Moving the Capital of US Literature from Boston to New York

Evidence from 11 million Library of Congress records

Erik Fredner

2024-11-20

1 Introduction

Literary historians of the United States broadly agree that, at some point during the nineteenth century, New York City overtakes Boston as the literary capital of the US. What Pascale Casanova argues Paris is for world literature, New York becomes for US literature specifically, and, eventually, Anglophone literatures more generally.[1]

Data recently released by the Library of Congress (LC) allows us to assess this historical claim in new ways, but not in the format in which it was originally released. By converting this data to a form useful for computational literary studies, I have evaluated this geographic shift, and made it possible for other scholars of nineteenth-century US literatures to use the same data for a wide range of purposes. Perhaps more importantly, the code that extracted the data used to answer this question about nineteenth-century places of publication can be used to get any fields of interest from any set of the 11 million book catalog records that the LC has made public.

[1] Pascale Casanova, *The World Republic of Letters*, Convergences (Cambridge, Mass: Harvard University Press, 2004).

Nancy Glazener argues that the *Atlantic* magazine "had greater authority over American literature than any other institution did" in the latter half of the nineteenth century.[2] Because of this, the shifting center of US literary culture from Boston to New York City is sometimes personified in *Atlantic* editor William Dean Howells. Howells moved from Ohio to Boston to make his literary mark, and then moved from Boston to New York in the 1880s.[3]

Yet the material infrastructures of literary culture involved in such a shift—the means of production, distribution, and the skilled workers involved—could not possibly move as easily as did Howells's prestige. While *The History of the Book in America* has already covered many details of this shift, it does not cover the changing imprint geographies of US publishing in general and US literary publishing in particular, in part because this data was neither publicly available nor in a computationally tractable form.[4]

[2] Nancy Glazener, *Reading for Realism: The History of a U.S. Literary Institution, 1850-1910*, New Americanists (Durham: Duke University Press, 1997), 5. See also Nancy Glazener, *Literature in the Making: A History of U.S. Literary Culture in the Long Nineteenth Century*, Oxford Studies in American Literary History (New York, NY: Oxford University Press, 2016).

[3] See, e.g., Nancy Bentley, "Literary Forms and Mass Culture, 1870–1920," in *The Cambridge History of American Literature: Volume 3: Prose Writing, 1860–1920*, ed. Sacvan Bercovitch, vol. 3, The Cambridge History of American Literature (Cambridge: Cambridge University Press, 2005), 65–286, 250. This symbolically important move cannot be pinned to a single year since, between 1886 and 1891, Howells moved back and forth between Boston and New York several times. Susan Goodman and Carl Dawson, *William Dean Howells: A Writer's Life* (Berkeley: University of California Press, 2005), xxiv, 279.

[4] Scott E. Casper, ed., *The Industrial Book, 1840-1880*, A History of the Book in America, v. 3 (Chapel Hill: Published in association with the American Antiquarian Society by the University of North Carolina Press, 2007); Carl F. Kaestle and Janice A. Radway, eds., *Print in Motion: The Expansion of Publishing and Reading in the United States, 1880-1940*, A History of the Book in America, v. 4 (Chapel Hill: Published in association with the American Antiquarian Society by the University of North Carolina Press, 2009).

2 LC MARC Distribution Services books data

For the first time in 2015 and most recently in 2020, the Library of Congress (LC) released its open access MARC Distribution Services (MDS) books dataset to the public, which contains more than 11 million book records.[5] Although these records cannot be said to list every relevant book, it nevertheless remains the largest such catalog to which there is public access.

Figure 1 shows how one very famous record among the 11 million appears on the web.

