The Women's Print History Project

Mind the (Data) Gaps, *The WPHP Monthly Mercury*

Moffatt, Kate, and Kandice Sharren, hosts. "Mind the (Date) Gaps." *The WPHP Monthly Mercury*, Season 1, Episode 6, 18 November 2020, https://womensprinthistoryproject.com/blog/post/51.

PDF Edited: 22 April 2024

Mind the (Data) Gaps
Kate Moffatt and Kandice Sharren

**Mind
the (Data)
Gaps**

episode   six
THE WPHP MONTHLY MERCURY

Have you ever wondered, *"Where does all the WPHP data come from?"* Well, look no further than this month's episode of *The WPHP Monthly Mercury*! In Episode 6, "Mind the (Data) Gaps", hosts Kate Moffatt and Kandice Sharren bring listeners behind-the-scenes to explore a few of the *many* superhero sources we use to create WPHP title data.

The WPHP collects data for titles produced by women between 1700 and 1836; accordingly, this episode briefly introduces you to sources like the ESTC, *The English Novel*, and the *Orlando: Women Writers in the British Isles from Beginnings to Present* database, to name just a few! Kate and Kandice also acknowledge where the data gaps lie in the WPHP, how they tend to come about, and how we grapple with them.

And there is grappling required on multiple levels when working with the sources of the WPHP: not only are we managing the gaps in our data, but we are also wrangling the different levels of information provided by each resource. Our data is not a simple replication of what is provided by our excellent sources—rather, we use multiple sources in conjunction with one another, amalgamating what we find to create coherent and cohesive data about the titles in the WPHP to create as full a picture as possible of women's involvement in print during our period. It is through these efforts, as well as our persistent transparency around our processes, our creation of subject guides and podcast episodes, and our use of widespread and various sources, that we hope the WPHP, too, can contribute to filling some of the existing gaps in women's book history.

For clarification on some of the bibliographical terms we use in this episode, you can find definitions of terminology in our documentation of title records here. You can also find a full list of our sources here.

***WPHP Monthly Mercury* Episodes Referenced**

Season 1, Episode 5: "Of Monks and Mountains!!!"
Season 1, Episode 4: "A Bibliographical Education"
Season 1, Episode 2: "Women in the Imprints"

WPHP Sources Referenced

Roy Bearden-White

Google Books

HathiTrust Digital Library

English Short Title Catalogue (ESTC)

The English Novel, 1770-1829

The English Novel, 1830-1836

The English Novel, 1800–1829 & 1830–1836: Update 7 (August 2009–July 2020)

Jackson Bibliography of Romantic Poetry

John Newbery and His Successors, 1750-1814

A Bibliography of Eliza Haywood

Osborne Collection of Early Children's Books

British Library

American Antiquarian Society

Eighteenth Century Collections Online (ECCO)

Nineteenth Century Collections Online (NCCO)

Sabin Americana: History of the Americas 1500–1926

Slavery and Anti-Slavery: A Transnational Archive

Archive.org

Orlando: Women's Writing in the British Isles from the Beginnings to Present
British Travel Writing: Women's Travel Writing, 1780–1840

WPHP Records Referenced

Burney, Frances (person, author)

Evelina (title)

Cecilia (title)

Camilla (title)

The Wanderer (title)

Ann Lemoine (firm, publisher)

Lemoine, Ann (person, publisher)

John Newbery (firm, publisher)
Francis Newbery (firm, publisher)
Elizabeth Newbery (firm, publisher)
Newbery, Elizabeth (person, publisher)
Haywood, Eliza (person, author)
Eliza Haywood (firm, publisher)
Radcliffe, Ann (person, author)

Works Cited

American Antiquarian Society catalog. American Antiquarian Society, https://www.americanantiquarian.org/.

Bearden-White, Roy. *How the Wind Sits; or, the History of Henry and Ann Lemoine, Chapbook Writers and Publishers of the Late Eighteenth Century*. 2007. Southern Illinois University Carbondale, M.A. Thesis.

British Library catalog. British Library, www.bl.uk/.

British Travel Writing: Women's Travel Writing, 1780–1840. U of Wolverhampton, 2014, www4.wlv.ac.uk/btw/.

Eighteenth Century Collections Online. Gale, www.gale.com/primary-sources/eighteenth-century-collections-online.

English Short Title Catalogue. British Library, www.estc.bl.uk/.

