, it was quickly apparent from relative frequency measures that the archival community did not refer to foundational archival concepts as “theory” in the literature until the late 1970s (see fig. 6).

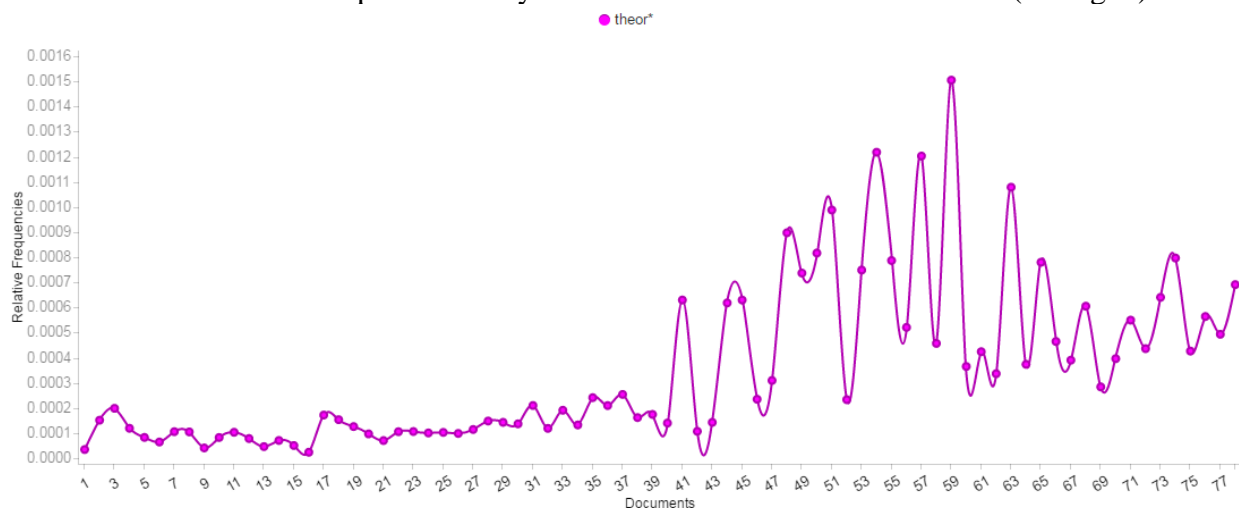


Figure 6. Relative frequency of term root “theor\*” by volume number.

This shift toward theory can also be seen in statistical figures for the term stem. According to the raw data, there were a total of 7,074 occurrences in the corpus as a whole. However, only 983 of these appeared before volume 40 (1977), while 6,091 were included in the journal after that time. While this dramatic increase does not necessarily suggest a lack of concern with theoretical

<sup>28</sup> John W. Roberts, “Archival Theory: Much Ado about Shelving,” *American Archivist* 50, no. 1 (Winter 1987): 67.

<sup>29</sup> John W. Roberts, “Archival Theory: Myth or Banality?,” *American Archivist* 53, no. 1 (Winter 1990): 119.

<sup>30</sup> For example, see Frederick Stielow, “Archival Theory Redux and Redeemed: Definition and Context Toward a General Theory,” *American Archivist* 54, no. 1 (Winter 1991): 14–26, and Terry Eastwood, “What Is Archival Theory and Why Is It Important?,” *Archivaria* 37 (1994): 122–30.

aspects of archival work prior to 1978, it does indicate that it became an increasingly important and explicit part of the professional discourse after that time.

Comparing the usage of “*theor\**” with other terms also indicates that this increase did not occur independently, but rather supports the view that archival theory was part of a larger movement within the community. Relative frequency graphs show a similar rise in publishing on professionalization, standardization, and certification beginning in the late 1970s, suggesting a connection between these concepts (see fig. 7). These relationships are also seen in the increased clustering of these terms in a Bubblelines view of the corpus (see fig. 8).

