

Library Circulation Histories Workshop White Paper
March, 2022

Table of Contents:

Project Summary	2
Project Origins and Goals	2
Project Activities	5
The Workshop	5
Challenges and Changes	5
Project Team	6
Project Outcomes	7
Website	7
Audience	7
Technical Details	8
Publications	8
Project Evaluation	9
Key Findings	9
-For Project Managers	10
-For Researchers	13
-for Instructors	15
Project Continuation/Long-Term Impact	15
Appendix 1	17
Appendix 2	19
Appendix 3	20

Library Circulation Histories Workshop White Paper

[Draft]

Project Summary

The Library Circulation History Workshop (LCHW) assembled 25 scholars, librarians, and data scientists to consider how to make library circulation data more analytically powerful. The initial phase of the event took place online from April 19-30, 2021. The program for this portion of the workshop featured a Keynote, 6 panels of two presentations each, and a summative plenary session. Presentations included project demonstrations, discussions of key analytical issues, and the presentation of original research based on circulation data. The workshop's second phase encompassed two live meetings via web conferencing as well as follow-up discussions via email and a short survey. The Center for Middletown Studies hosted the workshop, with support from Ball State University Libraries and BSU's Digital Scholarship Lab. The Center received funding from the National Endowment for the Humanities and the Institute for Museum and Library Services. All of the workshop presentations and discussions are preserved and remain publicly accessible on the project website (<https://lchw.bsudsl.org/>).

Project Origins and Goals

The LCHW originated out of conversations between the leadership team of the [What Middletown Read project](#) (WMR) and scholars engaged in similar endeavors. What Middletown Read is an NEH-funded project that has reproduced in digital form the circulation records of the Muncie (Indiana) Public Library between 1891 and 1902, documenting more than 174,000 loans involving more than 4,000 patrons and 6,000 books and magazines. Several other initiatives have produced comparable resources, including the [Australian Common Reader](#), the [Easton Company](#)

[Library Database](#), the [City Readers Project](#), [Dissenting Academies Online](#), and [Books and Borrowing: An Analysis of Scottish Borrowers Registers](#). All of them reproduce library circulation records that document individual borrowing in the English-speaking world during some portion of the period between the late seventeenth century and the early twentieth century.

These projects provide a rich trove of evidence about the history of reading, a key component of the larger field of print culture history and one with the potential to grow further. They represent some of the best evidence we have about broad patterns of historical reading choices. Yet opportunities to exploit them more intensively remain, particularly through the application of computational methods. They also face a common set of technical and interpretive challenges. The LCHW offered an opportunity to share lessons learned, exchange ideas, and explore opportunities for collaboration. The What Middletown Read team had a particular interest in developing ideas for updating its web interface and database and the workshop aimed to provide other project leaders with similar input. We also hope the workshop's online presence will serve as a resource for those seeking to develop new projects.

Since the larger purpose of analyzing data derived from circulation records is to investigate reading experiences, the workshop organizers also recruited scholars engaged in comparable work. These included [Christine Pawley](#) and [Jennifer Burek Pierce](#), who have analyzed circulation data in their published work, as well as the creators of several other digital resources or tools tied to the history of reading, such as [Reading What Middletown Read](#) (a suite of tools for visualizing What Middletown Read data), [Circulating American Magazines](#), the [Reading Chicago Reading Project](#), the [Reading Experience Database](#) and [READ-IT](#) (an investigation tool for exploring reading experience in modern Europe). Mark Towsey, Director

of the [Libraries, Reading Communities and Cultural Formation in the 18th-Century Atlantic](#) project, delivered the Keynote.

The workshop proposal emphasized two key research questions that have arisen concerning the use of circulation data to investigate reading history: how to use circulation data to inform computational text analyses, and how best to use evidence from circulation records to investigate the formation of reading communities. The workshop also addressed other issues that intersect with these core concerns, including the evolution of libraries, project sustainability, and the potential for linking datasets. Lastly, the group considered ways of making circulation data more useful and accessible for instructional purposes.

