Journal of Cultural Analytics

Kotin, Joshua, and Rebecca Sutton Koeser. 2026. *Shakespeare and Company Project* Data Sets, Version 2.0. *Journal of Cultural Analytics*, 11(1): pp. 1–10. DOI: https://doi.org/10.16995/jca.1114

# *Shakespeare and Company Project* Data Sets, Version 2.0

**Joshua Kotin,** Department of English, Princeton University, jkotin@princeton.edu
**Rebecca Sutton Koeser,** Center for Digital Humanities, Princeton University

The *Shakespeare and Company Project* data sets provide a detailed portrait of Shakespeare and Company, Sylvia Beach's bookshop and lending library in interwar Paris. This article outlines the research, data curation, and development that led to a major update of the data sets. Version 2.0 augments and refines the data sets included in Version 1.2 and adds two new data sets focused on the authors of the books that circulated in the lending library and the addresses of lending library members. The article should be read as a supplement to "*Shakespeare and Company Project* Data Sets," published in the *Journal of Cultural Analytics* in 2022.

## 1. Overview

"In the future, the *Project* team plans to release address and creator exports to facilitate further research." We made this statement—half prediction, half promise—in the final paragraph of our article, "*Shakespeare and Company Project* Data Sets," which appeared in the *Journal of Cultural Analytics* in 2022. Four years later, we have realized those plans, releasing a new version of the data sets, which updates the three original data sets and adds two new data sets: one focusing on the creators of the books that circulated in the lending library and the other on the addresses of lending library members. The release, *Shakespeare and Company Project* Datasets, version 2.0, has a single DOI and replaces version 1.2.[1] The release also includes a readme outlining the update and change log files for the revised data sets.[2]

In the supplement to our earlier article, we describe the new release. We begin by discussing the changes to the three original data sets, then the two new data sets. Together, the five data sets present the most complete and accurate portrait yet of the Shakespeare and Company lending library, which operated from 1919 until the death of its owner, Sylvia Beach, in 1962. The data sets are a unique resource, offering insights into the development of modernism, the history and sociology of reading, and the everyday lives of thousands of individual readers. There is no comparable source for understanding the period.

For an overview of the historical importance of Shakespeare and Company, we recommend our earlier article and the articles in *The World of Shakespeare and Company*, a special issue co-published by the *Journal of Cultural Analytics* and *Modernism/modernity* in 2024. The articles also provide a detailed account of the data sets and their research value, including their connection to archival documents in the Sylvia Beach Papers, 1872–1999 at Princeton University Library. One article from the special issue is especially relevant for understanding the data sets: Rebecca Sutton Koeser and Zoe LeBlanc's "Missing Data, Speculative Reading," which describes missing data in the Sylvia Beach Papers and its significance.

---

[1] When we published the original three data sets, we gave each one a distinct DOI and created an aggregate record with a fourth DOI. We anticipated publishing individual updates and wanted to permit researchers to work with and cite individual data sets. In practice, however, we have always published updated versions of the data sets together, and the multiple DOIs have caused confusion. For simplicity and clarity, we now treat the data sets as a single record with a single DOI.

[2] As in previous versions, the data sets are available in two formats: CSV (comma-separated values) and JSON (JavaScript Object Notation). We offer both formats for the convenience of researchers and developers. The CSV files are easy to open and use in a spreadsheet program such as Microsoft Excel and Google Sheets, but they do not work as well for multi-valued fields and do not preserve data type information. For interacting with the data via code, JSON is more efficient. We encourage researchers and developers to use the provided Frictionless Data Package file, which documents the structure and data types of the fields in each file.

## 2. Members, Books, and Events

Version 2.0 of the *Shakespeare and Company Project* data sets update the three original data sets, which focus, respectively, on the members of the Shakespeare and Company lending library; the books that circulated in the lending library; and the events—borrows, purchases, subscriptions, renewals, deposits, and reimbursements—that connected members and books.