Garside, Peter, James Raven, and Rainer Schöwerling, eds. *The English Novel, 1770–1829*. Oxford UP, 2000.

Garside, Peter and Anthony Mandal, eds. *The English Novel, 1830–1836*. Cardiff, Centre for Editorial and Intertextual Research, 2003.

Google Books, Google Inc, www.books.google.com/.

HathiTrust Digital Library. U of Michigan, www.hathitrust.org/.

Internet Archive. 1996, www.archive.org/.

Jackson, J. R. de J. *Jackson Bibliography of Romantic Poetry*. U of Toronto, 1975, www.jacksonbibliography.library.utoronto.ca/.

Leuner, Kirstyn, Deborah Hollis, and Kate Ozment, eds. *The Stainforth Library of Women's Writing*, www.stainforth.scu.edu.

Nineteenth Century Collections Online. Gale, www.gale.com/primary-sources/nineteenth-century-collections-online.

Nygren, Catherine. "Women's Travel Writing." Unpublished dataset, 12 September 2018.

Orlando: Women's Writing in the British Isles from the Beginnings to the Present. Cambridge UP, 2008, www.orlando.cambridge.org/.

Osborne Collection of Early Children's Books. Toronto Public Library, 1949, www.torontopubliclibrary.ca/osborne/.

Roscoe, Sydney. *John Newbery and His Successors*. Five Owls, 1973.

Sabin Americana: History of the Americas 1500-1926. Gale, https://www.gale.com/sabin-americana-history-of-the-americas-1500-1926.

Slavery and Anti-Slavery: A Transnational Archive. Gale, https://www.gale.com/primary-sources/slavery-and-anti-slavery.

Spedding, Patrick. *A Bibliography of Eliza Haywood.* Pickering & Chatto, 2004.

00:00:00	♫	[music playing]
00:00:09	Kandice Sharren (co-host)	During a recent bout of writing, I ran up against the question of how eighteenth-century novelist Frances Burney was identified towards the end of her writing career. I knew that as a general rule, she did not identify herself by name, preferring to describe her authorship on the title pages of her novels as the author of *Evelina*, *Cecilia* and *Camilla*, but her final novel, *The Wanderer*, published in 1814, has been widely understood as a critical disaster.
00:00:36	Kandice Sharren (co-host)	So I was curious whether that affected how her publishers marketed reprints of her novels in subsequent years. Naturally, I took to the WPHP to find out. We have a signed author field that captures how authors describe themselves on their title pages and occasionally in prefaces and dedications. So I was excited to check this information out. You can imagine my surprise when I discovered that post-1814 editions of Burney's two most popular works, *Evelina* and *Cecilia*, were not in the database.
00:01:13	Kandice Sharren (co-host)	How embarrassing! Burney was a major author of our period and editions of these novels appeared regularly throughout the 1810s and 1820s. Digitized copies of these editions are readily available on open-access resources, including *Google Books* and *HathiTrust*. How, then, did this oversight occur? It's easier than you might think.
00:01:35	Kandice Sharren (co-host)	One of the major difficulties that the WPHP faces is a lack of comprehensive resources covering the period between the turn of the nineteenth century and the beginning of the Victorian period. In this month's episode, we're going to talk about our sources: how they help us identify titles that we should include in the database, what their limitations are and what kinds of gaps we still need to fill.
00:01:59	♫	[music playing]
00:02:07	Kate Moffatt (co-host)	Hello and welcome to *The WPHP Monthly Mercury*, the podcast for *The Women's Print History Project*. The WPHP is a bibliographic database that collects information about women and book production during the eighteenth- and nineteenth centuries. My name is Kate Moffatt—
00:02:22	Kandice Sharren (co-host)	and I'm Kandice Sharren—