LCHW organizers sought to engage several overlapping audiences. Most directly, the presentations spoke to scholars and developers engaged with creating and sustaining digital resources that incorporate historical circulation data. The workshop program included a significant share of the humanities researchers, librarians, and project managers engaged in such work. The panels also addressed scholars who, while not involved in the development of such resources, had an interest in using them to investigate the history of reading. This category includes historians, literary scholars, communications scholars, and librarians. In some instances, researchers are interested in using circulation data in combination with computational text analysis to investigate broad patterns of borrowing choices, but in others, they seek more specific information about particular books, authors, borrowers, or publishers. An important thread within LCHW's deliberations involved developing resources that supported both kinds of research. In addition, the Workshop considered ways of providing resources for instructors seeking to incorporate circulation data into their teaching and encouraged project managers to consider the best ways to foster pedagogical uses of circulation data.

Project Activities, Team and Participants

The Workshop: Phase One of the Workshop took place in an online, asynchronous format over two weeks (April 19-30, 2021). Each session featured recorded video presentations of research findings and/or project histories and a forum for online discussion. To initiate exchanges, each panel had a designated discussant who presented a series of questions for consideration. Discussion took place through online commenting over two weeks. The project team released Towsey's Keynote on April 19th, one day before other sessions opened. This step enabled us to feature his talk, which highlighted some of the workshop's overarching themes. The remaining six panels went live the following day, April 20th. On April 27th, the workshop's lead organizer, James Connolly, posted a plenary talk that summarized the event's initial findings. The comment forums for all of the sessions remained open through April 30th. The full program is attached as Appendix 1.

The second phase of the LCHW consisted of two live, online gatherings for workshop participants, one on June 19th and another on July 19th, 2021. These sessions provided an opportunity to discuss issues raised in the online forums associated with each panel and to contemplate the next steps. Twelve of the eighteen presenters participated in one of both meetings. The Project Director pre-circulated questions to guide discussion (see Appendix 2). Those unable to attend had the option to share thoughts via email or by completing an online survey.

Challenges and Changes: The Covid-19 pandemic was the principal challenge facing the project team and necessitated the largest changes. Our original plan for the Workshop was to gather interested scholars and project leaders, including both those listed in the original grant proposal as well as researchers recruited through a public call for proposals, in May of 2020 at Ball State University. The pandemic forced the postponement and ultimately the reconfiguration of this event. Initially, we delayed the face-to-face event until September 2020 but continuing restrictions made meeting in person at that time impossible. After consulting with the NEH, we converted the workshop to an asynchronous online format and held it during the spring of 2021. We chose the asynchronous approach because participants were located in the U.S., Europe, and Australia and time differences made live interactions complicated.

There were several challenges associated with the conversion to an online format, which necessitated a more robust website. Chief among them was how to create a commenting system that would foster meaningful and timely exchanges but that would not become overrun by irrelevant, bot-driven comments. Others were labor-intensive tasks, including an arrangement of the recording and transfer of the fairly large video files created for each presentation, managing the registration process, and setting up hosting for the entire site. We were able to address these and other issues by drawing upon the model established for the [Nearly Carbon Neutral Conferences](#) at the University of California-Santa Barbara.

When we first decided to shift to an online format, we held out hope that we could hold a briefer in-person gathering during the early summer of 2021, even if only for North-American-based participants. Pandemic restrictions persisted, so we instead organized the two meetings that constituted Phase Two of the Workshop via videoconferencing. The start times varied to accommodate participants from different time zones.

The Project Team consisted of:

James J. Connolly, Project Director

Douglass Seefeldt, Co-Project Director

Frank Felsenstein, Project Consultant

Linda McCaw, Project Manager

David Rodriguez, Web Developer

Aidan McBride, Graduate Assistant/Web Manager

Megan Vohs, Graduate Assistant

All other participants are listed in the program (Appendix 1).

Project Outcomes

Workshop Website: Our principal product is the [workshop website](#). It contains all of the video recordings of the fourteen workshop presentations, texts that presenters wished to share, and full transcripts of the discussions associated with each panel. The project team also set up a “[Updates](#)” page that contains news about publications, new data releases, and other relevant activities by participants.