The most significant update to the original data sets concerns the quantity and quality of demographic information about individual lending library members. Version 1.2 provides demographic information for about 600 members. In version 2.0, that number approaches 1,800. Since the *Project*'s inception, the *Project* team has used information in the Sylvia Beach Papers (e.g., names and membership dates, and, when available, titles and addresses) to identify actual people. The members data set includes fields for gender, nationalities, birth and death years, and links to reference works and authority files.[3] Version 2.0 includes robust demographic information for over one third of the lending library's total membership.

Some identifications were easy to make. For example, lending library member "Ernest Hemingway" is clearly the famous modernist writer, Ernest Hemingway (1899–1961). To identify other members, a simple internet search was often sufficient. In a matter of minutes, team members identified "Madame Jean Prévost," who borrowed one book in 1931—Virginia Woolf's *A Room of One's Own* (1929)—as the French writer, Marcelle Auclair (1899–1983). But the majority of identifications required significant research. Using various sources—census records, genealogical databases, passport applications, transatlantic passenger lists, and digital libraries such as Gallica from the Bibliothèque nationale de France—team members identified "Madame Amor" as Hélène (de Yturbe) Amor (1881–1958), a refugee from the Mexican Revolution. Team members then composed a brief biography, connecting Amor to her granddaughter, the Mexican writer, Elena Poniatowska (1932–present). The notes field in the data set includes this biographical information. Much of this research was spearheaded by Yvonne Patch, a retired librarian in Hamilton, Ontario, and the mother of Project director, Joshua Kotin.

The updated data set illuminates hundreds of such biographies. But just as important, the data set allows for much more extensive analysis of lending library demographics (**Figures 1** and **2**). Links to reference works and authority files—Wikipedia and Virtual International Authority File (VIAF)—provide access to additional demographic and biographical information. The quantity of links reveals the prevalence of notable figures

[3] For a discussion of how the *Project* represents gender, see Budak. When possible, the members data set links to Virtual International Authority File (VIAF) and Wikipedia.

