

**Fiction and Memoir Inventory for the HBW Corpora:
The Black Book Interactive Project**

White Paper for Award PW-277575-21

History of Black Writing (HBW)

“Black Book Interactive Project III”

National Endowment for the Humanities: Division of Preservation and Access
Humanities Collections and Reference Resources

White Paper for Award HAA-281030-21

History of Black Writing (HBW)

“Building Literacy and Curating (Critical Cultural) Knowledge in
Black Humanities (BLACK DH)”

National Endowment for the Humanities: Office of Digital Humanities
Digital Humanities Advancement Grants

June 2025

Authors: Maryemma Graham, Ayesha K. Hardison, Brendan Williams-Childs,
Onat Kolcu, Mahala Higginbotham, Sarah Arbuthnot Lendt

Contents

- Project Summary 3
- Project Origins and Goals 4
- Project Activities, Team, and Participants..... 6
 - Identification 6
 - Gap-Filling 6
 - Active Identification 7
 - Verification 8
- Project Outcomes 10
 - HBW Corpora 10
 - Access 11
 - PhiloLogic 11
 - Black Lit Network 11
 - Training and Professionalization Opportunities for Students 11
- Project Impact, Evaluation, and Continuation 12

Project Summary

The History of Black Writing (HBW) is a research unit focused on literary recovery, preservation, and the digital humanities. HBW's long-term objectives are to enhance scholarly research and teaching as well as increase public access within the field of African American literary studies, particularly regarding Black-authored fiction and memoir. The project's specific aims are to develop (1) a comprehensive corpus of digitized texts and (2) a metadata schema—a blueprint for collecting descriptive information about the texts—that accounts for the complexities of identity and culture.

HBW's work addresses a critical need to document and analyze Black literature: a body of writing that has been underrepresented historically in literary scholarship as well as within contemporary digital humanities, what the American Libraries Association describes as a “dynamic dialogue between emerging technology and humanistic inquiry.” While building corpora of nearly 7000 fiction texts (encompassing publication dates 1853-2023) and 800 memoirs (encompassing publication dates 1937-2023), HBW has refined its inventory process over the years in terms of identifying texts and verifying their relevance to its corpora. In the future, HBW will complete the digitization and collection of metadata on these texts and build a searchable corpus with accompanying metadata that will support scholarly and teaching efforts.

This work was led by HBW Founding Director Maryemma Graham and her successor, HBW Director Ayesha Hardison.

Project Origins and Goals

Black authors and their contributions have been marginalized in mainstream literary canons, leading to an incomplete understanding of their impact on literary traditions and cultural history. This exclusion becomes notably more pronounced in digitally accessible sources and, as a result, digital scholarship. This lack of representation and availability not only limits scholarly research but also hinders broader access to these works by educators, students, and the public.

The History of Black Writing (HBW) works to bridge these gaps within and between literary studies and digital studies. HBW undertakes this work by identifying underrepresented authors, locating and digitizing texts, and implementing an original metadata schema. This schema curates descriptive information attentive to the texts' unique characteristics to build corpora that are searchable, manageable, and usable.

HBW brings the tradition of Black writing and futuristic thinking together. Viewing a corpus of novels or memoirs at scale with descriptive data will yield new information about the African American literary tradition. Nicole Brown et al addresses the problematic silences perpetuated in data with schemas that do not attend to cultural nuance in their essay "In Search of Zora/When Metadata Isn't Enough: Rescuing the Experiences of Black Women Through Statistical Modeling." Their study demonstrates how scholars might retrieve items buried in large archives.