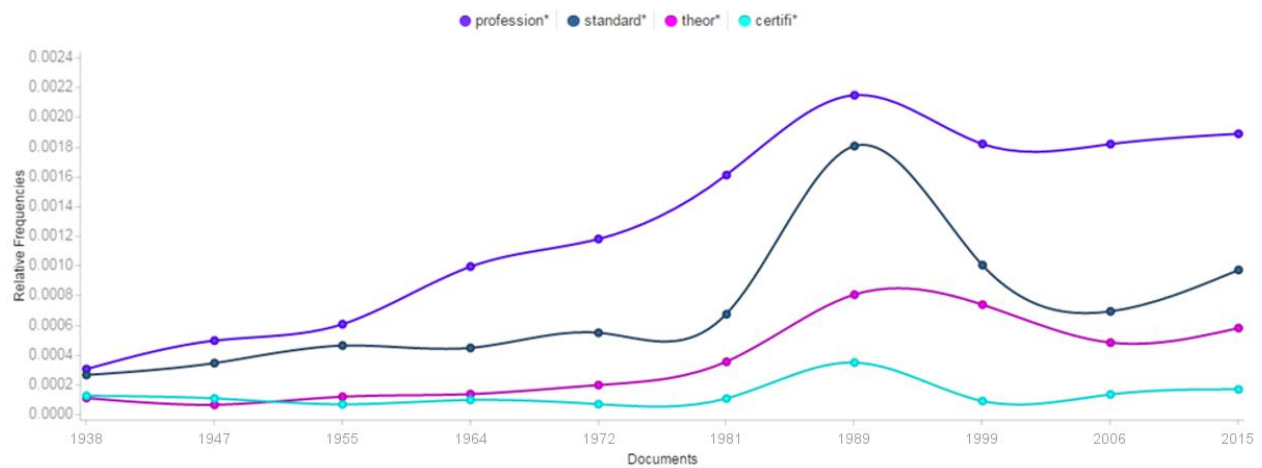


Figure 7. Relative frequency of the term stems “*profession\**,” “*standard\**,” “*theor\**,” and “*certifi\**” between volumes 1 and 78. Smoothing of trend line applied by reduction in number of x-axis segments.