Website Audience: The shift from in-person to online widened the workshop’s accessibility. In the end, we had 98 registered users for the workshop, 24 of whom were on the program. While registration enabled a user to participate in the online forums associated with each panel, those discussions included only people listed on the program. Nevertheless, there was a robust exchange, with the 24 participants posting 154 comments, many of which were several

paragraphs long. Our analytics indicate that the workshop’s site had 301 discrete users, who compiled a total of 572 sessions during the first phase. A clear majority—74%—were based in the U.S. During the workshop period, the site had a fairly low bounce rate (37.59%) suggesting a relatively high level of engagement. After the workshop, the site continued to receive interest, attracting 722 visitors over the period from the opening of the workshop through July 31, 2021.

Technical and Preservation Details: As noted above, the project website is hosted on WordPress, using the [Nearly Carbon Neutral Conferences](#) model established at the University of California-Santa Barbara. Our site designers used the [wpForo plugin](#) to manage the discussion boards, as well as anti-spam (Askimet), embedding (oEmbed), and PDF viewer (PDFjc) plugins. The Center for Middletown Studies will maintain the website for five years after which time it will work with Ball State University Archives and Special Collections to arrange archival storage that preserves access to the site’s contents via a [web archiving service](#). In addition, the closed-captioned video files and discussion transcripts for each presentation will be preserved separately in University Archives. Presenters provided the necessary permission for these uses.

Publications: The Project Director has discussed plans for a peer-reviewed special issue of *Libraries: Culture, History, and, Society* that would feature revised material from several workshop presentations. The journal has a publication backlog which means the special issue would appear in 2023 or 2024. The journal’s editors also propose that the issue include an open call for additional submissions, which may elicit additional scholarship that employs library circulation data or uses comparable methods.

Project Evaluation and Impact

As noted, workshop organizers received feedback through two live forums held after the initial workshop as well as via a short online survey (See Appendices 2 and 3 for the questions posed in this phase). In both venues, feedback from participants was largely positive. While they expressed a general preference for a face-to-face gathering, they were satisfied that the format for the workshop provided a vehicle for robust discussion. The website and moderation system worked as planned, enabling participants to post comments in the forums without difficulty and spam or other interference. One fortuitous result of the shift to a web-based format is that it has provided us with a workable alternative to in-person events that overcome the scheduling challenges created by time-zone differences. It also enabled us to preserve and share a more substantial record of the proceedings for those interested.

Key Findings: The Workshop's forums and follow-up meetings generated significant exchanges about several broad areas. Many of the most substantive discussions focused on how project leaders can best develop and sustain online resources that incorporate circulation data. Among the most notable were best practices for preparing, sharing, and preserving library circulation data and for developing interfaces that facilitated both traditional and computational approaches to research. Another significant vein of the conversation revolved around questions of interest to scholars seeking to use circulation data in their research. Key topics addressed in these exchanges included the utility of statistical modeling and the importance of understanding library infrastructure when analyzing circulation data. Several participants encouraged project teams to weigh the value of these datasets as pedagogical resources and issued a call to generate teaching modules and other tools to support classroom activities.

-For Project Managers: During the follow-up sessions, the discussion of how to prepare, share, and preserve data was perhaps the most wide-ranging. There is of course considerable variation in the form of the original records employed in the various projects connected to the workshop. In the case of older records such as the 18th-century records of the Library of Innerpeffray (Scotland) [described by](#) Matthew Sangster and Katie Halsey, register entries took quasi-narrative form. Nineteenth and early twentieth-century records, such as those captured in the What Middletown Read project, used printed ledgers that captured bibliographic and individual information in standardized categories. The more recent circulation data assembled for the Reading Chicago Reading project documents checkouts but not individual borrowers and originated in digital form. Privacy concerns and related legal questions will differ according to time and place and will have to be taken into account in developing a project. Given this variety, there are no uniform prescriptions for handling data. Tightening privacy laws, particularly in Europe, may also limit the collection and preservation of some data going forward.

Despite the differences in the nature of the circulation data from one project to the next, project leaders agreed on the value of providing access to the data set independently of the project's web interface. While an online search interface remains a vital element of these projects, particularly for users investigating specific book, author, patron, or publisher, direct access to data facilitates computational approaches that employ text analysis or network visualization. Downloadable data files, generally in a CSV format, enable users to analyze datasets independently of the project's primary interface, which can become dated quite rapidly. Given the constant evolution of computational tools and analytical methods, access to the underlying dataset(s) is especially useful for more complex undertakings. Participants also agreed that there is a necessity to engage in more systematic data practices, particularly careful

version control. Several recommended publishing and documenting multiple versions of a dataset in an openly accessible repository (e.g. GitHub), which provides users with access to raw data as well as the processed datasets that project web interfaces employ.