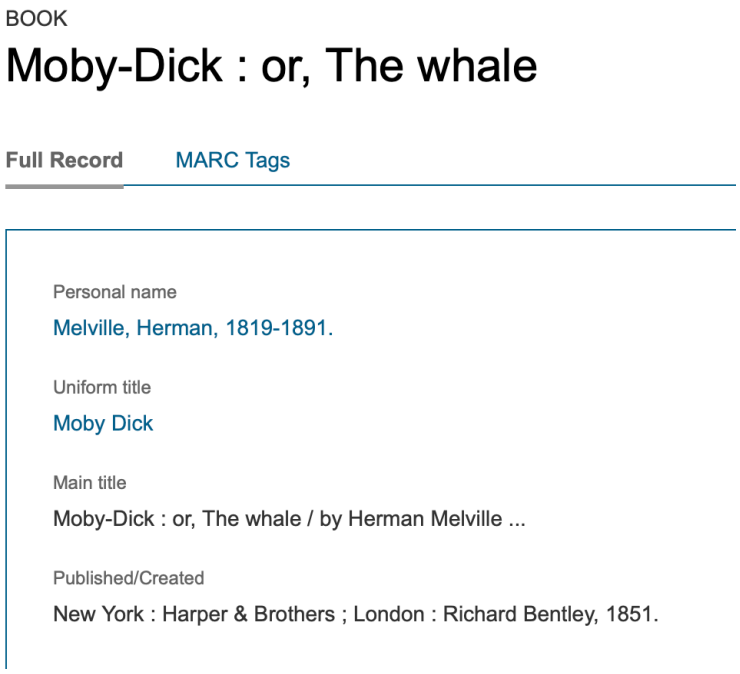

Figure 1: A screenshot of the LC catalog record for *Moby-Dick* (1851).

Below is some of the underlying XML that creates Figure 1, which exemplifies how all of the records are formatted:

[5] "Books All : MDSConnect Dataset," Compressed Data, *Library of Congress* (https://www.loc.gov/item/2020445551/, 2020). Thanks to Jaime Mears specifically and LC Labs generally for informing me about this dataset, as well as directing me to past work using this data by Matt Miller and Jer Thorp.

```
<datafield ind1="1" ind2="0" tag="245">
    <subfield code="a">Moby-Dick :</subfield>
    <subfield code="b">or, The whale /</subfield>
    <subfield code="c">by Herman Melville ...</subfield>
  </datafield>
  <datafield ind1=" " ind2=" " tag="260">
    <subfield code="a">New York :</subfield>
    <subfield code="b">Harper &amp;amp; Brothers ;</subfield>
    <subfield code="a">London :</subfield>
    <subfield code="b">Richard Bentley,</subfield>
    <subfield code="c">1851.</subfield>
  </datafield>
```

All of these records were created by librarians to describe "monographs written in nearly all the world's languages and published anywhere" including "Cataloging in Publication (CIP) records and minimal level cataloging records."[6] My contribution is extracting, cleaning, and analyzing the data that librarians created in a way that is immediately useful for scholars of US literature over the long nineteenth century. I provide both a working model and code for those in any other field who wish to work with any of the millions of LC MARC records I do not consider here, and publicize the uses of this resource to address questions

[6] "Books All (Database)," *The Library of Congress Cataloging Distribution Service* (https://www.loc.gov/cds/products/product.php?productID=5, June 2024). For more about the history of MARC records and their relation to the Library of Congress, see Betty Furrie, "Understanding MARC Bibliographic: Machine-Readable Cataloging" (https://www.loc.gov/marc/umb/, 2009) and MARC Standards Office, "A Brief Bibliography of Writings on MARC," *Library of Congress* (https://www.loc.gov/marc/readings.html, September 2011).

of literary history.

3 Measuring imprint geographies

Americanists will be familiar with the "PS" subclass of the LC's Language and Literature classification ("P"), which I use to study the changing imprint geographies of US literature over the long nineteenth century.[7] Generally, I check every classification value of each of the 11 million records in the MDS books dataset to see if any of its classifications are part of the "PS" subclass. I then extract selected fields from the relevant records. Finally, I clean the values for year and place of publication to make them computationally tractable. Readers interested in the details should consult the repository, which includes code to reproduce these results locally, and which can be trivially modified to extract any other records and fields in which they may be interested.[8]

3.1 Findings

Given the received narrative from US literary histories, I expected to see New York's publishing dominance rise consistently over the century, reaching a more or less unchallenged position after Reconstruction begins in 1865. But the data only partially support that narrative.

Figure 2 shows a five-year rolling average of the absolute growth in the amount of annual publishing by imprint city in American literature from 1750 to 1920. Note that this

[7] "Library of Congress Classification Outline," *Library of Congress* (https://www.loc.gov/catdir/cpso/lcco/, May 2024).

[8] Erik Fredner, "Getting Nineteenth Century Data from Library of Congress Book Records" (https://github.com/erikfredner/c19dc, June 2024).