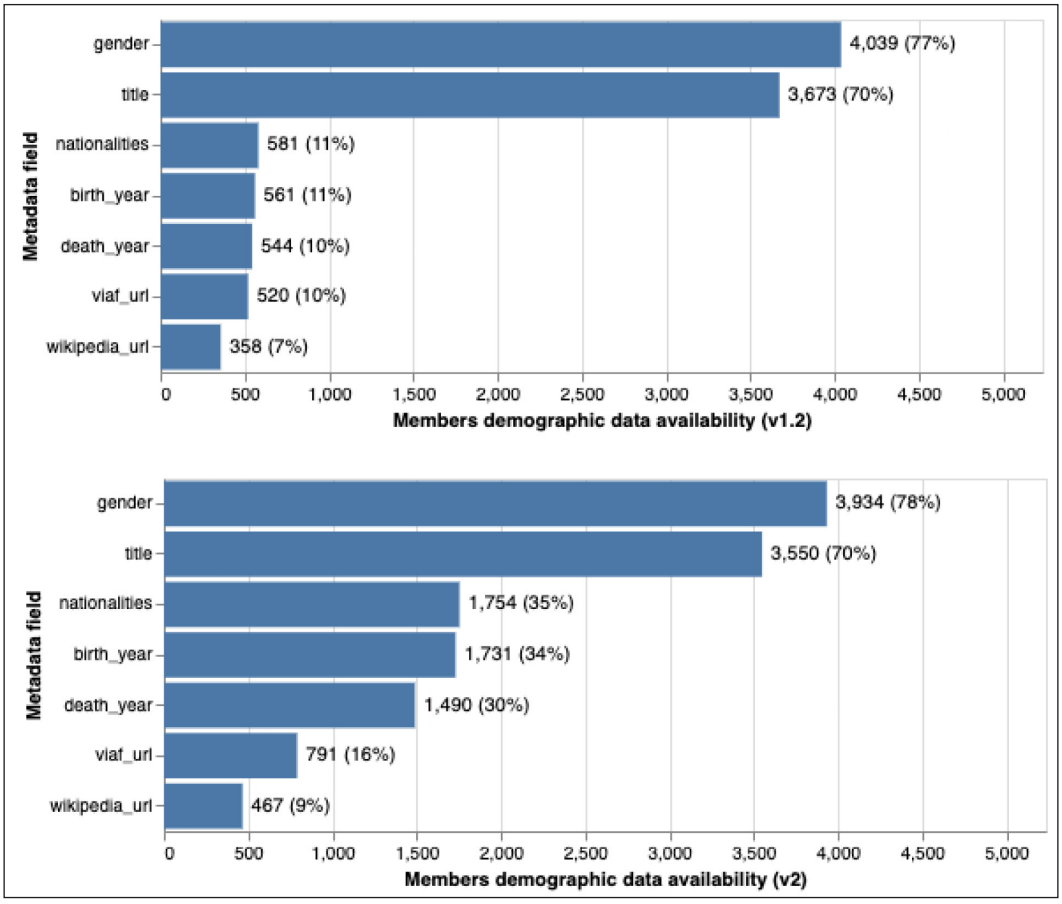


**Figure 1:** This pair of graphs shows the changes in available demographic data for lending library members in the 2.0 data sets as compared to the 1.2 version, based on raw numbers of members.

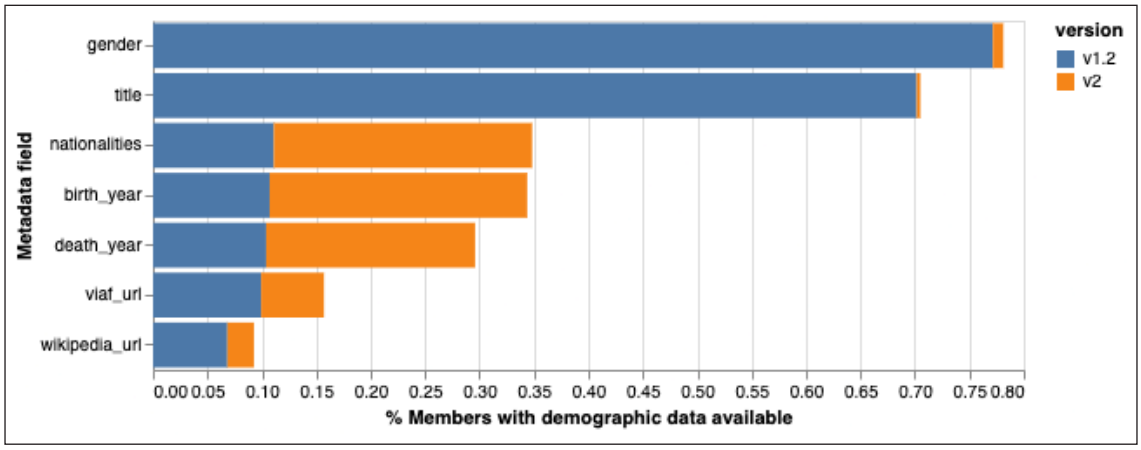


**Figure 2:** This graph shows changes in available demographic data for lending library members in the 1.2 and 2.0 data sets, as a percentage of total members in each version of the dataset.

at the Shakespeare and Company lending library. About one-sixth of its members were published writers.

The Project team also uses the notes field to suggest likely identifications. For example, "Mademoiselle M. Borsoupzky," who was a member in 1934, is almost certainly the French-Romanian filmmaker Myriam Borsoutsky, but the team cannot definitively place Borsoutsky at Shakespeare and Company that year. Instead of introducing possible errors into the data set, the team used the notes field to suggest the identification. Version 2.0 includes likely identifications for 146 members, up from forty-six in version 1.2. (Between the two versions, the team confirmed some likely identifications and added many others.) Researchers can use this information to conduct their own research.