00:02:24	Kate Moffatt (co-host)	and we are longtime editors of the WPHP and the hosts of this podcast. On the third Wednesday of every month, we'll introduce you to anecdotes, puzzles and problems related to recovering evidence of women's involvement in print.
00:02:36	♫	[music playing]
00:02:42	Kate Moffatt (co-host)	The WPHP has more than 60 sources listed in its sources table, from the comprehensive—the *Eighteenth Century Short Title Catalogue*—to the relatively niche—Roy Bearden-White's MA thesis on Ann Lemoine, a chapbook seller and publisher—and we add new sources to the list all the time. We use our sources to identify and verify the titles that we include in the database. In order for a title to be verified, we have to consult at least two sources. One of which is preferably a digital copy that we can inspect, although a detailed and reliable bibliographical resource will do in a pinch.
00:03:14	Kate Moffatt (co-host)	When we have the opportunity to visit a library or collection and inspect a book by hand, we don't need to use a second source to verify it. Our sources can be grouped into four main categories: bibliographies; digitized full-text resources; library catalogues and libraries; and resources that have some limited bibliographical information but are more focused on other things, like authorship. The range of sources available to us is both a blessing and a curse: most of the information that we are looking for is available *somewhere*.
00:03:43	Kate Moffatt (co-host)	However, finding ways to work through it systematically is an ongoing challenge. We're going to introduce you to some of the sources that we use specifically for the title records of the WPHP. These are our superheroes and our rock stars. And we'll talk through some of the ways that these sources support our work and some of the specific limitations that we face and how we tackle them.
00:04:01	♫	[music playing]
00:04:11	Kandice Sharren (co-host)	We're going to start with our bibliographical resources. These guys make up the heart of the WPHP data because they're focused on the same kind of data we're interested in collecting. We need these sources at the bare minimum to provide us with titles, authors when they're known, and dates of publication. Ideally, they also include imprint information, pagination, format, volume information, that kind of detail, but they don't have to. So, our bibliographical resources, because our period is so expansive, tend to focus on particular time periods within the period we're looking at as well as specific genres or specific individuals involved in production.

00:04:57	Kate Moffatt (co-host)	One of our bibliographical resources that tends to focus on a particular time period is the *English Short Title Catalogue*. This is arguably our single most important resource because it is relatively comprehensive and it covers 1473 to 1800. At present, we have 4,968 title records in the WPHP that cite the ESTC as a source. And we're constantly adding more. The ESTC contains almost complete bibliographical information. Sometimes, it abbreviates titles and imprints. So we can't rely on imprint information from the ESTC, but it's one of the most bibliographically detailed.
00:05:32	Kandice Sharren (co-host)	We also have a lot of data from the ESTC because we had a data ingest early in our history that ended up forming the backbone of our data. Our developer at the time searched the author field for women's names and imported all of the results. While this strategy resulted in some false positives and didn't capture every female author by any means, it gave us a lot of titles to start editing and to start trying to verify. At that early stage, we didn't know how many women were actually involved in the book trades. Turns out it's a lot. So we didn't complete the same process for the titles that include women's names in their imprints, but we're working on doing that now.
00:06:11	Kate Moffatt (co-host)	Luckily their database is highly searchable. We're actually able to pretty much replicate the process. I can search the imprint field specifically using names of women we know were in the book trades and ran firms and move those titles in accordingly.
00:06:24	Kandice Sharren (co-host)	We also have a lot of genre-specific resources. So, genre-specific bibliographies and checklists have long been a useful tool in literary scholarship. They list all the titles that they can identify in a given genre, thereby producing researchers who are interested in moving beyond the canon with suggestions for fruitful places to look, and some of the most valuable resources that we use fall into this category. Because they have clearly defined parameters, they can often offer more detail than a general resource or be more comprehensive within the bounds that they've chosen.
00:07:01	Kate Moffatt (co-host)	*The English Novel* is a two-volume print bibliography that accounts specifically for novels written in English published between 1770 and 1829. We have 2,468 titles in the database that list *The English Novel* as a source. Along with the ESTC, this was one of the first resources that we used. It has complete bibliographical information for first editions, including volumes, format, pagination, imprints, title, price when available, the author, et cetera.
00:07:29	Kate Moffatt (co-host)	Subsequent editions, however, are not fully represented. It indicates that they exist up to a point and then it will provide very limited data. So it'll tell us how many further editions there were, for example, but not give any more detailed information than

that. Because *The English Novel* is a print resource, it makes it very difficult to correct or to update. So they actually have digital addendums. There is a digital addendum in particular that further includes fiction published between 1830 and 1836. And there have been other updates since the book was published.

