

Searching for Anglo-American Digital Legal History

STEPHEN ROBERTSON

As the fields of digital humanities and digital history have grown in scale and visibility since the 1990s, legal history has largely remained on the margins of those fields.¹ The move to make material available online in the first decade of the web featured only a small number of legal history projects: *Famous Trials*; *Anglo-American Legal Tradition*; *The Proceedings of*

1. I make no claim to have comprehensively surveyed digital legal history. There is no register or compilation of work in digital history, let alone of digital legal history; therefore, this overview by necessity is focused on the area of legal history in which I work and with which I am familiar: Anglo-American legal history. In looking for digital legal history, I drew on the list of projects compiled in 2013 by Kaci Nash and William G. Thomas III for Thomas' "The Promise of the Digital Humanities and the Contested Nature of Digital Scholarship," in *A New Companion to Digital Humanities*, ed Susan Schreibman, Ray Siemens, and John Unsworth (Chichester: Wiley Blackwell, 2016), 603–17. Their *Digitalhistory Zotero Library* can be found at <https://www.zotero.org/groups/digitalhistory>. For an overview of digital history, see Stephen Robertson, "The Differences between Digital Humanities and Digital History," in *Debates in the Digital Humanities 2016*, ed. Matt Gold and Lauren Klein (Minneapolis: University of Minnesota Press) <http://dhde.bates.gc.cuny.edu/debates/text/76>. Accessed 24 July 2016.

Stephen Robertson is the director of the Roy Rosenzweig Center for History and New Media, and a professor, in the Department of History and Art History at George Mason University <srober30@gmu.edu>. He is the author of *Crimes against Children: Sexual Violence and Legal Culture in New York City, 1880–1960* (Chapel Hill: University of North Carolina Press, 2005) and, with Shane White, Stephen Garton, and Graham White, *Playing the Numbers: Gambling in Harlem Between the Wars* (Cambridge, MA: Harvard University Press, 2010) as well as of *Digital Harlem: Everyday Life, 1915–1930*, <http://digitalharlem.org>. He thanks Lincoln Mullen for his comments on earlier versions of this article.

the *Old Bailey Online, 1674–1913*.² Early efforts to construct hypertext narratives and scholarship also included some works of legal history: “Hearsay of the Sun: Photography, Identity and the Law of Evidence in Nineteenth-Century Courts,” in *Hypertext Scholarship in American Studies; Who Killed William Robinson?* and *Gilded Age Plains City: The Great Sheedy Murder Trial and the Booster Ethos of Lincoln, Nebraska*.³ In the second decade of the web, the focus shifted from distributing material to exploring it using digital tools.⁴ The presence of digital history grew at the meetings of organizations of historians ranging from the American Historical Association to the Urban History Association, but not at the American Society for Legal History conferences, the annual meetings of the Law and Society Association, or the British Legal History Conference.⁵ Only a few Anglo-American legal historians took up computational tools for sorting and visualizing sources such as data mining, text mining, and topic modeling; network analysis; and mapping.⁶ Paul Craven and Douglas Hay’s *Master and Servant* project text mined a comprehensive database of 2,000 statutes and 1,200,000 words to explore similarities and influence among statutes.⁷ *Data Mining with Criminal Intent* mined and visualized the

2. *Anglo-American Legal Tradition: Documents from Medieval and Early Modern England from the National Archives in London* <http://aalt.law.uh.edu>; *Famous Trials* <http://law2.umkc.edu/faculty/projects/frtrial/frtrial.htm>; and *The Proceedings of the Old Bailey* <http://www.oldbaileyonline.org>. Accessed 24 July 2016.

3. Thomas Thurston, “Hearsay of the Sun: Photography, identity and the Law of Evidence in Nineteenth-Century Courts,” *Hypertext Scholarship in American Studies* <http://chnm.gmu.edu/aq/photos/index.htm>; *Who Killed William Robinson?* <http://historyarthistory.gmu.edu/people/shamdani> (now part of a collection of *Great Unsolved Mysteries in Canadian History* <http://www.canadianmysteries.ca/en/>); *Gilded Age Plains City: The Great Sheedy Murder Trial and the Booster Ethos of Lincoln, Nebraska*, <http://gildedage.unl.edu>. Accessed 24 July 2016.

4. Daniel Cohen nicely captures this change as involving a conceptual shift from discussing the web using nouns such as “web pages” and “web sites,” to using verbs such as “searching,” “sorting,” “gathering,” and “communicating.” See Daniel J. Cohen, “History and the Second Decade of the Web,” *Rethinking History* 8 (2004): 295.

5. For digital history at the meetings of American historical organizations, see Robertson, “Differences,” para. 2.

6. These tools are defined and discussed later. For an introduction to text mining and topic modeling for historians, see Shawn Graham, Ian Milligan, and Scott Weingart, *The Historian’s Macroscope: Exploring Big Historical Data* (London: Imperial College Press, 2015).

7. Paul Craven and William Traves, “A General-Purpose Hierarchical Coding Engine and its Application to Comparative Analysis of Statutes,” *Literary and Linguistic Computing* 8 (1993): 27–32; Paul Craven and Douglas Hay, “Computer Applications in Comparative History: The Master & Servant Project at York University (Canada),” *History and Computing* 7 (1995): 69–80; and Douglas Hay and Paul Craven, “Introduction,” in *Masters, Servants, and Magistrates in Britain and the Empire, 1562–1955*, ed. Hay and

words in trial records using structured data from *The Proceedings of the Old Bailey Online, 1674–1913*. *Locating London's Past*, a project that mapped resources relating to the early modern and eighteenth century city, and also made use of the Old Bailey records.⁸ *Digital Harlem* mapped crime in the context of everyday life in the 1920s.⁹ Only in the past few years has more digital legal history using computational tools begun to appear, and like many of the projects discussed in this special issue, most remain at a preliminary stage.¹⁰ This article seeks to bring into focus the constraints, possibilities, and choices that shape digital legal history, in order to create a context for the work in this special issue, and to promote discussion of what it means to do legal history in the digital age.

The dearth of digital legal history is particularly striking, given that legal history is better positioned for a digital turn than most historical fields when it comes to the amenability of legal sources to computational analysis and the availability of those sources in digitized forms. The consistent forms of legal sources such as statutes, court records, trials, and judicial opinions give structure to the information they contain. Although those records have changed in shape and substance, those variations retain sufficient structure to allow them to be compared over time. Legal sources were published frequently, and survive in comprehensive collections. The language of those records is marked by a repetitious and highly specialized vocabulary that gives legal texts many standardized elements. In addition, that language also mitigates one of the limitations of computational text analysis: that it processes words, but because words and

Craven (Chapel Hill: University of North Carolina Press, 2004). This database has not been made available online.

8. *The Proceedings of the Old Bailey Online, 1674–1913* <http://www.oldbaileyonline.org/>; *Locating London's Past* <http://www.locatinglondon.org/index.html>; *Data Mining with Criminal Intent* <http://criminalintent.org>. Accessed 24 July 2016.

9. *Digital Harlem: Everyday Life, 1915–1930* <http://digitalharlem.org>. Accessed 24 July 2016.