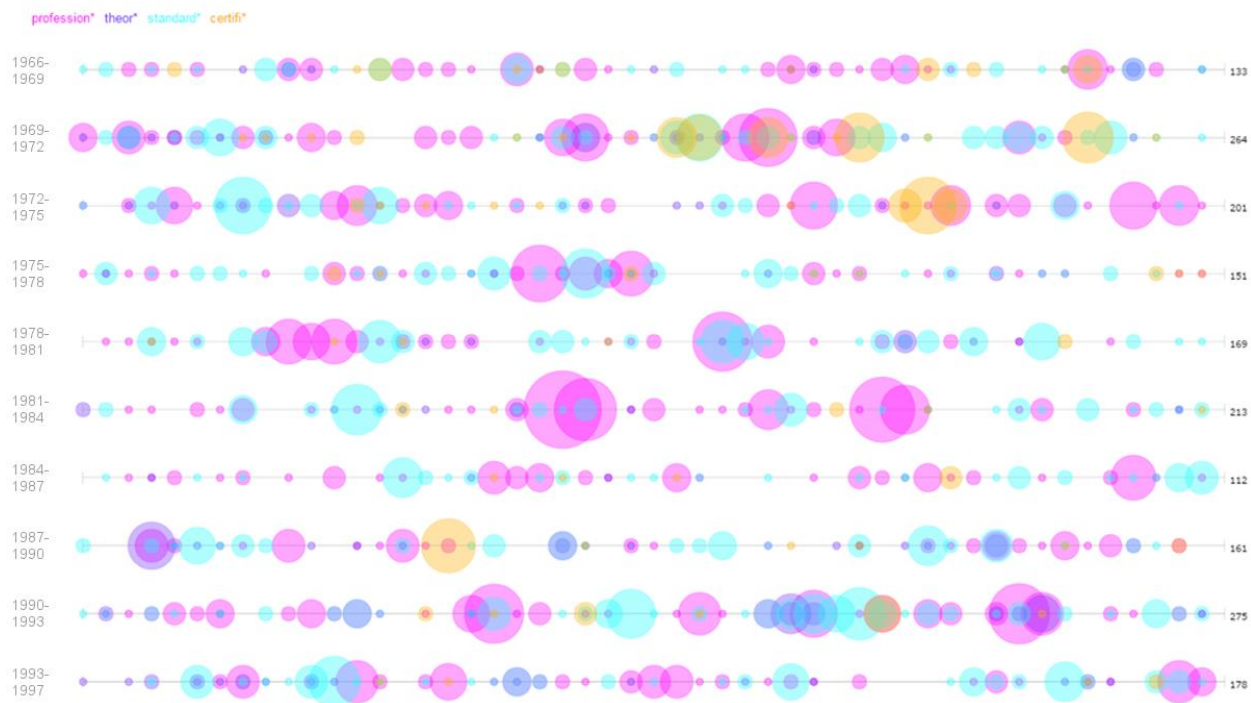


Figure 8. Bubblelines visualization of term frequency and collocation, between approximately volumes 31 and 60 (1968–1997).

In addition to examining theory as a concept, we also used Voyant Tools to examine how authors in *The American Archivist* have treated those theorists commonly identified by historians of the profession. Comparing a relative frequency graph of the use of prominent individuals’ names we found that there has been a declining interest in Margaret Cross Norton and Oliver Wendell Holmes over time, while the *Dutch Manual*, Sir Hilary Jenkinson, and T. R. Schellenburg have remained influential (see fig. 9).

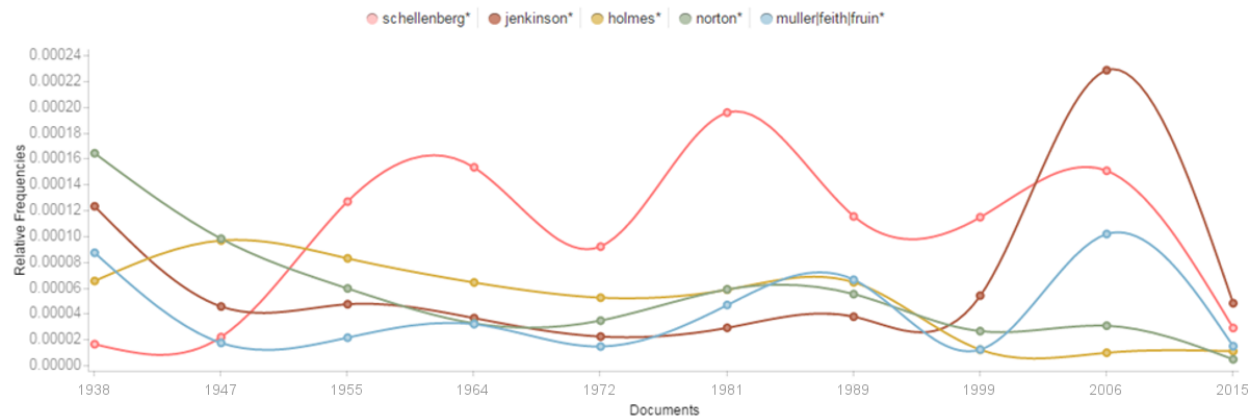


Figure 9. Relative frequency of archival theorist names in the *American Archivist* between volumes 1 and 78. Smoothing of trend line applied by reduction in number of x-axis segments.

Based on these Voyant Tools results, it would appear that archival practice as “theory” was not a subject of contemporary discussion among the founding members of the profession. Rather, these results suggest that as archivists sought to consolidate and reorganize through the Society of American Archivists in the late 1970s, the concept of archival theory was developed in the context of professionalization. This effort involved a reconsideration of existing archival practice, and a reinterpretation of earlier archivists as archival theorists.

The focus on theory and professionalization during this period was not necessarily specific to archives, as other disciplines sought to formalize their methods in order to improve their status. The literature of closely allied fields such as history and sociology include similar discussions of the role of theory.<sup>31</sup> Looking at the overall corpus of English-language text illustrates the increased focus on “theory” in the postwar period (see fig. 10).