00:07:58	Kate Moffatt (co-host)	Because the original resource was in print, we also couldn't automate this process at all, or, copy and paste anything in. We had to type everything ourselves. It's a pretty minor point, but it is worth noting that this creates more opportunities for human error. Not only because we're potentially copying mistakes that were printed in the book, but because we might have made our own typos.
00:08:16	Kandice Sharren (co-host)	We have a couple other genre-specific bibliographies as well. Another major one is the *Jackson Bibliography of Romantic Poetry*, which covers the period between 1770 and 1835. So it's very similar to *The English Novel* in terms of its date range. It also has a genre focus—in this case it is poetry specific. Although it's worth noting that the category of 'poetry' here is a little bit flexible. So a lot of the titles that are included in *The Jackson Bibliography* are actually 'drama.'
00:08:49	Kandice Sharren (co-host)	So things like 'tragedies' written in verse get included in *The Jackson Bibliography of Poetry*. *The Jackson Bibliography* is not nearly as complete in terms of its bibliographical data. So it doesn't always include the full title or the full imprint information, but it does provide the basic information as well as things like date of publication, pagination, et cetera. Instead of format, it actually gives us the size in centimeters, which is an interesting difference.
00:09:17	Kandice Sharren (co-host)	And this one is actually a little bit of an *odd* source because in our database, what we refer to as *The Jackson Bibliography* actually kind of includes complementary digital and print sources. The digital sources have been drawn from these original print sources, but they've been maintained and updated. So as a general rule, we took those as the more reliable records. However, at times the print bibliographies that were published first offered slightly different or more detailed information about things like format that weren't included in the database.
00:09:58	Kandice Sharren (co-host)	So, when we were entering these, we ended up cross-referencing between digital and print records pretty constantly. This was a necessary and helpful exercise in making sure that we had all of the possible information that we could find. However, because these resources are so complementary, we ended up not listing them separately, partly because at an earlier stage in the database, we could only ever list two sources. So, we didn't want both of our sources to be *Jackson*.

00:10:27	Kate Moffatt (co-host)	We also have individual specific resources for the database. An example is Sidney Roscoe's *John Newbery and His Successors*, which we mentioned during our episode on education and children's literature with our previous RA, Reese, and this resource accounts for works published by popular children's publisher, John Newbery, and more interesting to us, Elizabeth Newbery, who was the wife of his nephew Francis, who started his own publishing business, which Elizabeth continued very successfully after her husband's death. Wow. We have 396 titles with this listed as a source.
00:11:00	Kate Moffatt (co-host)	We have also used an MA thesis by Roy Bearden-White on Ann Lemoine, and this has provided us with 213 titles produced by Lemoine, who was a chapbook seller and publisher. We've also made use of Patrick Spedding's bibliography of Eliza Haywood's works, which is particularly interesting because she was an author, but also a bookseller. These resources are great because they're very comprehensive for a single individual, but that can also be a bit of a problem. It starts to sort of *skew* our data. Right now we have almost all the titles for Ann Lemoine but we have very limited information about other chapbook publishers, for example.
00:11:35	Kandice Sharren (co-host)	Library catalogues and the physical spaces of the libraries themselves can also be a really important and useful resource for us. So at one point, actually, we wanted to use *WorldCat*, which pulls together library catalogue information from different libraries across the world as a kind of authority record for our titles. However, we ended up finding that because it is drawing on so many different library catalogues, there wasn't enough consistency across records for this to make sense.
00:12:07	Kandice Sharren (co-host)	So we couldn't always tell when we were looking at a *WorldCat* entry, whether it was a first edition, a second edition, a third edition, especially if it was a title where there were three editions within the space of a certain year. That said, it is a really useful sort of tertiary resource for us. If we're trying to find out where we might find fuller records or fuller bibliographical data for a title, because it indicates which libraries hold the text. We've also used data for specific collections.
00:12:39	Kandice Sharren (co-host)	So the Osborne Collection is a collection contained in the special collections of the Toronto Public Library. And this focuses on children's books specifically. So they actually gave us a bunch of their data to import, which Reese, our former RA, who focused on children's literature, did. And, we use library catalogues such as the catalogues for the British Library and the American Antiquarian Society as fairly major resources. So, as we have mentioned in a couple of episodes now, the American Antiquarian Society is one of the few resources that lets us search by gender.