10. William Thomas, *O Say Can You See: Early Washington, D.C., Law and Family* <http://earlywashingtondc.org>; Kellen Funk and Lincoln Mullen, "A Servile Copy: Text Reuse and Medium Data in American Civil Procedure," in Forum: Die geisteswissenschaftliche Perspektive: Welche Forschungsergebnisse lassen Digital Humanities erwarten? [Forum: With the Eyes of a Humanities Scholar: What Results Can We Expect from Digital Humanities?], *Rechtsgeschichte* 24 [Legal History] (forthcoming, 2016); Adam Badawi and Rend Bod, "Legal Structures," *2013 Digging Into Data Challenge* <http://diggingintodata.org/awards/2013/project/legal-structures>; Lea VanderVelde, *The Law of the Antebellum Frontier* <http://web.stanford.edu/group/spatialhistory/cgi-bin/site/project.php?id=1057>; Stephen Berry, *CSI: Dixie* <https://csidixie.org>; and John Blanton, Micki Kaufman, and Nora Slonimsky, "An Analysis of Three Editions of the Blackstone Legal Commentaries Using Computational Text Analysis" <http://mickikaufman.com/BLACKSTONE-POSTER.pdf>. Accessed 24 July 2016.

meaning have no easy correspondence, identifying patterns in words does not always offer a clear picture of the information that documents contain. However, the highly technical nature of legal language means that the correlation between words and meaning is much higher than in most textual sources, and the results of computational text analysis of legal sources are more revealing of their contents.

Many legal history sources already exist in digital formats that can be used with computational tools. Legal records were some of the first historical documents transformed by databases and digitization. To take just the American example, LexisNexis and Westlaw introduced computerized databases in the 1970s, and have progressively expanded them to include all published federal court decisions and the decisions of the higher state courts. Gale's *Making of Modern Law* databases include 22,000 English and American legal treatises, published trials, and United States Supreme Court briefs and records. HeinOnline includes a law reports, treatises, and a wide range of session bills, statutes, and published legal sources. Much of the published material from the years before 1923 can also be accessed through HathiTrust's digital library and Google Books.¹¹

The nature of the legal sources that have been digitized has contributed to the limited amount of digital legal history. These sources are overwhelmingly from inside the law—case law, statutes, treatises, trials—whereas legal history since the 1970s has been increasingly focused on the relationship between law and the wider society of which it is part.¹² The “law and. . .” approach requires additional sources beyond those generated by the legal system. Far less of that material has been digitized. Much of the digital historical record that does exist consists of periodical literature.¹³ The vast majority of archival material has not been digitized.¹⁴

Although all the sources for legal history are not digitized, the wealth of databases of legal material nonetheless means that scholars who study the law are likely among the historians who frequently conduct searches of

11. *The Making of Modern Law* <http://gdc.gale.com/products/the-making-of-modern-law-primary-sources-1620-1926/>; HeinOnline <http://heinonline.org>; HathiTrust Digital Library <https://www.hathitrust.org>; and Google Books, <https://books.google.com>. For additional online primary sources, see the list on *Legal History on the Web* https://law.duke.edu/legal_history/portal/primary-sources.html. Accessed 24 July 2016.

For an older survey, see Morris Cohen, “Researching Legal History in the Digital Age,” *Law Library Journal* 99 (2007): 377–93.

12. For a recent summary overview, see Alfred L. Brophy and Stefan Vogenauer, “Introducing the Future of Legal History: On Re-launching the *American Journal of Legal History*,” *American Journal of Legal History* 5 (2016): 1–5.

13. Lara Putnam, “The Transnational and the Text-Searchable: Digitized Sources and the Shadows They Cast,” *American Historical Review* 121 (2016): 390.

14. A common rough estimate is that at most 5% of archival material has been digitized.

databases as part of their research. That is an assumption because, as is the case across historical fields, there is no discussion of search as a research method in legal history scholarship. Nonetheless, studies of the research practices of historians report widespread use of searches, beginning with Google searches to identify sources, and proceeding to full text keyword searching to research within digitized collections and documents.¹⁵ The failure to discuss these searches in scholarship implicitly treats them as “a finding aid analogous to a catalog.” That characterization was somewhat true when searching focused on metadata. However, from the mid-2000s, the widespread use of optical character recognition (OCR) software to turn images of documents into machine readable text brought a shift to full-text search, “a name for a large family of algorithms that humanists have been using for decades to test hypotheses and sort documents by relevance to their hypotheses.”¹⁶

Recognizing that full-text search is a computational tool highlights that much legal history is at least inflected with digital history, and requires more attention to be given to how database searches work and what it means to use searching as a research tool.¹⁷ A search will produce different results depending upon how the searchable text was generated. Although the document—the image of the page—is the same regardless of the database, machine-readable words can be generated from that image in

15. Jennifer Rutner and Roger Schonfeld, *Supporting the Changing Research Practices of Historians*, ITHAKA S+R, 2012 <http://www.sr.ithaka.org/sites/default/files/reports/supporting-the-changing-research-practices-of-historians.pdf>; Max Kemman, Martijn Kleppe, and Stef Scagliola, “Just Google It,” in *Proceedings of the Digital Humanities Congress 2012*, ed. Clare Mills, Michael Pidd, and Esther Ward (Sheffield: HRI Online Publications, 2014) <http://www.hrionline.ac.uk/openbook/chapter/dhc2012-kemman> (accessed 24 July 2016); and Alexandra Chassanoff, “Historians and the Use of Primary Sources in the Digital Age,” *The American Archivist* 76 (2013): 458–80.

16. Putnam, “Transnational and the Text-Searchable”; and Ted Underwood, “Theorizing Research Practices We Forgot to Theorize Twenty Years Ago,” *Representations* 127 (2014): 65. On OCR, see Simon Tanner, “Deciding Whether Optical Character Recognition is Feasible,” 2004 http://www.odl.ox.ac.uk/papers/OCRFeasibility_final.pdf. Accessed 24 July 2016.

17. By contrast, the impact of searching and computerized databases on legal research has provoked extensive and ongoing discussion and debate. See, for example, Carol Bast and Ransford Pyle, “Legal Research in the Computer Age: A Paradigm Shift?” *Law Library Journal* 93 (2001): 285–302; F. Allan Hanson, “From Key Numbers to Keywords: How Automation Has Transformed the Law,” *Law Library Journal* 94 (2002): 563–600; and John McGinnis and Steven Wasick, “Law’s Algorithm,” *Florida Law Review* 66 (2014): 991–1050. It is hoped that the appearance of Lara Putnam’s article on the impact of searchable databases on transnational history in the *American Historical Review*, the field’s leading journal, will provoke an overdue discussion of searching (Putnam).

different ways: by transcription (the results of which vary depending upon the individual transcriber) and by OCR, the results of which vary by software and the extent of efforts to correct errors. A search of the same documents in different databases can, therefore, produce different results.¹⁸ The results a search returns also depend upon what word or phrase is searched for and in which part of a document, and, depending upon the database, whether searches for phrases are “Boolean searches (United AND States) or exact-phrase searches made with database-specific delimiters (United States).”¹⁹ Searching also struggles to deal with what lies outside a set of results. In returning only the terms one enters, a search filters out any alternative hypotheses. For historians, this poses particular challenges, as the language and ways of organizing knowledge in the past often differ significantly from contemporary terms and patterns of thought. If scholars use the wrong search terms, they literally misread their sources, and might not read them at all. Moreover, when working with interfaces that indicate how many results were found without reference to how many results were possible, it is not always clear just how significant those results might be. As Ted Underwood notes, “in a database containing millions of sentences, full-text search can turn up twenty examples of anything.”²⁰