Workshop discussants also raised the possibility of providing access to a corpus (or corpora) of text files for users interested in linking computational text analysis and circulation data. Alternatively, a Hathi Trust capsule may serve this purpose. Ready access to these kinds of resources would facilitate such an approach. None of the projects presented at this workshop included such an affordance, although scholars from Washington University in St. Louis's Humanities Digital Workshop have compiled a corpus of texts that circulated in the Muncie Public Library between 1891 and 1902 and which are documented in the What Middletown Read database. The [presentation](#) by Lynne Tatlock, Doug Knox, and Steve Pentecost illustrates the kind of analysis that combines textual and circulation data. One challenge in creating this kind of resource is matching available digital texts to the specific edition of a book that a library held.

Since each project differed, there was no single interface template that is likely to work in every case. Naturally, consideration of the needs of various groups of users as well as of the amount and kinds of information contained in the data should inform the design. Beyond developing means to query circulation data, there are opportunities to provide basic visualization tools (see for example Julieanne Lamond's [discussion](#) of the Australian Common Reader's visualization tools) and other complementary resources, such as digitized library reports, marginalia, or lesson plans for instructors. When possible, project developers should provide access to digitized versions of the original library records so users can examine them. Including a field or fields in the data structure for comments about books, borrowers, and loans is advisable, either to explain data anomalies or to provide useful contextual information.

Several presentations, including those by [Kyle Roberts](#), [Julieanne Lamond](#), and [Edmund King](#), explored questions of sustainability and the need to plan for the sunsetting or evolution of projects. These issues emerged as another key thread in LCHW discussions. For most of the projects represented in the workshop, the challenges of creating and launching a digital resource based on circulation records occupied so much attention that consideration of how to conclude these efforts rarely arose during planning. Lamond highlighted the [Endings Project](#) at the University of Victoria, which compiles resources and advice for building sustainable digital resources, including technical guidance for shifting from a dynamic to a static website. Project leaders also discussed the importance of securing long-term institutional support and developing a clear understanding of how a project will be maintained or archived without reliance on a particular individual or individuals. In one instance, increased emphasis on the pedagogical uses of circulation data and associated resources helped project leaders secure more durable support from their home institution. Several people cited the importance of approaching the issue of sustainability in human-centered terms, considering who will continue a project once its original leaders move on, as well as in terms of funding and technical resources. [Minimal computing](#) also warrants consideration in this context, both as a response to limited infrastructural resources as a means of reducing obsolescence.

Finally, the opportunity to link together the datasets developed for various projects received consideration. The wide differences between 18th-century university libraries, private subscription libraries, public libraries of the nineteenth and twentieth centuries, and libraries in the digital age create large obstacles to such connections. So too do the different approaches outlined in workshop presentations about the [Reading Experience Database](#), [READ-IT](#), and [Circulating American Magazines](#). Moreover, the precise research questions animating

scholarship that employs these varied datasets differ considerably. However, the idea of employing specific books as nodes for assembling scattered borrowing and reception data drew some interest as a basis for connecting projects that document activity from the same era. The discussion following Mark Towsey's [Keynote](#) and subsequent conversation suggested the potential for linking the resources compiled for his broader project with those of Dissenting Academies Online, the Books and Borrowing Project, the Reading Experience Database. Linking these and other resources on both sides of the Atlantic offers a means of exploring the histories of particular books and authors during the long 18th century. Exploring the potential for such connections represents one avenue of ongoing work emerging from the workshop, although the practical challenges facing such an undertaking are substantial.

-For Researchers: For scholars interested in using circulation data for their research, several considerations arose. For some, quantitative analysis, often linked with distant reading methods or network visualization (or both), will serve as the principal method. For others, the primary approach may involve documenting the experiences of individual borrowers or groups of borrowers, or examining the circulation and reception of particular books or the corpus of an author. Instructors in several disciplines have used library circulation data to model these and other research methods in their courses. Projects that reproduce and analyze circulation data should consider each of these sets of users as they develop their resources.