00:13:21	Kandice Sharren (co-host)	So as we expand and begin to include American titles using these search results and refining them by date is a really exciting opportunity to be comprehensive from the get-go. We can identify as many women who are writing texts as we possibly can, and we can include them in our database. The British Library we use in a slightly different way. So their online records, their online catalogue contains some bibliographical data that we can use.
00:13:42	Kandice Sharren (co-host)	But it's usually partial. So, we have short titles, imprints. It does give us the format and pagination, but the most useful thing that they do is include links to open-access digitized copies for many of the books that they have in their holdings. And this allows us to really look at the books themselves and confirm the data that is being provided in their catalogue records.
00:14:21	Kate Moffatt (co-host)	Because we don't usually consider a title entry verified until we've seen a copy, either the scan digitization online or in person, the sources that provide titles and authors are often then used in conjunction with other sources, like the British Library, that have the digitizations of the works that we're looking at. The *English Short Title Catalogue* is a great example.
00:14:38	Kate Moffatt (co-host)	It provides almost all of the information that we're looking for, but we often use it alongside, say, the British Library or the *Eighteenth Century Collections Online* database because they hold digitizations that allow us to verify for ourselves the information that we're trying to put into the WPHP and to include information that we look for, like colophons, that don't get included in ESTC records, or for that matter, the majority of the resources that we use.
00:15:02	Kandice Sharren (co-host)	*Eighteenth Century Collections Online*, or ECCO, is one of our most frequently used full-text resources. And this one in particular has a lot. It also really usefully allows us to search by ESTC number, which is a wonderful way that our sources speak to each other and allow us to verify things more quickly and easily. We can very quickly determine if a book that is recorded in the ESTC also appears in ECCO.
00:15:45	Kandice Sharren (co-host)	However, like the ESTC, this resource only covers the eighteenth century, but it is fortunately one of a few resources that we use that is a Gale database. Others include the *Nineteenth-Century Collections Online* database, the *Sabin Americana* database, and *Slavery and Anti-slavery* database. And it is now actually possible to search simultaneously across different Gale resources, which allows us to more efficiently determine if one of the Gale collections has a title that we're looking for.

00:16:05	Kandice Sharren (co-host)	Wonderfully, all of their documents also have relatively stable ID numbers, which makes them easy to identify and refine later, if we need to check something else or add additional information. However, all of the Gale resources also require a subscription, which means that they're not available to people without an institutional affiliation or a lot of money [Kate laughs]. So, this is a problem, because ideally we do want to be as transparent about our data as possible.
00:16:36	Kate Moffatt (co-host)	In order to achieve that transparency, we try to include links to open-access resources. And luckily we have been able to find quite a few that do this work for us. It's worth noting that open-access resources are slightly limited because they can only include digitizations that are legally able to be included in an open-access resource. Not all libraries allow their digitizations to be available that way. But we've been able to find a lot.
00:17:00	Kate Moffatt (co-host)	One of our favorites is *HathiTrust Digital Library*, which is by far the most searchable. And it also breaks down the bibliographical information for us. Most of our open-access resources don't do that. It also compiles the different digitizations available into one record, which we *love*, and it tells you which library each digitization comes from as well. We also make use of *Archive.org*, which doesn't break down the bibliographical data for us, but we know how to use the resources, so long as we're given a full-text digitization, we know how to use that in order to find some of the work or data that we're looking for. Like pagination, for example.
00:17:36	Kate Moffatt (co-host)	*Google Books* is another one that we turn to a lot just because there are so many digitizations available. It's similar to *Archive.org* in that it doesn't break down bibliographical data for us, but we can work with that. And while it does have quite a lot of digitizations, it can be really difficult to navigate finding each volume of a multi-volume work, for example, can be really tricky because they have different records for each volume and no connections between them that are easy to find.
00:18:03	Kate Moffatt (co-host)	It's also, I think, important to mention that we only use *Google Book'*s records that actually include full-text digitizations Because not all of their records do. They create records for books that exist, but aren't available digitally or in a digitization form. And if we find a book that we're looking for on *Google* records, but it doesn't have a full-text digitization, then we don't consider it a valid source for the WPHP. We do have to be able to see a title page and see the pages themselves in order to consider it a source.