In addition to enabling full-text search, the words produced by digitization can be used with other computational tools. As Lara Putnam points out, this “mass data-fication of words” both associates digital history with earlier forms of quantitative history and distinguishes it from them. It is not historians’ ability to undertake computational analysis that is new; it is that the data available for computational analysis now includes

18. For example, Drew VandeCreek noted that the corrected text of the *Congressional Record* available in *Proquest Congressional* “contained a very small amount of scanning errors, significantly fewer than those found in the portion of the [University of North Texas Libraries uncorrected] data that I reviewed, and about the same as the Hein materials.” See VandeCreek, “Text Mining at an Institution with Limited Financial Resources,” *D-Lib Magazine* 22 (2016) <http://www.dlib.org/dlib/july16/vandecreek/07vandecreek.html>. Accessed 24 July 2016. Crucially, commercial vendors do not provide information on the OCR accuracy of their products, or on how they correct the text. This situation is complicated by the fact that, as Matthew Jockers and Ted Underwood note, “since different kinds of errors have radically different effects, there is no single accuracy percentage that proves a text is good enough to support analysis.” Jockers and Underwood, “Text-Mining the Humanities,” in *The New Companion to Digital Humanities*, ed Susan Schreibman, Ray Siemens, and John Unsworth (Chichester: Wiley Blackwell, 2016), 359. The *Beyond Citation* project (<http://www.beyondcitation.org>) is attempting to address scholars’ need for more information about the proprietary databases of digitized material on which humanities scholars rely.

19. Caleb McDaniel, “The Digital Early Republic,” *Offprints*, April 7, 2011 <http://mcdaniel.blogs.rice.edu/?p=150>. Accessed 24 July 2016.

20. Underwood, “Theorizing Research Practices,” 66.

words.²¹ However, few legal historians have used computational tools other than searchable databases because most digitized legal sources have not been made available for use with such tools. The databases of vendors such as LexisNexis, HeinOnline, ProQuest, and Gale currently neither offer tools for text mining nor allow the access to their contents that is necessary to obtain data for text mining. As Tanenhaus and Nystrom put it in their article in this issue, “The sort of digital access appropriate for a traditional user, including limitations posed by the quantity of material available, the size and format of permitted downloads, the ease of machine interaction with the download interface, and the cost of licensing, can be a stumbling block for legal historians who want to conduct additional computational analysis.” They were able to obtain the session laws they needed by negotiating to exceed the download limit imposed by HeinOnline, but could not undertake digital analysis of newspapers because their access was blocked after downloading only 18 months of publications.²² Currently, vendors such as ProQuest and Gale are imposing additional charges for the access required for text mining, and even then, delivering that content on hard drives, with terms of use that limit the text mining that can be done.²³ Developing large-scale open access collections as an alternative to proprietary databases is a challenging and expensive activity. As Andrew Prescott notes, *The Proceedings of the Old Bailey Online, 1674–1913* required funding from four separate agencies and three different universities, and involved a team of twenty-two people, an infrastructure that “has more in common with filmmaking than old-style academic publishing.”²⁴ In the absence of sufficient public funding, an alternative is to partner with commercial vendors, as Harvard Law School Library has with Ravel Law to digitize United States case law. This project will make approximately 40,000,000 pages of material freely available; however, the bulk access

21. Putnam, “Transnational and the Text-Searchable” 400.

22. Eric Nystrom and David Tanenhaus, “The Future of Digital Legal History: No Magic, No Silver Bullets,” *American Journal of Legal History* 56 (2016): 158.

23. VandeCreek, “Text Mining.”; Andrew Prescott, “What Price Gale Cengage?” *Digital Riffs*, July 15, 2016 <https://medium.com/digital-riffs/what-price-gale-cengage-668d358ce5cd#11e0j5xp>. Accessed 24 July 2016.

24. Andrew Prescott, “Beyond the Digital Humanities Center: The Administrative Landscapes of Digital Humanities,” in *The New Companion to Digital Humanities*, ed. Susan Schreibman, Ray Siemens, and John Unsworth (Chichester: Wiley Blackwell, 2016), 536. The early collaborative digital history projects undertaken by the Center for History and New Media and the American Social History Project used the language of film production to describe the roles of project members. See Stephen Robertson, “CHNM’s Histories: Collaboration in Digital History,” October 14, 2014 <http://drstephenrobertson.com/blog-post/chnms-histories-collaboration-in-digital-history>. Accessed 24 July 2016.

required for computational analysis will be restricted for 8 years by the exclusive commercial license granted to Ravel.²⁵

Given how little of the abundance of digitized material is accessible for computational analysis, and consequently, how important digitization continues to be to digital history, it is appropriate that several of the articles in this special issue discuss projects to make legal history sources available online. Surveying the changes being wrought by digitization in 2003, Roy Rosenzweig asked, “Should the work of collecting, organizing, editing, and preserving of primary sources receive the same kind of recognition and respect that it did in earlier days of the profession?”²⁶ As the profession has moved toward the recognition of digital history, that question has been answered in the negative. The American Historical Association, in its recent *Guidelines for the Evaluation of Digital Scholarship in History* defines digital history as “scholarship that is either produced using computational tools and methods or presented using digital technologies.” Other projects involving digital tools are bracketed off as service.²⁷ One way to bridge that gap is to publish peer-reviewed scholarship about those projects. This issue of *Law and History Review* is an important venue for such work.

Eiseman and Seipp offer accounts of two different approaches to delivering legal sources online. Comparing those two projects illustrates the different character of digitization undertaken by libraries and archives and by researchers, the consequence of cultural institutions that hold source materials having “their own ways of organising and describing source materials which may be quite different from the information produced by the research process.”²⁸ Those differences impact the extent to which the

25. Jennifer Dixon, “Harvard Launches “Free the law” Digitization Project,” *Library Journal*, December 12, 2015 <http://lj.libraryjournal.com/2015/12/oa/harvard-launches-free-the-law-digitization-project/#>. Accessed 24 July 2016. The agreement does allow Harvard Law Library to provide bulk access to researchers, if it so chooses. This model of partnerships involving periods of restricted access also characterizes arrangements between Ancestry.com and its related entities and the United States National Archives and various state archives.

26. Roy Rosenzweig, “Scarcity or Abundance? Preserving the Past in a Digital Era,” *American Historical Review* 108 (2003): 760.

27. American Historical Association, *Guidelines for the Evaluation of Digital Scholarship in History* <https://www.historians.org/teaching-and-learning/digital-history-resources/evaluation-of-digital-scholarship-in-history/guidelines-for-the-evaluation-of-digital-scholarship-in-history>. Accessed 24 July 2016. See also Thomas, “Promise of the Digital Humanities.”