One promising avenue for further research involves marrying text analysis with circulation data. This approach permits scholars to weight texts according to the frequency with which they circulated as part of broader investigations that employ distant reading techniques or to analyze themes and other literary elements that attracted particular demographic groups (see for instance the workshop [presentation](#) by Tatlock, Knox and Pentecost). As circulation data

becomes more widely accessible, it will be helpful to develop a set of best practices, or at least more consistent methods, for this kind of analysis. The workshop initiated discussions among several participating scholars about the development of methodological work in this area.

Other forms of computational work that uses circulation data can help us better understand borrowing choices and, with the caveat that borrowing and reading are not the same things, reading experiences. Modeling that employs textual elements of books, demographic data, and other factors may enable predictive insight that helps us get a strong sense of readers' tastes. This question animates a considerable amount of research on reading, as print culture historians seek to understand the extent to which reading choices and interpretive experiences are shaped by the wider culture and social order. There are complexities to this approach, as noted in Alex Leslie's [workshop presentation](#) and the subsequent discussion. But it helps us sort out the variety of factors shaping borrowing, including but not limited to the themes of a text or the social background of the borrower. As Robin Burke and John Shanahan [explain](#) in the case of the One Book One Chicago program, the nature of the holdings, shelf effects (when a library patron seeks one book but selects another book shelved near it), and a text's level of reading difficulty are among the factors that also affect borrowing patterns.

These findings also point us to another key element of circulation data analysis, the role of contextual elements such as library infrastructure. Christine Pawley [lays out](#) the importance of taking into account the organizational framework in which borrowing occurs. Careful attention to the nature of different types of libraries as well as the history and character of specific institutions is necessary to a fully developed analysis of circulation data. Giving weight to a library's holdings and its physical organization, for example, is necessary for assessing circulation patterns. Attention to formal and informal rules governing access to the library use

should also influence demographic analyses of borrowing, as should the relationship between circulation via libraries and other forms of distribution and recommendation. The latter might include published book reviews, as Julieanne Lamond [suggests](#), the circulation of periodicals [described](#) by Brooks Hefner and Edward Timke, or, in the digital age, social-media recommendations such as those described in Jennifer Burek Pierce's [workshop talk](#). More broadly, Shafquat Towheed's [presentation](#) about READ-IT provides an instructive model for considering the range of textual and contextual elements that intersect to create reading experiences.

-For Instructors: Several presentations took note of the pedagogical potential of these projects. In addition to obvious opportunities to explore various aspects of print culture history, these resources also provide students with the chance to engage in various forms of data analysis and archival research that have wider applications for digital humanities education. Frank Felsenstein's [presentation](#) offers the most extensive consideration of instructional approaches, although several other talks discuss them as well. During wrap-up conversations, several workshop participants noted the importance of incorporating teaching resources into a project as a way not only to attract a broader audience but also to help sustain it.

Project Continuation and Long-Term Impact

Perhaps the most significant long-term outgrowth of the LCHW is cementing connections among scholars who use circulation data to explore reading experiences. There are plans to submit proposals for conference sessions, as well as for joint publications that explore the history of libraries in relation to the evidence provided by circulation records. Workshop participants also proposed a one-year-on meeting where those interested can share updates about their work,

which the LCHW project director will propose during the spring of 2022. The Center for Middletown Studies will maintain the workshop website, including the Updates page, as an embryonic version of a broader online clearinghouse for a scholarship that uses library circulation data and similar material. We hope to create such a resource under the aegis of the Society for the History of Authorship, Reading, and Publishing (SHARP) or another institution.

A key reason for organizing the LCHW was to gather insights that will inform an update of the What Middletown Read database, which launched in 2011. Our plan is to prepare one or more funding proposals for a WMR reboot that capitalizes on recent advances in digital humanities and access to new resources. Revisions include an updated interface, enhanced datasets, ready access to corpora that facilitate distant-reading methods, and a more robust sustainability plan. Other participants are also in the midst of preparing new funding proposals for their projects that will employ insights from the workshop. Workshop participants have already signed on as consultants or provided letters of support for other projects.

A core aim of the LCHW was to develop ideas for making circulation data more analytically powerful and more accessible to a wide range of scholars. The ideas we developed about how to provide data access that serves the needs of a broad range of researchers and instructors, facilitate computational work, design interfaces, and pursue opportunities for collaboration should help achieve this end. While the upsets of the Covid-19 pandemic created obstacles to our efforts, it also pushed us to employ a forum that preserves its substance more thoroughly and makes it more widely accessible.