00:18:31	Kandice Sharren (co-host)	So, in addition to those resources that provide the detailed bibliographical information that we're looking for, as well as the digitized versions that allow us to collect that bibliographical data ourselves. We do use some resources that have extremely limited bibliographical information because they allow us to identify the titles that need to be added in the first place.
00:18:56	Kandice Sharren (co-host)	One of these is *Orlando: Women's Writing in the British Isles from the Beginnings to the Present*, which is a great resource because it's actually focused on women writers and offers a pretty comprehensive list of the women who wrote in literary genres during our period, as well as their books. It is, however, primarily a *biographical* database, not a bibliographical one. So, it's focused on individuals and their lives rather than the details of how their books were produced. This means that we have to supplement the data it provides with our other resources in fairly involved ways.
00:19:33	Kate Moffatt (co-host)	And this brings us to how we grapple with the limitations of our resources, which are inevitable. *Orlando's* a particularly interesting example because while it clearly has limitations for us, given that it doesn't include the bibliographical information that we're actually looking for, it does focus on gender in a way that a lot of our other resources don't. This is a pretty major limitation that we face in the majority of the resources that we use. We have to normally sort through them for the women ourselves. *Orlando* has also allowed us to identify a number of titles after 1800, which is our biggest challenge because...
00:20:08	Kandice Sharren (co-host)	The ESTC ends in 1800!
00:20:11	Kate Moffatt (co-host)	No! [both laugh]
00:20:15	Kandice Sharren (co-host)	Behind-the-scenes look at data entry post 1800! So we often bemoan the state of our post-1800 data because of this. After the ESTC ends, our records become spotty. There's no longer a centralized resource that accounts for all of the editions of a given work that we can easily import into the database. As an example of the kinds of gaps that this leads to, last month, our team developed a spotlight series that focused on some of the gothic works in our database and the authors and publishers who produced them.
00:20:49	Kandice Sharren (co-host)	Our RA, Victoria DeHart, wrote a post about Ann Radcliffe, one of the most important writers writing in the gothic mode. If you want to learn more about her,

we'll link to Victoria's spotlight in the blog post that accompanies this episode, and you can also check out last month's episode "Of Monks and Mountains", where we discuss Ann Radcliffe in a little bit more detail.

00:21:12	Kandice Sharren (co-host)	So it was only because of the work Victoria did for her spotlight that we noticed the WPHP was actually missing editions of Ann Radcliffe's titles from after 1800. Much like Frances Burney's *Evelina* and *Cecilia*, Radcliffe and her titles are very familiar ones, but because those later editions fall between the cracks of what's included in the ESTC, so up to 1800, and *The English Novel* bibliography, which has a focus on first editions, we missed ever entering them in the database in the first place. Sometimes it takes this kind of in-depth author specific work like a spotlight to find out what we just didn't think to look for in the first place.
00:21:59	Kate Moffatt (co-host)	These gaps are as much a result of searching capacities as they are of the limitations of our specific resources. So most of the sources that we use don't allow you to search by gender as we've mentioned. So we can't, for example, search *HathiTrust* or *Google Books* for all titles published between 1700 and 1836 that were written by women. So while there are, I mean, guaranteed a lot of digital copies of books that aren't in the database yet, we can't actually find out about them yet.
00:22:27	Kate Moffatt (co-host)	We need to have a source that tells us they exist in the first place. And we can't use *HathiTrust* or *Google Books* to do that. As I'm finding out, it's sometimes also hard to search by imprint, not all of our resources will allow you to search by imprint. And that limits our ability to find titles that were produced by women in the book trades, rather than just books that were edited or authored by women, which is often more searchable.
00:22:48	Kandice Sharren (co-host)	So how do we handle all of these limitations? Great question [laughs]. A lot of the time the issue is that we don't actually *know* what we're missing until we *find* something that we're missing, which will then allow us to know that we need to search for this specific thing. So now that we know that we're missing all of these later editions of Frances Burney's works, or Ann Radcliffe's works, we know we should probably keep an eye out for other popular titles or other popular authors to see how frequently they were reprinted after 1800.
00:23:24	Kandice Sharren (co-host)	However, if we hadn't noticed that we were missing those in the first place, we wouldn't know to search for them. A lot of the time when we realize there's a gap it's kind of fortuitous or by accident. Some gaps, however, we can see more clearly, for example, because so many of our sources are genre-specific, we have skewed data

towards certain genres. We can address this, though, by intentionally searching out sources that deal with other genres that we haven't covered yet. So for example, we're currently entering a bunch of women's travel writing books into the database, using a new-to-*us* resource about women's travel writing in Great Britain that we found.