28. Toby Burrows, “Sharing Humanities Data for E-Research: Conceptual and Technical Issues,” *Sustainable Data from Digital Research: Humanities Perspectives on Digital Scholarship. Proceedings of the Conference Held at the University of Melbourne, December 12–14, 2011* <https://ses.library.usyd.edu.au/handle/2123/7938>. Accessed 24 July 2016.

digitized sources can be explored using digital tools. As archivists and librarians, Eiseman and his colleagues began with the notebooks of students at the Litchfield Law School, not their contents, and with data about those legal sources, not the data in the sources. Their approach to making the notebooks accessible to researchers was to focus on the library catalog records, and then create a portal to “enhance access” to the notebooks. The portal includes valuable contextual information, but in its current preliminary form, the data and its format offer only a limited ability to explore the contents of the notebooks. The information about the notebooks is sortable rather than searchable. A user can click on “different legal titles such as Baron and Feme, Real Property, Powers of Chancery” associated with a section of a notebook to see all lectures on that subject across all the notebooks. That format limits a user to following a single path to exploring the collection, the one created by the scholar Whitney Bagnall. Moreover, as Eiseman notes, the subjects are not in a standardized form; therefore, using them to sort the contents of the notebooks does not necessarily gather all the related documents.²⁹ A transcription of the first line of each notebook section is included, but the records cannot be sorted based on that text, nor can that text be searched; therefore, those transcriptions cannot be readily used as a way of exploring the notebooks. Further stages of the project will address some of these limitations, with plans to standardize the subject categories used to describe the contents of the notebooks; however, the emphasis is on using catalogue metadata to expand access to the notebooks not to expand the ability to explore their contents.³⁰

By contrast, Seipp’s project is primarily concerned with the data in a source, rather than data about the source and the source itself. To more quickly and easily find information in a corpus of eleven large volumes of black-letter text and sixty-seven modern scholarly editions of reports, Seipp “compil[ed] a database of 22,318 records indexing and paraphrasing every printed Year Book report from the years 1268 to 1535 in England.” Databases require information to be organized into a tabular form following a set of rules—columns for different types of information, rows for each instance or record—and standardized. Many historical sources are difficult to convert into a database resource. They contain unstructured

29. Documents Collection Center, Yale Law School, Lillian Goldman Law Library. <http://documents.law.yale.edu/litchfield-notebooks/subjects>. Accessed 24 July 2016.

30. Eiseman does make brief mention of the “possibility” of crowdsourcing transcription of all the text; that is, building an online platform and recruiting volunteers to transcribe the documents using that platform. For examples of crowdsourced transcription projects, see Sharon Leon, “Build, Analyse and Generalise: Community Transcription of the *Papers of the War Department* and the Development of *Scripto*,” in *Crowdsourcing Our Cultural Heritage*, ed Mia Ridge (Farnham, UK: Ashgate, 2014).

information, and mix different types of information, producing ambiguous and inconsistent data. However, those problems occur less often with legal sources. Consistent forms such as statutes, court records, trials, and judicial opinions give structure to the information they contain. They also employ a repetitious and highly specialized vocabulary that gives legal texts many standardized elements. Those records were published frequently, and survive in comprehensive collections.

Seipp's database reflects these characteristics of legal sources. The Year Books have sufficient structural consistency to allow records that span more than 250 years to be included in a single database. The thirty-nine fields (columns) of data Seipp defined for the reports include fields for types of information produced by the structure of a report: "the name of the court, the writ, the names of the parties, if disclosed in the report, and other persons and places named." Also included are types of information that are features of legal proceedings: "the full names and abbreviated titles of all judges and lawyers quoted or mentioned every time they appear, statutes mentioned or hinted at... a field of keywords that lists every legal term in the report (in noun form), and a field that I call 'process' in which I include legal steps of pleading and procedure." A field somewhat misleadingly labeled "Commentary" contains "summaries and rough translations." Those "paraphrases," the information in the reports in an unstructured form, highlight that there is a range of information related to the facts of the cases, not the workings of the law, that is less easily organized, and which Seipp has chosen not to include in the database. The remaining fields are devoted to information in the traditional form of citation (akin to the catalog metadata created by Eiseman), contextual information and relationships drawn from sources other than the reports themselves.

In describing the database as an index, Seipp signals his focus on using searching to explore the Year Books. The interface offers a full text/keyword search (of all the information in the database), and a search limited to each field. Search results are delivered in a form that mirrors an index: as a list of individual reports, for which only the citation information is displayed, each linked to a single page containing all the fields of information relating to that report, and a link to an image of the document. The online database directs users to individual records, but offers no other ways to explore the reports. However, another way databases can be used to explore sources can be glimpsed in the total number of results that appears at the top of the list of search results.

A count of search results is an instance of how a database can group as well as retrieve information. Aggregating records allows them to be counted and patterns identified within the information. This simple data mining

is a quantitative approach, but does not rely on the statistical analysis that characterized the quantitative history of the 1960s and 1970s. As Fred Gibbs and Trevor Owens note, “The mere act of working with data does not obligate the historian to rely on abstract data analysis. Historical data might require little more than simple frequency counts, simple correlations, or reformatting to make it useful to the historian looking for anomalies, trends, or unusual but meaningful coincidences.” Such data mining is a technique focused on discovering and framing research questions, rather than generating evidence to confirm or refute a hypothesis.³¹

The Prosecution Project described by Mark Finnane involves creating a database that can be explored to discover patterns in Australian trials. Like Eiseman and Seipp, Finnane describes the creation of this digital resource, but he also offers the results of some preliminary explorations of the database to identify patterns in the information. The information in the database is drawn from registers of trials, a source even more structured than case reports such as those with which Seipp worked. In the example given by Finnane, Victoria, “the data by early twentieth century includes name, committal date and location, trial date and location, judge, prosecutor and defending counsel, names of witnesses (including their title if a police officer or medical expert), plea and outcome including sentence when convicted and appeal outcome when that applies.” Registers of cases also have the advantage of being available for all the Australian states, a comprehensive source that offers the possibility of creating a database of all of the country’s criminal trials. However, there is “sufficient variety in format that information across every category is not available for all 52,495 trials [currently in the database].” More broadly, the lack of narrative text that makes the registers relatively easy to convert into a database does also make them less rich in information than a case report.

Notwithstanding the limits on the types of information provided by the registers, Finnane’s preliminary data mining identified a range of patterns that provide directions for further research: variations in judicial sentencing, in granting bail, in legal representation for defendants, in which offenses involved co-defendants, and in the proportion of defendants who pled guilty. For example, defendants were “significantly more likely to

31. Fred Gibbs and Trevor Owens, “The Hermeneutics of Data and Historical Writing,” in *Writing History in the Digital Age*, ed. Kristen Nawrotzki and Jack Dougherty (Ann Arbor: University of Michigan Press, 2013), <http://quod.lib.umich.edu/d/dh/12230987.0001.001/1:7/-writing-history-in-the-digital-age?g=dculture;rgn=div1;view=fulltext;xc=1#7.3>. Accessed 24 July 2016. See also Jockers and Underwood, “Text-Mining the Humanities,” and Stephen Robertson, “Finding Questions As Well As Answers: Conceptualizing Digital Humanities Research,” May 2, 2016 <http://drstephenrobertson.com/blog-post/finding-questions/>. Accessed 24 July 2016.

be released on bail if charged with crimes against the person than crimes against property,” and “being released on bail was very strongly associated with a higher likelihood of acquittal or the abandonment of the case by the prosecution.” For Finnane

such evidence prompts of course further inquiry into the reasons for that association—how far were doubts about the strength of a case already in play at the committal stage, thereby shaping a bail decision? Or, as contemporary criminologists who have discerned similar trends have speculated, did pre-trial release better enable defendants to present themselves and their case in a more favorable light, while prolonged detention hampered their efforts to properly consult with legal counsel and perhaps encouraged them to simply plead guilty?