The books and events data sets have also been updated. Minor errors have been corrected and additional information provided. Most significantly, the books data set now includes a "genre" field, which identifies whether a book is fiction, nonfiction, poetry, drama, a periodical, or a combination thereof (**Figure 3**). The information should help analyze the reading practices and preferences of lending library members. Version 2.0 also includes twenty-eight additional events based on information in a notebook from the Harry Ransom Center at the University of Texas. An important reminder to researchers: The book purchase information in the events data set *only* includes purchases recorded on the lending library cards. Beach kept separate records for the bookshop. Accordingly, the data sets cannot be used for a comprehensive account of book sales at Shakespeare and Company.

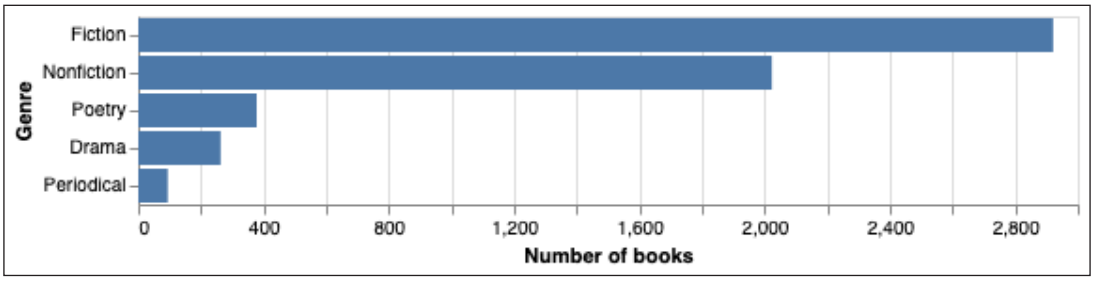


**Figure 3:** This graph shows the number of books that circulated in the lending library by "genre."

## 3. Book Creators and Addresses

Version 2.0 of the data sets also includes two new data sets. "Book creators" focuses on the authors and other creators of books that circulated in the Shakespeare and Company lending library. The term "creators" encompasses many roles, such as editors, illustrators, translators, and so forth. (Researchers can use the book creators and books data sets together to analyze specific roles.) The data set includes 2,530 creator names with fields for gender, birth and death years, nationalities, VIAF URL, and member URI (in case the creator was a member of the lending library: 113 creators were also members). Like the

members dataset, an additional field, "is organization," distinguishes individual and collective creators. There is only one of the latter, which is the Fabian Society.

The book creators data set was developed by Fedor Karmanov and Joshua Kotin for their article "A Counterfactual Canon," published in the *World of Shakespeare and Company*, the special issue of *JCA*. The article analyzes connections between gender and reading preferences at Shakespeare and Company, and includes an analysis of the gender distribution of the authors of the books held by the lending library. Almost 80 percent of the books known to circulate in the library were written by men. In contrast, over 62 percent of lending library members were women. Perhaps unsurprisingly, women were almost twice as likely as men to borrow books by female authors.

Birth years in the creators data set confirm that Shakespeare and Company was predominately a library of contemporary literature (**Figure 4**).[4] The average creator birth year is 1849, and the mode is 1889. For members, the average birth year is 1894 and the mode is 1897. The Chinese philosopher Laozi has the earliest birth year, 571 BCE, seventy-five years before Sophocles. The English playwright John Osborne has the latest, 1929. Beach loaned Osborne's *Look Back in Anger* (1957) to Hélène Baltrusaitis (1908–2004), daughter of the French art historian Henri Focillon (1881–1943), in 1961, twenty years after the bookshop and lending library officially closed. (Beach continued to loan books from her apartment until her death in 1962.)[5] Osborne's play marked a breakthrough for a new generation of English writers, working-class and left-wing, who were developing new forms of social realism—a departure from the literature most often associated with Shakespeare and Company.