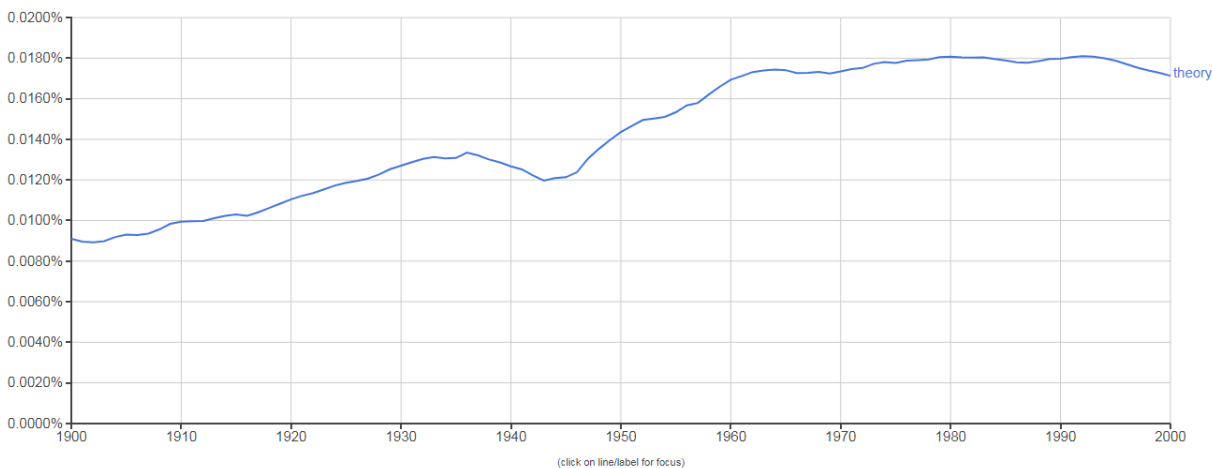


Figure 10. Google Ngram of occurrence of the term “theory” during the period between 1900 and 2000.

Corpus analysis tools suggest a close association between theory and discussions about professionalization in the archival community in the 1970s, 1980s, and early 1990s. Additional research methodologies should be used to explore this suggested association between theory and professionalization. The major point of this limited study is to introduce software applications such as Voyant Tools that might help us to visualize potential connections between discussions in the professional literature. These tools can be extremely useful for identifying patterns and trends that deserve deeper exploration.

<sup>31</sup> For examples, see “History and Theory,” *History Workshop* 6 (Autumn 1978): 1–6, or Kenneth C. Land, “Formal Theory,” *Sociological Methodology* 3 (1971): 175–220.

## Conclusion

As seen through this limited study, corpus analysis provides new tools for considering the role of *The American Archivist* on the development of the profession. It helps us better understand what the content of the journal reveals about the importance of various topics and shows how discussions in *The American Archivist* can better illuminate the growth of the profession. Linguistic approaches allow researchers a way other than traditional bibliometric approaches to explore sources, point out emerging trends, and suggest alternative narratives.<sup>32</sup> By pursuing a fuller analysis of the journal's content, archivists will be able to better understand its role and impact on the profession.

It is our hope that other archivists will utilize corpus analysis tools to explore not only *The American Archivist* but other related journals in order to discover and better understand our professional community. There is great value in connecting the past to the present and corpus analysis tools give us another way to do this. The archival professional community has a rich heritage and dynamic literature that deserves further exploration and explication.

## Bibliography

Baayen, R. Harald. *Word Frequency Distributions*. Dordrecht, Netherlands: Kluwer Academic Publishers, 2001.

Berner, Richard C. *Archival Theory and Practice in the United States: A Historical Analysis*. Seattle, WA: University of Washington Press, 1983.

Blouin, Francis X., Jr. and William G. Rosenberg. *Processing the Past: Contesting Authority in History and the Archives*. Oxford: Oxford University Press, 2012.

Budgell, Brian S., Alice Kwong, and Neil Millar. "A Diachronic Study of the Language of Chiropractic," *Journal of the Canadian Chiropractic Association* 57, no. 1 (2013): 49–55.

Burke, Frank G. "The Future Course of Archival Theory in the United States," *American Archivist* 44, no. 1 (Winter 1981): 40–46.

Cappon, Lester. "What, Then, Is There to Theorize About?," *American Archivist* 45, no. 1 (Winter 1982): 19–25.