The *Digital Panopticon* illustrates another way that databases can be used to explore digital sources, by facilitating record linkage.³² The project aims to “trace the criminal and wider life histories of the 90,000 or so offenders sentenced at the Old Bailey to transportation to Australia or imprisonment within Britain between 1780 and 1925.” The evidence for those histories appears in more than forty different sets of judicial and civil records. Organized into databases, those documents can be linked by algorithms that use names to identify all the documents related to an individual. Ward and Williams do not discuss how that record linking is achieved; however, details of this central element of their digital method can be found on the project blog. Historical records are particularly challenging to link because of variations in spelling, the lack of unique identifiers, and imprecise dating. To increase the number of matches, the project has used algorithms that identify names that sound similar when spoken, but might be (accidentally) spelled differently, and that quantify and identify variance to match names with very small variance, in which only a single letter is different or omitted. Information from other sources is added to try to verify matches, and they are manually checked. An ongoing challenge of record linkage is trying to find the optimal, complementary balance of automated and manual work.³³

32. Finnane also mentions plans to enrich the *Prosecution Project* database by linking the data from registers to information from other sources, including “semi-automated linking” to the wealth of digitized newspapers available in *Trove*, <http://trove.nla.gov.au>.

33. “Adventures with Data Linkage,” *The Digital Panopticon*, <http://www.digitalpanopticon.org/?p=546>; “What’s in a Name?: Details and Data Linkage,” *The Digital Panopticon*, <http://www.digitalpanopticon.org/?p=707>; “Record Linkage Workshop Report, Part 2,” *The Digital Panopticon*, <http://www.digitalpanopticon.org/?p=717>; “James Littleton and the Problems of Automatic Record Linkage,” *The Digital Panopticon*, <http://www.digitalpanopticon.org/?p=1098>. Accessed 24 July 2016.

The “life archives” produced by the *Digital Panopticon*’s record linking can be mined for quantitative patterns. In that way, as Ward and Williams note, record linkage combines quantitative breadth with qualitative depth. It allows the recovery of individual lives alongside the patterns shared by thousands of lives. Ward and Williams offer a preliminary exploration of patterns in the current, incomplete data, using data mining in the same manner as Finnane. Only an outline is visible in their article of the complementary approach, being pursued by Tim Hitchcock, of using record linkage to orientate digital history toward the lives of ordinary individuals, and give greater shape to “history from below.”³⁴ Somewhat oxymoronically, big data is a powerful tool for producing small stories.

Textual data that has not been converted into a database, that remains unstructured in documents, can also be mined: this subfield of data mining is known as “text mining.”³⁵ Whereas patterns in structured data can be found by counting fields in a database, with unstructured text it is necessary to identify what will be counted. “A computer doesn’t know what a word is and certainly has no sense of what words might refer to,” as Stefan Sinclair and Geoffrey Rockwell note. “A computer “reads”—processes—text as a meaningless string of characters.” To demarcate words, computational tools look for spaces and punctuation, a process called “tokenization.”³⁶ In this issue, Tim Hitchcock and William Turkel use text mining to revise a narrative of court behavior being transformed in eighteenth and nineteenth century London by “the development of the ‘adversarial trial’, the changing role of legal counsel, the rise of ‘plea bargaining’ and summary justice, and the evolving functions of both judge and jury.”³⁷ They first explore what kind of evidence *The Proceedings of the Old Bailey Online, 1674–1913* can provide of court behavior, particularly the extent to which the changing nature of the *Proceedings* as a document reflects changes in court behavior rather than other forces. The existing understanding of this source is based on small samples and impressions; it is not possible for a researcher to look at all of the 127,000,000 words and 197,745 trials that make up the *Proceedings*. Text mining allows a comprehensive view: a computer can process every word, making it possible “to locate patterns made invisible by the sheer volume of inherited text.”

34. Tim Hitchcock, “Voices of Authority: Toward a History from Below in Patchwork,” *Historyonics*, April 27, 2015 <http://historyonics.blogspot.com/2015/04/voices-of-authority-towards-history.html>. Accessed 24 July 2016.

35. Jockers and Underwood, “Text-Mining the Humanities,” 351.

36. Stefan Sinclair and Geoffrey Rockwell, “Text Analysis and Visualization: Making Meaning Count,” in *The New Companion to Digital Humanities*, ed Susan Schreibman, Ray Siemens, and John Unsworth (Chichester: Wiley Blackwell, 2016), 339–40.

37. This work originated in *Data Mining with Criminal Intent*.

Hitchcock and Turkel use trial length—the number of words—in aggregate in each year, and in each trial, to explore the nature of the *Proceedings*. They found that in the eighteenth century, “the relationship between what was published and what occurred at the Old Bailey changed from decade to decade and from year to year,” which “makes their use as evidence for the rise of legal counsel and the adversarial trial difficult to sustain.” By contrast, text mining showed that the *Proceedings* give a fuller account of nineteenth century trials, contradicting the impressions of legal historians, which have led them to ignore those later records in favor of those from the eighteenth century. Text mining the nineteenth century *Proceedings* reveals “a mixture of longer and shorter trial reports between the early 1830s and 1850 with relatively few trials occupying,” and that trials that resulted in verdicts of not guilty in this period are “reported at much greater length than those resulting in a ‘guilty’ verdict.” Together with data mining that shows that guilty pleas and verdicts rose in the same period, Hitchcock and Turkel’s text mining confirms the growing importance of plea bargaining. Explaining the forces that brought about this change requires pairing the use of computational tools, with close reading and archival research.

David Tanenhaus and Eric Nystrom use another text mining technique, grouping documents together based on a measure of their similarity.³⁸ Algorithms measure similarity in different ways. Tanenhaus and Nystrom use the Jaccard coefficient, “which is the number of elements [words or phrases] the two documents have in common, divided by the number of elements found in both documents (with those appearing in both documents only counted once).”³⁹ Given the importance of questions about precedent and influence to legal history, and the specialized and repetitive vocabulary used in legal sources and settings, a computational tool that measures similarity has obvious value to legal historians. In this combination of counting and calculation, a computer is performing a form of close reading, and doing so with accuracy that a human would be hard pressed to match.

Tanenhaus and Nystrom’s article offers only a glimpse of how measuring similarity can be used for discovery, with text mining playing a minor role in their argument. Although the frame of the article discusses the use

38. Jockers and Underwood, “Text-Mining the Humanities,” 352. The other approach in digital humanities to analyzing word frequencies is visualization, particularly using *Voyant*, a tool developed by Stefan Sinclair and Geoffrey Rockwell that offers a variety of charts and graphs. See <http://voyant-tools.org> and Sinclair and Rockwell “Text Analysis and Visualization.”