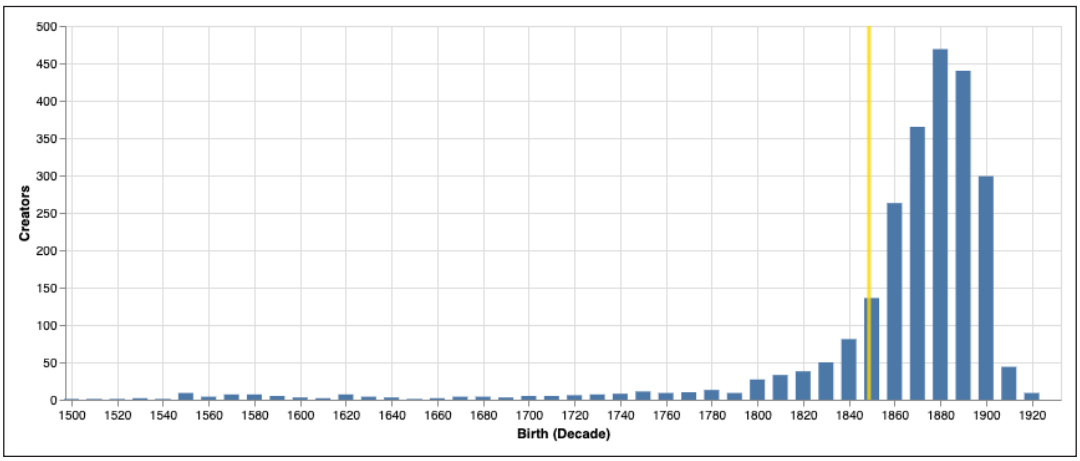


**Figure 4:** This graph shows the age distribution of creators in the book creators data set, with average age marked in yellow (and omits thirty-four creators born before 1500).

[4] For a discussion of Shakespeare and Company as a library of contemporary literature, see Antoniak et al.

[5] For a discussion of the last book that Sylvia Beach loaned at Shakespeare and Company, see Walsh.

The book creators data set mirrors the members data set. But there is one important exception. In the members data set, "United Kingdom" encompasses the nationalities of members from England, Scotland, and Wales. In the book creators data set, the team distinguishes among these nationalities. There are sixty-nine Scottish writers, including Edwin and Willa Muir, the English-language translators of Franz Kafka; and sixteen Welsh writers, including Dylan Thomas. Nevertheless, the most represented nationality is still English, with authors from the United States a distant second (**Figure 5**). The data set also reveals that the lending library held books by 129 French authors from Abelard and Héloise to Marguerite Duras (1914–1996). The circulation of these books reveals a segment of the Shakespeare and Company membership who likely couldn't read French. As in the members data set, nationalities is a multi-value field. Seventy-nine creators have multiple nationalities. For example, Joseph Conrad is Polish and English; Claude McKay is Jamaican and American; and Vladimir Nabokov is Russian and American.

The second new data set, "members addresses," includes information about the addresses of lending library members. Much of this information is available in the members data set but is difficult to analyze. Members often have multiple addresses, each comprising multiple fields: street address, postal codes, arrondissements, coordinates. In the new members addresses data set, each address is a top-level entity