00:24:07	Kandice Sharren (co-host)	We've also made connections to other projects and we're using their data and sharing ours so that we can kind of mutually help plug some of the holes that we have. So for example, Kirstyn Leuner has shared her data from the *Stainforth Library of Women's Writing* database, which has allowed us to identify women writers that we were missing. And Catherine Nygren has shared data that she collected about women's travel writing.
00:24:33	Kate Moffatt (co-host)	We're also trying to account more fully for the role that women-run firms played. As Kandice mentioned, we did not do a data ingest for women in the imprints in the ESTC. So that is a gap that has become fairly clear to us, and has become a new task for us to take on. We have many of Ann Lemoine's titles in the database because of Roy Bearden-White's MA thesis, and she's a woman-run firm, but as we've identified, these kinds of sources are few and far between and that can skew our data. So, I am currently using the ESTC instead to search the imprint field using the names of women that we've identified as working in the book trades so that we can include the titles that they produced.
00:25:10	Kandice Sharren (co-host)	That's slow, though.
00:25:11	Kate Moffatt (co-host)	Oh God, yeah [laughs]. I can only do this because the ESTC allows us to search by imprint. This also has its limitations, of course; it's the ESTC, so it only works for pre-1800 titles—sad. And, it also requires these women to sign their names in imprints in a way that's actually identifiable. If someone prints their name as 'A. Smith', it's very difficult for me to figure out which ones were printed by women or produced by women.
00:25:38	Kate Moffatt (co-host)	But, we talk more about this particular issue in our second episode, "Women in the Imprints", if you're interested. But the work that I'm doing, as slow and laborious as it is, does begin to balance the firm's data that currently seems unreasonably skewed towards a handful of women in the book trades, such as Ann Lemoine or Elizabeth Newbery.
00:25:56	♫	[music playing]

00:26:04	Kate Moffatt (co-host)	The print and digital sources that we have access to really do make the WPHP world turn 'round! As indicated by our introductions to these sources, however, as well as by the issues that we've identified with capturing data, for things like post-1800 editions of Frances Burney and Ann Radcliffe's works, the process of finding and inputting data into the WPHP is not a perfect nor straightforward process. Because the majority of our resources are formed around particular people, particular genres, or particular time periods, we inevitably have gaps in our data that reflect the areas our sources don't cover.
00:26:38	🎵	[music playing]
00:26:45	Kandice Sharren (co-host)	So in conclusion, we'd love for the listeners of this podcast to share their favorite post-1800 pre-1836 bibliographical resources with us. No, really we'd love it. We'll send you a Christmas card.
00:26:58	Kate Moffatt (co-host)	We're just kidding.
00:26:59	Kandice Sharren (co-host)	Are we though?
00:27:01	Kate Moffatt (co-host)	Maybe not!
00:27:02	Kandice Sharren (co-host)	[laughs]. But in all honesty, our sources are a source of great excitement and great frustration. The different forms that they take mean that we have to adopt a flexible approach to data collection that accounts for the different searching capacities and scopes of other bibliographies, databases, and projects.
00:27:24	Kandice Sharren (co-host)	Just as we know that our resources likely contain most of the information that would fill our gaps, we also know that first, those gaps need to be identified. As we develop the project further and identify new sources, we must continually develop new searching strategies. And this has made us in turn aware of the potential limitations of our own database's searchability.
00:27:49	Kate Moffatt (co-host)	To supplement our search function, we are currently working on creating a couple of subject guides for the WPHP, including one on how to search for works by people of color and another for how to identify modes of writing not fully captured by our genre tags. By offering meta commentaries on our sources and strategies through these subject guides and podcast episodes like this one, we hope to highlight the

interconnections that form this ecosystem of sources through which we can capture women's book history and their involvement in print.

| 00:28:15 | 🎵 | [music playing] |

00:28:21 Kandice Sharren (co-host)

This has been the sixth episode of *The WPHP Monthly Mercury*. We will be releasing an episode every third Wednesday of the month. If you're interested in learning more about what we discussed today, we've compiled a list of suggestions for further reading and links to some relevant entries in the WPHP in a blog post that you can find at womensprinthistoryproject.com.

| 00:28:45 | 🎵 | [music playing] |

00:28:57 Kate Moffatt (co-host)

[outtakes] The WPHP has more than 60 sources [Hippo meows] listed in its sources table from the comprehensive [Hippo meows], the eighteenth century—

00:29:03 Kandice Sharren (co-host)

[Hippo meows] Hang on—Hippo is...Hippo is having a—she came over to chat...she's been having a nap all afternoon...[laughs].

00:29:16 Kandice Sharren (co-host)

Sorry. She was being really chatty.