Cline, Scott. "Archival Ideals and the Pursuit of a Moderate Disposition," *American Archivist* 77, no. 2 (Fall-Winter 2014): 444–458.

---

<sup>32</sup> In another recent project, Voyant Tools was used to explore the relationship between finding aid content and online use of that content (see Ryan K. Lee et al., "Using Google Analytics, Voyant and Other Tools to Better Understand Use of Manuscript Collections at L. Tom Perry Special Collections," *Practical Technology for Archives* 7 (January 2017), accessed March 7, 2017, [https://practicaltechnologyforarchives.org/issue7\\_lee](https://practicaltechnologyforarchives.org/issue7_lee)).



Conway, Paul and William E. Landis. "Open-Access Publishing and the Transformation of the *American Archivist* Online," *American Archivist* 74, no. 2 (Fall-Winter 2011): 482–505.

Cox, Richard J. "American Archival Literature: Expanding Horizons and Continuing Needs, 1901–1987," *American Archivist* 50, no. 3 (Summer 1987): 306–323.

Cox, Richard J. "An Analysis of Archival Research, 1970–92, and the Role and Function of the *American Archivist*," *American Archivist* 57, no. 2 (Spring 1994): 278–288.

Drouin, Jeffrey. "Close- and Distant-Reading Modernism: Network Analysis, Text Mining, and Teaching the *Little Review*," *The Journal of Modern Periodical Studies* 5, no. 1 (2014): 110–135.

Eastwood, Terry. "What is Archival Theory and Why Is It Important?," *Archivaria* 37 (1994): 122–130.

Harper, Wakefield. "A Content Analysis of Archival Journal Literature" Master's thesis, University of North Carolina at Chapel Hill, 2010.

"History and Theory," *History Workshop* 6 (Autumn 1978): 1–6.

Hohmann, Paige. "On Impartiality and Interrelatedness: Reactions to Jenkinsonian Appraisal in the Twentieth Century," *American Archivist* 79, no. 1 (Spring-Summer 2016): 14–25.

Hussain, Akhtar, Nishat Fatima, and Devendra Kumar. "Bibliometric Analysis of the 'Electronic Library' Journal (2000–2010)," *Webology* 8, no. 1 (June 2011). Accessed January 20, 2017, <http://www.webology.org/2011/v8n1/a87.html>.

Land, Kenneth C. "Formal Theory," *Sociological Methodology* 3 (1971): 175–220.

Lee, Ryan K., Cory L. Nimer, J. Gordon Daines III, and Shelise Rupp. "Using Google Analytics, Voyant and Other Tools to Better Understand Use of Manuscript Collections at L. Tom Perry Special Collections," *Practical Technology for Archives* 7 (January 2017). Accessed March 7, 2017, [https://practicaltechnologyforarchives.org/issue7\\_lee/](https://practicaltechnologyforarchives.org/issue7_lee/).

Livelton, Trevor. *Archival Theory, Records, and the Public*. Lanham, MD: Society of American Archivists and Scarecrow Press, 1996.

McCarthy, Michael and Ann O'Keefe, "Historical Perspective: What Are Corpora and How Have They Evolved?" In *The Routledge Handbook of Corpus Linguistics*, edited by Anne O'Keefe and Michael McCarthy. London and New York: Routledge, 2010.

Mayer-Schönberger, Viktor and Kenneth Cukier. *Big Data: A Revolution That Will Transform How We Live, Work, and Think*. Boston, MA: Houghton Mifflin Harcourt, 2012.

Michel, Jean-Baptiste, Yuan Kui Shen, Aviva Presser Aiden, Adrian Veres, Matthew K. Gray, The Google Books Team, Joseph P. Picket, Dale Hoiberg, Dan Clancy, Peter Norvig, Jon

Orwant, Steven Pinker, Martin A. Nowak, and Erez Lieberman Aiden. "Quantitative Analysis of Culture Using Millions of Digitized Books," *Science* 331, no. 6014 (January 2011): 176–182.