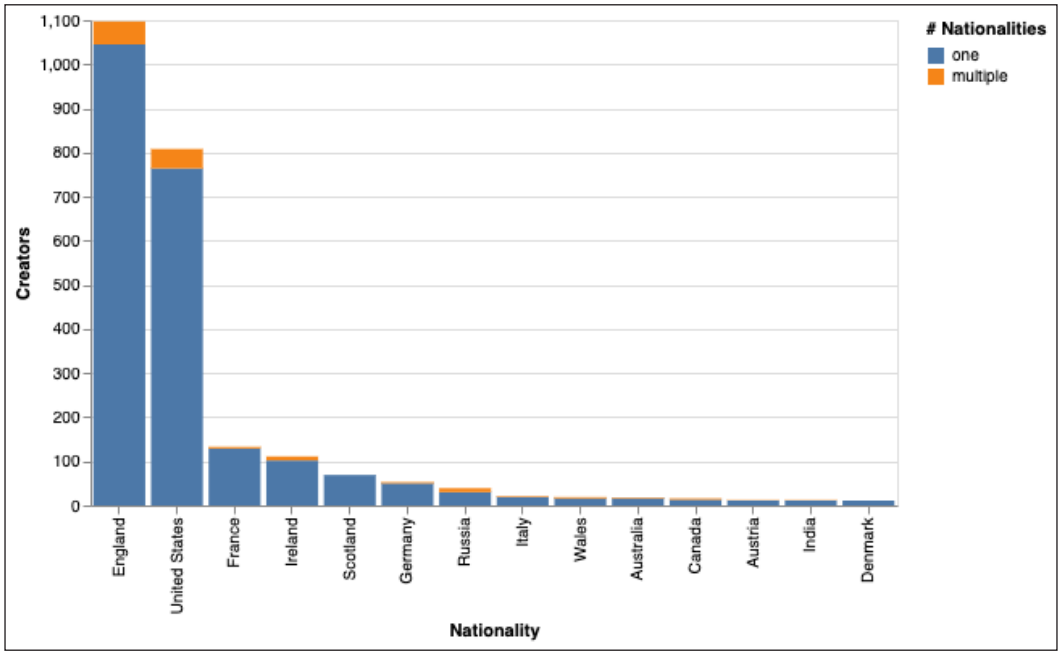


**Figure 5:** This graph shows the nationalities of book creators (limited to nationalities that occur at least ten times).

represented by a single row in the tabular data with a member URI so the data can be used in combination with the members data set. Members with multiple addresses have multiple entries. The data set should make geographical analysis easier. Unsurprisingly, members concentrate on the Left Bank, especially in the sixth arrondissement (**Figure 6**). But as Jesse McCarthy and Moacir P. de Sá Pereira show in "The Literary Right Bank" (2021)", a significant number of members lived on the Right Bank, especially in the wealthy sixteenth arrondissement.