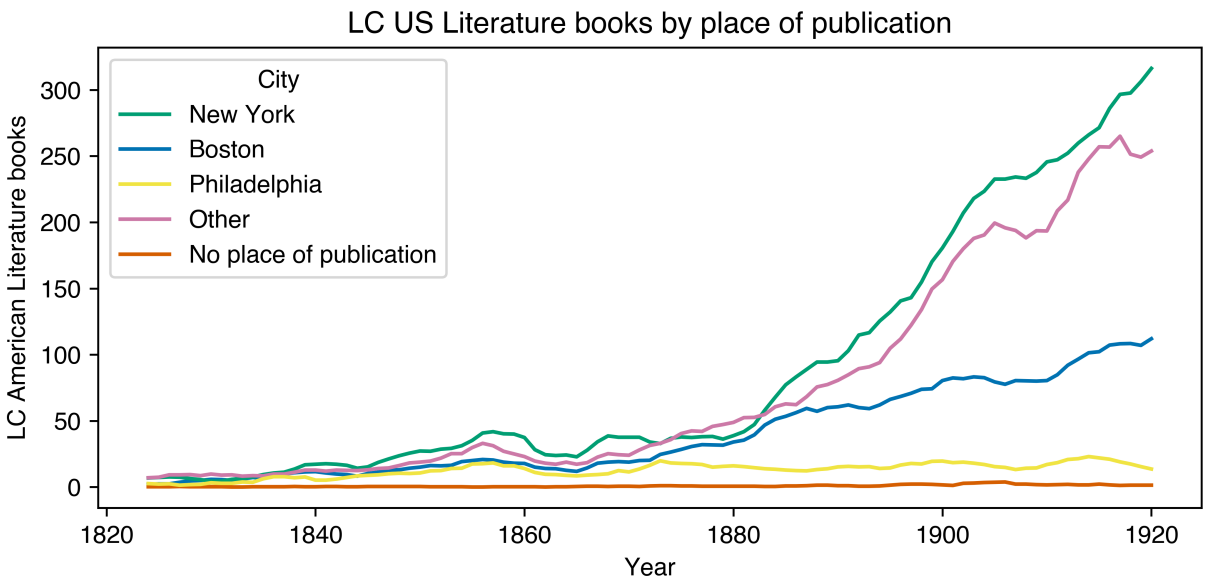

Figure 2: Five-year rolling average of LC US literature books by place of publication

represents *cataloged* publications, not sales or any other measure of readership beyond what is implied by cataloging librarians' judgment. While it is quite likely that books printed in New York sold better than books printed in, say, Indianapolis, this data cannot tell us much about that disparity.

However, the absolute growth in publication makes it difficult to see the changing geographies of US imprints. Measuring proportional change, as I do in Figure 3, suggests that uncertainty as to whether Boston, Philadelphia, or New York would become the publishing capital of US literature is over as early as the 1840s, not in the post-Civil War period as is often assumed.[9] In fact, the period after Reconstruction shows a brief but real *decline* in New York's importance.

More important than revising a timeline, however, is this: The data suggest that it is never the case that more than half of nineteenth-century US literature has a New York

[9]Unlike Figure 2, which begins at 1750, Figure 3 begins at 1820. This is because the number of publications per year in the period prior to 1820 is so low that proportional representation is noisy and misleading.