O'Reilly, Dermot and Mike Reed. "The Grit in the Oyster: Professionalism, Managerialism, and Leaderism as Discourses of UK Public Services Modernization," *Organization Studies* 32, no. 8 (2011): 1079–1101.

O'Toole, James M. and Richard J. Cox. *Understanding Archives & Manuscripts*. Chicago, IL: Society of American Archivists, 2006.

Pinkett, Harold T. "American Archival Theory: The State of the Art," *American Archivist* 44, no. 3 (Summer 1981): 217–222.

Poole, Alex H. "Archival Divides and Foreign Countries? Historians, Archivists, Information-Seeking, and Technology: Retrospect and Prospect," *American Archivist* 78, no. 2 (Fall-Winter 2015): 375–433.

Ridener, John. *From Polders to Postmodernism: A Concise History of Archival Theory*. Sacramento, CA: Library Juice Press, LLC, 2009.

Roberts, John W. "Archival Theory: Much Ado about Shelving," *American Archivist* 50, no. 1 (Winter 1987): 66–74.

Roberts, John W. "Archival Theory: Myth or Banality?," *American Archivist* 53, no. 1 (Winter 1990): 110–120.

Shetler, Brian. "Archives and Recordkeeping: Theory into Practice," *American Archivist* 78, no. 2 (Fall-Winter 2015): 581–584.

Sinclair, John. *Corpus, Concordance, Collocation*. Oxford: Oxford University Press, 1991.

Sinclair, Stéfan and Geoffrey Rockwell. "Teaching Computer-Assisted Text Analysis: Approaches to Learning New Methodologies." In *Digital Humanities Pedagogy: Practices, Principles, and Politics*, edited by Brett d. Hirsch. Cambridge, MA: Open Book Publishers, 2012.

Sinclair, Stéfan and Geoffrey Rockwell. "Text Analysis and Visualization: Making Meaning Count." In *A New Companion to Digital Humanities*, edited by Susan Schreibman, Ray Siemens, and John Unsworth. Malden, MA: Wiley/Blackwell, 2016.

Singh, K. P. and Harish Chander. "Publication Trends in Library and Information Science: A Bibliometric Analysis of *Library Management* Journal," *Library Management* 35, no. 3 (2014): 134–149.

Stephenson, Mary Sue. "The American Archivist, 1971 to 1990: A Demographic Analysis of the Articles," *American Archivist* 55, no. 4 (Fall 1992): 538–561.

Stielow, Frederick. "Archival Theory Redux and Redeemed: Definition and Context Toward a General Theory," *American Archivist* 54, no. 1 (Winter 1991): 14–26.

Tranter, Kate. "Corpus Analysis of Historical Documents." In *The Encyclopedia of Applied Linguistics*, edited by Carol A. Chapelle. Hoboken, NJ: John Wiley and Sons, 2014.

Trever, Karl L. "The American Archivist: The Voice of a Profession," *American Archivist* 15, no. 2 (April 1952): 147–155.

Tsay, Ming-yueh and Zhu-ye Shu. "Journal Bibliometric Analysis: A Case Study on the *Journal of Documentation* 67, no. 5 (2011): 806–822.

Uboldi, Giorgio and Giorgio Caviglia. "Information Visualizations and Interfaces in the Humanities." In *New Challenges for Data Design*, edited by David Bihanic. London: Springer, 2015.

Vanchena, Lorie A. "Reading German Culture, 1789–1918." Presentation at "Distant Readings/Descriptive Turns: Topologies of German Culture in the Long Nineteenth Century," 21st St. Louis Symposium on German Literature & Culture, Washington University in St. Louis, March 29–31, 2012.

Voyant Tools, "Tools Index." Accessed March 3, 2017, <http://docs.voyant-tools.org/tools/>.

Voyant Tools, "Voyant: See Through Your Text." Accessed February 2, 2017, <http://voyant-tools.org/>.

Webster, Jessica Wagner. "'Filling the Gaps': Oral Histories and Underdocumented Populations in *The American Archivist*, 1938–2011," *American Archivist* 79, no. 2 (Fall-Winter 2016): 254–282.

Welsh, Megan E. "Review of Voyant Tools," *Collaborative Librarianship* 6 (2017): 96–97.