39. Tanenhaus and Nystrom offer only a brief summary of their research in the article in this issue. The details of their use of computational tools can be found in Nystrom and Tanenhaus, “The Future of Digital Legal History,” at 161.

of digital tools, the body is a traditional, dense narrative argument. It is digitally inflected history rather than digital history, to borrow a distinction used in teaching. Computational methods clearly play a greater role in their larger project—as they described in another recently published article—and the conclusion of this article lays out how the authors plan to develop their digital tools. However, in the narrative, text mining is used only to confirm the choice of Arkansas as a representative case study, and to point to possible sources for the new language that appeared in state law in 1991. In neither case did the text mining reveal strong similarities. Four state laws passed after the 1997 Arkansas law shared significant elements with that law. Testing phrases did not reveal a source for the new sections for the 1991 law; therefore, Tanenhaus and Nystrom tested for the frequency of individual words, and weighted those common in one document but rare in the corpus more heavily, an approach effective only in giving a sense of common ideas, not whether language was borrowed. That text mining found similarity with many of the transfer laws passed during the mid-1990s, but especially in Virginia, which helped focus the analysis on the role of prosecutors.

Kellen Funk and Lincoln Mullen's work in progress on the transmission of the Field Code (New York's Code of Civil Procedure) also measures the similarity of documents, but not for discovery, as Tanenhaus and Nystrom did. Instead, Funk and Mullen seek to answer a specific question: how did the Field Code influence other American jurisdictions? That approach requires knowledge of the nature of a set of sources. Rather than mining all nineteenth century statutes of procedural law, Funk and Mullen worked with only potentially relevant laws: 135 statutes from the nineteenth century, amounting to 7,700,000 words organized into 98,000 regulations. Measuring the similarity of the sections of those laws reveals patterns in how law migrated at several different scales of analysis. An overview of the relationships among codes as a whole shows a network that features several different branches. Looking at borrowings in each code reveals a variety of different patterns in how many sections each code borrowed from another code. Finally, to find small changes in the wording and substance of the law, sections are grouped based on their similarity to one another, regardless of which code they come from, putting them in the context of their particular variations, not particular codes. This clustering shows, for example, that the Field Code expanded witness competency so that it excluded only the insane and very young children, but legislators in California grafted on older racial exclusions from Midwestern states, which were then adopted in the codes of many other Western states. Similar bars on testimony by nonwhites appeared in Iowa's code, which

relied exclusively on understanding of a legal oath to establish competency, a model reproduced a small number of other states.⁴⁰

Whereas Tanenhaus and Nystrom and Mullen and Funk use computational tools to measure the similarity of documents, Charles Romney measured the similarity of the context in which words appear as a means of discovering the similarity of concepts used in legal decisions.⁴¹ What this technique counts are the words that co-occur with the key word. He argues that using this computational tool is an approach analogous to the contextual close reading of Skinner's Cambridge School of intellectual history. Calculating similarity helped him identify both persistent concepts and moments of conceptual changes in the law. Specifically, he found in the Hawaii decisions a stable language of liberty across disparate fields of law and different periods of time, and the moment when the legal discourses about labor and habeas corpus crossed.⁴²

It is striking that, to date, legal historians have not used topic modeling, the computational text analysis tool mostly widely used in the digital humanities.⁴³ These algorithms produce possible topics by identifying clusters of words that appear in proximity to each other, which are in the same context. The algorithm divides the texts into as many topics as the user specifies to produce a model of probable topics; not a picture of the topics in a corpus. It is the "the task of the interpreter [researcher] to

40. Funk and Mullen, "A Servile Copy."

41. For an explanation of this form of vector space modeling, see Michael Gavin, "The Arithmetic of Concepts: a response to Peter de Bolla," *Modeling Literary History*, September 18, 2015 <http://modelingliteraryhistory.org/2015/09/18/the-arithmetic-of-concepts-a-response-to-peter-de-bolla/>. Accessed 24 July 2016.

42. A different approach to measuring similarity has been employed by Tim Hitchcock and his collaborators to explore the extent to which the treatment of crime in the *Proceedings of the Old Bailey Online* changed in line with the argument that a civilizing process that changed cultural norms produced a dramatic decline in violence. Like Funk and Mullen, they curated two corpuses, violent and nonviolent crimes, from the *Proceedings*, using the offense category tags. Rather than using word frequency as the basis for meaning, they coarse-grained the words in trials into named categories based on similarity of meaning using the nineteenth century Roget's *Thesaurus*. That process produced 1040 synonym sets, which nested inside 116 categories. They found "an increasingly clear distinction, within the record of spoken language, between trials associated with violent and nonviolent indictments." Sara Klingenstein, Tim Hitchcock, and Simon DeDeo, "The Civilizing Process in London's Old Bailey." *Proceedings of the National Academy of Sciences* 111 (2014): 9419–24, quote at 9419.

43. The most closely related work I could find is an analysis of popular constitutional discourse in United States newspapers in the years 1866–84. See Daniel Taylor Young, "How Do You Measure a Constitutional Moment? Using Algorithmic Topic Modeling to Evaluate Bruce Ackerman's Theory of Constitutional Change," *Yale Law Journal* 122 (2013): 1990–2054.

decide, through further investigation, whether a topic's meaning is overt, covert, or simply illusory."⁴⁴ A tool for topic modeling *The Proceedings of the Old Bailey Online, 1674–1913* does exist, but as yet no scholarship making use of it has appeared.⁴⁵ Tanenhaus and Nystrom rejected topic modeling because they were concerned that the results of their computational work be reproducible, and topic models "are not designed to give the same answers each time—that is, with the same inputs, and the same set of procedures, the outcomes can vary from one run to the next." However, the results of topic modeling are not answers, in the sense of providing evidence of the meaning of a set of sources. They are a place to begin, a pathway for discovery. Any answers they offer will come only when they are explored through close reading.

Given the current digital projects of legal historians, a more compelling explanation for topic modeling not having been used in digital legal history is that there is no need for it. The nature of legal sources—their consistent structures, comprehensiveness, and distinctive and repetitive language—means that tools based on word frequency are effective in discovering what is in sources, in revealing semantic information. Historical sources are more often characterized by a lack of structure and a variety of different types of information and language that causes text mining to be less revealing of their contents and meaning. It is historians exploring newspapers, magazines, and State Department memoranda and teleconference transcripts who have turned to topic modeling.⁴⁶

The nature of legal sources, together with the questions legal historians ask, likewise explains the relative lack of digital legal history projects using digital mapping tools. Drawn both by spatial questions, and by the lack of digitized historical sources, historians have "turned to digital mapping to a

44. Lauren Klein, Jacob Eisenstein, and Iris Sun, "Exploratory Thematic Analysis for Digitized Archival Collections," *Digital Scholarship in the Humanities* 30, Supplement 1 (2015): 131.

45. *Topic Explorer*, <http://inphodata.cogs.indiana.edu/oldbailey/40/?topic=6>. Accessed 24 July 2016.