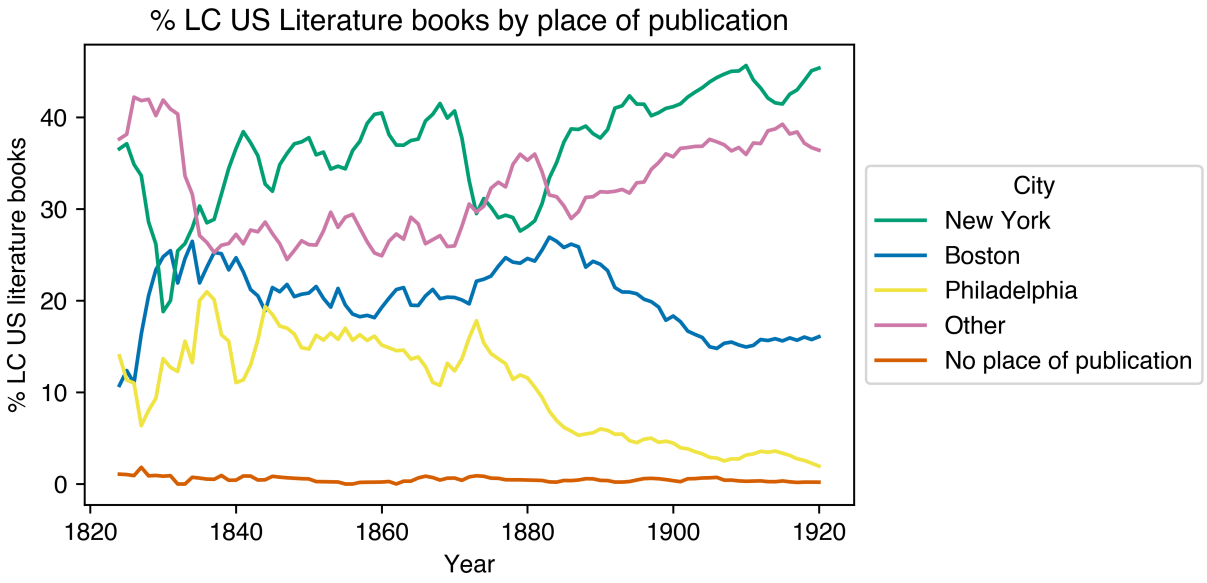

Figure 3: Five-year rolling average of the percentage of LC US literature books by place of publication

imprint. Measured in catalog records, most of US literary publishing happens elsewhere. Relatedly, the big surprise of these charts is the size of "Other," which groups together every place of publication not named in the legend. Even as Boston and Philadelphia's shares of US literary publication decline, Other keeps pace with (and occasionally exceeds) New York. New York may be the most important single place, but the data suggest that is far from the only place.

There is much more to be said about these charts elsewhere. For now, this example shows how one can use the transformed MDS Connect books data to study questions of historical and critical interest for nineteenth-century literatures. Other possible applications of the data—modeling republications, finding the distribution of unique authors by period, mining the text of the titles, etc.—abound within the code and data published with this short essay. Other LC record fields not included like subject headings (e.g., "Ahab, Cap-

tain (Fictitious character)–Fiction. Mentally ill–Fiction.") could be of interest. By means of small adjustments to the code in the repository, anyone interested in other geographies, other periods, or other fields can extract data from any subset of the LC's 11 million public records in a format suitable for analyses like the ones above.

Works Cited

Bentley, Nancy. "Literary Forms and Mass Culture, 1870–1920." In *The Cambridge History of American Literature: Volume 3: Prose Writing, 1860–1920*, edited by Sacvan Bercovitch, 3:65–286. The Cambridge History of American Literature. Cambridge: Cambridge University Press, 2005.

"Books All : MDSConnect Dataset." Compressed Data. *Library of Congress*. https://www.loc.gov/item/202 2020.

"Books All (Database)." *The Library of Congress Cataloging Distribution Service*. https://www.loc.gov/cds/products/product.php?productID=5, June 2024.

Casanova, Pascale. *The World Republic of Letters*. Convergences. Cambridge, Mass: Harvard University Press, 2004.

Casper, Scott E., ed. *The Industrial Book, 1840-1880*. A History of the Book in America, v. 3. Chapel Hill: Published in association with the American Antiquarian Society by the University of North Carolina Press, 2007.

Fredner, Erik. "Getting Nineteenth Century Data from Library of Congress Book Records." https://github.com/erikfredner/c19dc, June 2024.

Furrie, Betty. "Understanding MARC Bibliographic: Machine-Readable Cataloging."

https://www.loc.gov/marc/umb/, 2009.

Glazener, Nancy. *Literature in the Making: A History of U.S. Literary Culture in the Long Nineteenth Century*. Oxford Studies in American Literary History. New York, NY: Oxford University Press, 2016.

———. *Reading for Realism: The History of a U.S. Literary Institution, 1850-1910*. New Americanists. Durham: Duke University Press, 1997.

Goodman, Susan, and Carl Dawson. *William Dean Howells: A Writer's Life*. Berkeley: University of California Press, 2005.

Kaestle, Carl F., and Janice A. Radway, eds. *Print in Motion: The Expansion of Publishing and Reading in the United States, 1880-1940*. A History of the Book in America, v. 4. Chapel Hill: Published in association with the American Antiquarian Society by the University of North Carolina Press, 2009.

"Library of Congress Classification Outline." *Library of Congress*. https://www.loc.gov/catdir/cpso/lcco/, May 2024.

Office, MARC Standards. "A Brief Bibliography of Writings on MARC." *Library of Congress*. https://www.loc.gov/marc/readings.html, September 2011.