Appendix 1: Workshop Program

Keynote (Released April 19)

Mark Towsey (University of Liverpool), History Now is the Favourite Reading: From Library Circulation Records to A History of Reading.

Discussant: Frank Felsenstein, Ball State University

Panels (Panels 1-6 released on April 20)

Panel 1: Building and Sustaining Databases

Frank Felsenstein (Ball State University), The What Middletown Read Database as an On-Line Resource for Gauging Readership Trends in Late Nineteenth-Century America: J.T. Trowbridge's *The Tinkham Brothers' Tide-Mill* as Exemplar

Julieanne Lamond (Australia National University), The Legacy Database and the Question of the "Representative" Reader

Discussant: Raphael Acosta, University of Kansas

Panel 2: Eighteenth-Century Readers and Circulation Data

Kyle Roberts (American Philosophical Library), Growing Pains and Digital Library Projects: The Opportunities and Challenges of Follow-on Initiatives

Matthew Sangster (University of Glasgow) and Katie Halsey (University of Stirling), Books and Borrowing Across Scotland, 1750-1830

Discussant: Mike Sanders, University of Manchester (UK)

Panel 3: Placing Readers and Reading in Context

Christine Pawley (University of Wisconsin), Linking Ordinary Readers with Texts:
American Public Libraries and the Infrastructure of Print

John Shanahan (DePaul University) and Robin Burke (University of Colorado),
Modeling Contemporary Reading Behavior at City-Scale: The “Reading Chicago
Reading” Project

Discussant: Melanie Walsh, Cornell University

Panel 4: Circulation Data Beyond the Library

Jennifer Burek Pierce (University of Iowa), Reading Data Documenting 21st Century
Reading: *Vlogbrothers*, The Nerdfighter Census, and Actual Readers

Brooks E. Hefner (James Madison University) and Edward Timke (Duke University),
Circulating American Magazines: Lessons from the Audit Bureau of Circulations Data

Discussant: Kalani Craig, Indiana University

Panel 5: Using What Middletown Read

Lynne Tatlock, Steve Pentecost, Doug Knox (Washington University at St. Louis),
Reading the American South in the Muncie Library

Alexander Leslie (Rutgers University), Patterns and Predictability in Borrower Behavior
Discussant: Jordan Bratt, Ball State University

Panel 6: From RED to READ-IT

Edmund King (Open University), Looking Back on The Reading Experience Database,
1450-1945

Shafquat Towheed (Open University), From UK-RED to READ-IT (2018-2021): Insights and Challenges for the Future

Discussant: Doug Seefeldt, Clemson University

Plenary Session (Released on April 27)

James Connolly (Ball State University), Circulation Histories: Initial Findings

Discussion forums for each panel remained open through April 30.

Appendix 2. Live Forum Discussion Questions

1. What do you see as the best opportunities for us to collaborate or work cooperatively? This might include everything from a loose scholarly network to a clearinghouse for untapped data, to a long-term effort to link (some of) our datasets.
2. For those who employ computational methods, what are the best practices for making library circulation data most usable for these techniques (e.g. distant reading, network analysis, modeling)?
3. What can we learn from each other about (re)designing projects (and making data accessible) in ways that can accommodate our varied users (scholars, students, members of the public)? And what are the connections between these issues and questions of sustainability that we all face?
4. Christine Pawley has noted the importance of “looking at the whole picture of the library” when analyzing circulation data. Most of us do a pretty good job of that, but how do we make it so other users will appreciate institutional, community, and other contexts in which borrowing occurred?

Appendix 3: Online Survey Questions:

1. A key aim of the Library Circulation Histories Workshop was to share and develop ideas about how to make circulation data more analytically powerful. How well do you think the Workshop achieved that end? What aspects of the Workshop were most useful for you?
2. What topics or issues related to circulation data do you think to require further exploration?
3. What do you think were the strengths and/or weaknesses of the asynchronous online format we adopted in response to Covid restrictions?
4. What do you see as desirable next steps for the group participating in the Workshop?
5. Do you have other comments or suggestions?