46. Sharon Block, "Doing More with Digitization: An Introduction to Topic Modeling of Early American Sources," *Common-Place* 6 (2006) <http://www.common-place.org/vol-06/no-02/tales/>; Robert Nelson, *Mining the Dispatch* <http://dsl.richmond.edu/dispatch/pages/home/>; Micki Kaufmann, "Everything on Paper Will Be Used Against Me:" *Quantifying Kissinger* <http://blog.quantifyingkissinger.com/category/methods/topic-modeling/>; Andrew Torget and Jon Christensen, *Mapping Texts* <http://language.mappingtexts.org>; E. Thomas Ewing, Samah Gad, Bernice L. Hausman, Kathleen Kerr, Bruce Pencek, and Naren Ramakrishnan, *An Epidemiology of Information: Datamining the 1918 Flu Pandemic*, 2014 http://vtechworks.lib.vt.edu/bitstream/handle/10919/46991/An%20Epidemiology%20of%20Information%20Project%20Research%20Report_Final.pdf?sequence=1. Accessed 24 July 2016.

greater extent than other disciplines in the digital humanities, adopting it as their favored computational tool.”⁴⁷ Neither of the factors that draw historians in other fields to mapping have the same pull on legal historians. The relative wealth of digitized legal sources instead has led digital legal historians toward text mining, and the character of those sources has provoked questions about the law and legal practice. *Locating London's Past* extends the *Proceedings of the Old Bailey Online, 1674–1913* to a mapping project that places the crimes tried in Old Bailey in the city of London. However, that site has not been as generative of scholarship as have the online *Proceedings*. One prominent digital history mapping project, *Digital Harlem*, which I created with colleagues at the University of Sydney, does map crime, using prosecutors' records of felonies, but it is not a legal history. The focus of *Digital Harlem* is everyday life, and our concern has been more with what legal records tell us about the neighborhood than what they reveal about the law and legal process.⁴⁸ For example, our book about numbers gambling—*Playing the Numbers: Gambling in Harlem Between the Wars*—is a cultural history, exploring how the game permeated all aspects of life. The locations in which bets were placed, winnings collected, and games organized are only a small part of that story, as are the arrests that revealed that information, and the hearings and prosecutions that resulted.⁴⁹

Ng's article points to spatial dimensions to legal practice and the legal process that could be mapped. At this early stage of his project, Ng has mapped only the locations of lawyers' offices. The future directions for the project that he mentions are related to policing; however, there are other legal spaces and places that could be mapped, such as courts, prisons, probation and parole departments, and medical and psychiatric clinics. Mapping these places would highlight how the movement of individuals through the legal process was spatial as well as bureaucratic. *Digital*

47. Robertson, “Differences.”

48. Stephen Robertson, “Putting Harlem on the Map,” in *Writing History in the Digital Age*, ed Kristen Nawrotzki and Jack Dougherty (Ann Arbor: University of Michigan Press, 2013), <http://quod.lib.umich.edu/d/dh/12230987.0001.001/1:8/--writing-history-in-the-digital-age?g=dculture;rgn=div1;view=fulltext;xc=1#8.2> (accessed 24 July 2016); and Stephen Robertson, “Digital Mapping as a Research Tool: *Digital Harlem: Everyday Life, 1915–1930*,” *American Historical Review* 121 (2016): 156–66.

49. Shane White, Stephen Garton, Stephen Robertson, and Graham White, *Playing the Numbers: Gambling in Harlem Between the Wars* (Cambridge, MA: Harvard University Press, 2010). See also Stephen Robertson, “Arrests for Numbers Gambling,” *Digital Harlem Blog*, April 17, 2009 <https://digitalharlemblog.wordpress.com/2009/04/17/numbers/>; Stephen Robertson, “Numbers on Harlem's Streets,” *Digital Harlem Blog*, December 1, 2011 <https://digitalharlemblog.wordpress.com/2011/12/01/numbers-on-harlems-streets/>. Accessed 24 July 2016.

Harlem, for example, shows that the two magistrate's courts that processed cases from the neighborhood were located beyond the boundaries of black settlement, to the south and north, in overwhelmingly white neighborhoods. Tim Hitchcock's current work creating a three-dimensional (3D) model of the Old Bailey courtroom is effectively a spatial history of the interior of the places of the legal process, one that places the words recorded in the *Proceedings* in the spatial context in which they were spoken. The model highlights the changing layout of the court, and the relative position of the different speakers, and allows for an exploration of how the different actors in a legal proceeding heard themselves and each other.⁵⁰

Mapping the legal process does not require the tools and approaches of geographic information systems (GIS) that Ng uses. Much of the concern with geometric and mathematical precision and technical challenges that occupies Ng in this article is related to working with GIS, which is ultimately unnecessary given the kind of point mapping that he is undertaking. His account reflects the constraints that limited historical mapping before the Internet more than the possibilities opened up by web mapping. Ng is trying to bring "the quantitative methods stemming from geographic information systems (GIS) . . . to bear on qualitative, historical and interpretive methods from the humanities." Mapping in digital humanities involves something more, as Todd Presner and David Shepherd argue; a "reconceptualization of the significance of place in relationship to narrative, practices of representation, and digital technologies."⁵¹ It has been catalyzed by the availability of alternatives to GIS that are easy to use, and by web map mashups, such as *Digital Harlem*, projects built on top of platforms such as Google Maps. A map that effectively shows the patterns that Ng discusses could be created with a fraction of the effort it took him using a free web-mapping platform such as CartoDB, Neatline, Google Maps or Google Earth. Such a map would be dynamic and interactive, with users able to modify what they saw on a map using options such as filters, time sliders, panning, and zooming. It would be a research tool; a means of discovering as well as displaying knowledge.⁵²

50. Hitchcock, "Voices of Authority;" Tim Hitchcock, "Re-imagining the Voice of the Defendant at the Old Bailey," *The History of Crime and the Courts in Three Dimensions: a Half-Day Workshop*, October 20, 2015 www.hrionline.ac.uk/san/wp_digitalpanopticon/hitchcockVoiceDefendant.pdf. Accessed 24 July 2016. An example of a project that reconstructs a historical soundscape in this way is the *Virtual Paul's Cross Project*, a digital re-creation of John Donne's Gunpowder Day Sermon in 1622 (<https://vpcp.chass.ncsu.edu>).

51. Todd Presner and David Shepard, "Mapping the Geospatial Turn," in *The New Companion to Digital Humanities*, ed Susan Schreibman, Ray Siemens, and John Unsworth (Chichester: Wiley Blackwell, 2016), 247.

52. Robertson, "Finding Questions;" Robertson, "Differences;" and Presner and Shepard, "Mapping the Geospatial Turn," 247, 251.

One way to think about digital mapping is as a form of data visualization “that uses levels of abstraction, scale, coordinate systems, perspective, symbology, and other forms of representation to convey a set of relations.”⁵³ Other forms of visualization are deployed in the articles in this issue, most notably the graphs that Hitchcock and Turkel produce to “allow all the available data to be viewed at a single glance, and to facilitate an open-eyed engagement with the patterns revealed.” I could identify only one legal history project centered on visualizations: William Thomas’s *O Say Can You See? Early Washington, D.C., Law and Family*. Dynamic network graphs offer a way to “explore the web of litigants, jurists, attorneys, and community members present in the court records,” and “legal, occupational, family, and social connections to each other.”⁵⁴ However, as Fred Gibbs recently noted, “as the volume of digitized historical data grows, the visualizations that help make sense of data at large scales will play an increasingly significant role in our analyses and interpretations of the historical record. They have a new element of necessity.”⁵⁵ Network graphs have a particular potential for legal history, offering a visualization of relationships that can be applied to questions of precedent and influence.⁵⁶

Although data visualization has only a small footprint to date in digital legal history, it is poised to become a prominent part of legal research. Ravel Law is undertaking the digitization of Harvard Law Library’s collection of American case law discussed earlier to obtain data for a visualization product that creates visual maps of search results and networks of cases, as well as data visualizations of judges’ careers, showing all their decisions and citations, and the specific language that they use.⁵⁷ That product is only one of many launched recently to visualize legal research.⁵⁸ Individual legal researchers are also developing visualization tools. Colin Starger, for example, has developed software to create interactive visual citation networks for the United States Supreme Court.⁵⁹

53. Presner and Shepard, “Mapping the Geospatial Turn,” 247.

54. *O Say Can You See* <http://earlywashingtondc.org/>. Accessed 24 July 2016.