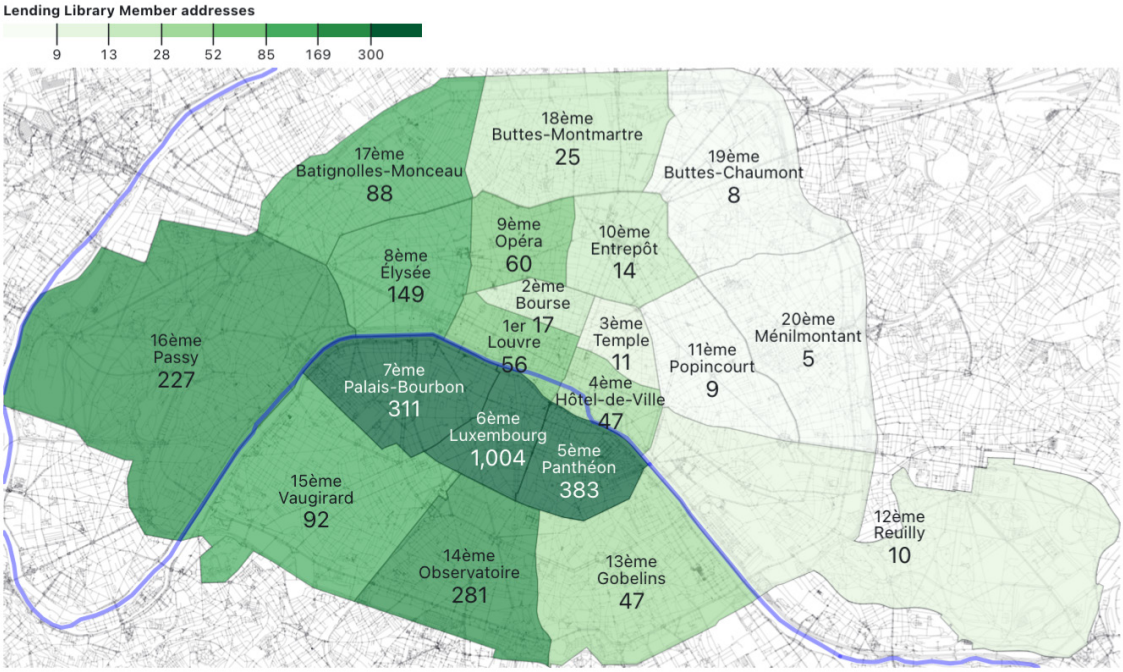


**Figure 6:** This map of Paris shows members by arrondissement and the predominance of members on the Left Bank.

## 4. Future Plans

We are hesitant to make additional promises or predictions about the future of the *Shakespeare and Company Project* data sets. The *Project* team intends to release annual updates. Version 2.1 will add to the demographic information in the members dataset but will not include any structural changes. Since the release of version 2.0, the *Project* team has already matched an additional ninety-seven names in the Sylvia Beach Papers to biographies of real people. These identifications are live on the *Project* website.

The Sylvia Beach Papers still have much to teach us about modernism, book history, and social networks. Small- to medium-sized improvements to the current data sets would open new avenues of research. The members addresses data set, for example,

could be improved by adding occupancy dates for members with multiple addresses. The information would help reveal connections between geographical proximity and taste. An entirely new project could focus on book sales at Shakespeare and Company. Beach recorded every sale in logbooks. (She didn't, however, record purchaser names.) A separate book sales data set, combined with the events data set, would offer a near-complete account of Shakespeare and Company's revenue. An expenses data set would then become the final puzzle piece needed for a full accounting of Beach's three businesses: lending library, bookshop, and publishing house.

At present, the *Shakespeare and Company Project* has shifted from developing the data sets to using them. In 2024, Kotin founded the Shakespeare and Company Project Lab at Princeton, and in 2025, the Lab completed its first article, an account of the circulation of James Joyce's *Ulysses* in Paris. It is our hope that the *Shakespeare and Company Project* data sets will continue to inspire new research and research collaborations.

## Competing Interests

The authors have no competing interests to declare.

## Works Cited

Antoniak, Maria, David Mimno, Rosamond Thalken, Melanie Walsh, Matthew Wilkens, and Gregory Yauney. "The Afterlives of Shakespeare and Company in Online Social Readership." *Journal of Cultural Analytics*, vol. 9, no. 2, 2024, https://doi.org/10.22148/001c.116919.

Budak, Nick. "Representing Gender in the Shakespeare and Company Project." *Shakespeare and Company Project*, December 12, 2019, https://shakespeareandco.princeton.edu/analysis/2019/12/representing-gender-in-the-shakespeare-and-company-project/.

Karmanov, Fedor, and Joshua Kotin. "A Counterfactual Canon." *Journal of Cultural Analytics*, vol. 9, no. 2, 2024, https://doi.org/10.22148/001c.116915.

Koeser, Rebecca Sutton, and Joshua Kotin. "Shakespeare and Company Project Datasets." Version 2.0, Princeton University, 24 Feb. 2025, https://doi.org/10.34770/kf6c-b079.

Koeser, Rebecca Sutton, and Zoe LeBlanc. "Missing Data, Speculative Reading." *Journal of Cultural Analytics*, vol. 9, no. 2, 2024, https://doi.org/10.22148/001c.116926.

Kotin, Joshua, and Rebecca Sutton Koeser. "*Shakespeare and Company Project* Data Sets." *Journal of Cultural Analytics*, vol. 7, no. 1, 2022, https://doi.org/10.22148/001c.32551.

———. "The World of Shakespeare and Company: An Introduction." *Journal of Cultural Analytics*, vol. 9, no. 2, 2024, https://doi.org/10.22148/001c.116905.

McCarthy, Jesse, and Moacir P. de Sá Pereira. "The Literary Right Bank." *Shakespeare and Company Project*, April 5, 2021, https://shakespeareandco.princeton.edu/analysis/2021/04/literary-right-bank/.

"Sylvia Beach Papers, 1872–1999." Manuscripts Division, Department of Special Collections, Princeton University Library, http://arks.princeton.edu/ark:/88435/7h149p866.

Walsh, Keri. "Sylvia Beach's Final Book." *Journal of Cultural Analytics*, vol. 9, no. 2, 2024, https://doi.org/10.22148/001c.116913.