55. Fred Gibbs, “New Forms of History: Critiquing Data and Its Representations,” *The American Historian* 7 (2016), <http://tah.oah.org/february-2016/new-forms-of-history-critiquing-data-and-its-representations/>.

56. See Funk and Mullen, “A Servile Copy” for a network graph showing code to code borrowings.

57. *Ravel, Data Driven Research* <https://www.ravellaw.com>

58. Robert Ambrogi, “Visual Law Services Are Worth a Thousand Words—and Big Money,” *ABA Journal* (2014) http://www.abajournal.com/magazine/article/visual_law_services_are_worth_a_thousand_words--and_big_money/. See also CODEX, the Stanford Center for Legal Informatics, <https://law.stanford.edu/codex-the-stanford-center-for-legal-informatics/>. Accessed 24 July 2016.

59. *Court Listener, Supreme Court Citation Networks* <https://www.courtlistener.com/visualizations/scotus-mapper/>. Accessed 24 July 2016.

These new kinds of data visualizations are helping renew interest in the possibilities for publishing history in digital formats. Print forms of scholarly communication struggle to accommodate visualizations. The problem exceeds the cost of reproducing images, especially the color images needed to capture visualizations, in scholarly journals. It is also that many of those visualizations are dynamic, and readers need to be able to interact with them to explore and assess their value as evidence. Currently, the solution is to have authors create and host online supplements.⁶⁰ Taken together with researchers' increasing reliance on online sources—which remains somewhat occluded by the widely shared practice of citing print editions of sources even when an online edition was used—and on accessing secondary sources online, the growing use of visualization means that print forms are not able to accommodate the core elements of much historical scholarship. The limits of what can be done in print are coupled with new recognition of what is possible in digital forms. Cheil van den Akker has argued that, “the digital environment supports, indeed demands, new narrative forms that are more participatory, dialogic, procedural, reciprocal, and spatial.”⁶¹ Thanks to a funding initiative launched by the Andrew W. Mellon Foundation, several university presses are developing digital platforms for scholarly publication. These projects range from a system and framework for publishing digital-born scholarship, and a platform for multimedia online journals, to iterative, networked, electronic versions of scholarly monographs to appear alongside the print edition of the book, and platforms for digitally enhanced monographs, and open-access monographs.⁶² However, for the moment, the nearest most historical scholarship

60. See Cameron Blevins, “Space, Nation, and the Triumph of Region: A View of the World from Houston,” *Journal of American History* 101 (2014): 122–47; and Cameron Blevins, “Mining and Mapping the Production of Space: A View of the World from Houston,” 2014 <http://web.stanford.edu/group/spatialhistory/cgi-bin/site/pub.php?id=93>. Accessed 24 July 2016.

61. Cited in Thomas, “Promise of the Digital Humanities,” 606–7.

62. Stanford University Press is developing the system and framework for publishing digital-born scholarship (<http://library.stanford.edu/news/2015/01/stanford-university-press-awarded-12-million-publishing-interactive-scholarly-works>). West Virginia University is developing Cairn, an online, free, and open-source system that will help editors of scholarly multimedia journals, books, and data sets engage in building and reading multimedia-rich, peer-reviewed content (<http://wvutoday.wvu.edu/n/2015/02/03/wvu-receives-1-million-grant-from-mellon-foundation-for-first-of-its-kind-digital-publishing-system>). The University of Minnesota Press and the GC Digital Scholarship Lab at the Graduate Center of the City University of New York (CUNY) are developing Manifold Scholarship (<http://manifold.umn.edu>), a platform for an alternative iterative, networked, electronic versions of scholarly monographs alongside the print edition of the book (<http://www.upress.umn.edu/press/press-releases/manifold-scholarship/>). The presses at Indiana, Michigan, Minnesota, Northwestern, and Penn State are developing a new platform using Hydra/Fedora that will

gets to a digital format is the PDF file in which print articles are delivered online.

Searching for digital legal history reveals that what currently distinguishes digital legal history as a field of digital history is its focus on text mining. Legal sources skew the field toward that approach, which offers a way of exploring questions about the re-use of text and concepts that characterizes key facets of legal process and legal publication. A wealth of opportunities exists to expand the use of this approach in analyzing the bills, statutes, and judicial opinions, the sources discussed in this issue, and beyond, to trials and treatises. To do so, legal historians must join with academic librarians in pushing commercial vendors to provide bulk access to their products, and to provide more information about the quality of their contents. It will also be necessary for legal historians to create and curate digital collections of their own, and here again, they will need to work with librarians, to ensure that their projects meet data standards. They also need to work with libraries and other institutions undertaking digitization to bridge the gap between digital collections and research data. A team at the Roy Rosenzweig Center for History and New Media that Sean Takats and I are leading is developing software that will help connect the digitization work of researchers and institutions. Tropy provides researchers with a tool to manage the digital photographs that they take in archives, and facilitates sharing that material with the archives that hold those sources.⁶³

Other forms of digital history have yet to engage the attention of legal historians, although that may be about to change. Through the multifaceted *Digital Panopticon* project, the team that played a pivotal role in introducing text mining to legal history through its creation of *The Proceedings of the Old Bailey Online, 1674–1913* is poised to play a similar role with regard to data visualization, mapping, and 3D modeling. Developments in legal research are likewise putting visualization tools in front of legal historians. The results will likely echo what happened when LexisNexis and Westlaw introduced computerized databases. Legal historians will

enable the publication and preservation of digitally enriched humanities monographs (<http://www.publishing.umich.edu/2015/04/01/mellon-grant-funds-u-m-press-collaboration-on-digital-scholarship/>). The University of California Press and the California Digital Library are developing a system to support the publication of open-access monographs (<http://www.cdlib.org/cdinfo/2015/03/05/uc-press-and-the-california-digital-library-receive-750k-grant-from-the-andrew-w-mellon-foundation/>).

63. Stephen Robertson, "Tropy – Digital Image Management for the Humanities Research Community," October 8, 2015 <http://drstephenrobertson.com/news/tropy-a-tool-to-organize-describe-share-digital-images-taken-in-research/>. The development of this software can be followed at <http://tropy.org> and @tropychmm.

have access to computational tools ahead of scholars in other fields. As they start using those tools, they need to avoid the mistake we made when we started searching databases as part of our research. Legal historians need to explore how new tools are transforming their practice. And while they are doing that, it is important to start discussing how we as historians have been using searching to do legal history, and bring our current digital research practices and their consequences into our scholarship.