Comparable to Brown et al's study, HBW recognizes that inventory (the process of identifying and verifying titles by Black authors), digitization, and metadata collection are crucial for ensuring works are discoverable and accessible for scholarly activity and teaching in this technology age. These efforts will allow researchers to locate and analyze materials, providing the basis to engage with individual texts and look for literary trends. It will also allow instructors to convey more complicated, relational histories of the literary tradition that take into account the many ways texts interact with shared cultural values or historical phenomena. Furthermore, HBW's inventory, digitization, and metadata collection supports the integration of these works into larger datasets, such as HathiTrust¹, to enable comparative studies and interdisciplinary research that challenge the existing paradigm. In alignment with the longstanding mission of HBW, this work contributes to the preservation of these texts, helps to ensure that Black literature receives critical attention, enriches the broader field of literary studies, and fosters a general reading public.

Founded in 1983 at the University of Mississippi, HBW was established with the belief that what was being read and taught represented a small percentage of a larger tradition of

¹ HBW's partnership with the HathiTrust Research Center (HTRC) (2020-2023), supported by a grant from the Mellon Foundation, allowed a large portion of the HBW Novel Corpus to serve as a flagship project. This entry into the HTRC ecosystem helped HBW to better understand what can and must be done when sparse, incomplete documentation and metadata limit discoverability within a massive digital collection. Learn more about the outcomes of this partnership here: <https://htrc.github.io/scwared/>

Black writing, and that computer technology could provide more insight and access to that tradition. An initial grant from the National Endowment for the Humanities (NEH) in 1988 allowed the recovery of 800 published novels that were mostly unknown and unaccounted for in the literary tradition. As part of that grant, HBW created a database that recorded bibliographic information about the texts, i.e., publisher and publication date. This effort was HBW’s earliest “metadata” effort before the term became more commonly used among digital humanists. During its development, HBW’s inventory, digitization, and metadata objectives have been funded by internal grants from its former institutional home at the University of Kansas and by external grants, such as the American Council of Learned Societies, the NEH Office of Digital Humanities, and the Mellon Foundation. This history of funding underscores the significance of the HBW Novel Corpus as the first and largest digital archive of its kind with nearly 7000 titles.

The significance of inventory, digitization, and metadata collection extend far beyond mere organization; they are key elements supporting scholarly inquiry and public engagement. By making underrepresented works more accessible, HBW’s inventory, digitization, and metadata collection provides learning and professional opportunities for scholars, including faculty and emerging student-scholars at the graduate and undergraduate levels. In turn, student staff refine their humanities research skills, acquire knowledge in new technologies, and learn to work collaboratively. The inventory and digitization of texts and the creation and standardization of metadata for Black literary works are essential for preserving this vital aspect of literary heritage.

Two different NEH funding sources provided support for much of HBW’s collection building efforts for nearly four years. The following chart shows the two NEH funding sources, dates, and the respective corpora supported:

NEH Funding Source	Funding Dates	HBW Corpus
Humanities Collections and Reference Resources (HCRR)	7/1/2021-6/30/2023	Novel (fiction)
Digital Humanities Advancement Grant (DHAG)	1/1/2022-12/31/2024	Memoir (non-fiction)

The work on the HBW Novel Corpus was a continuation of efforts already in progress while the creation of the HBW Memoir Corpus was an entirely new initiative.

Project Activities, Team, and Participants

HBW has been building its Novel Corpus and formalizing its collection process for decades. The process of identifying titles and determining their appropriateness for the HBW Corpus is a process called *inventory*. Inventory is completed in two stages: *identification* and *verification*. *Identification* refers to the discovery of new titles, whether from authors already in the Corpus or newly added. *Verification* refers to the confirmation of the authors' race and thus their inclusion in the Corpus.

Titles are identified from a variety of sources. After identification, the author's race must be verified, usually by at least three sources, before it is added to the Corpus. All of this identification and verification work is completed by HBW student staff under the direction of HBW Founding Director Maryemma Graham, succeeded by HBW Director Ayesha Hardison.

Student contributors of note during the grant terms (2021-2024) include:

Memoirs

Mahala Higginbotham
Onat Kolcu

Novels

Abisola Akinsiku
Katie Bryan
Jade Harrison
Mahala Higginbotham
Andi Krebs
Brendan Williams-Childs

Identification

There are two tracks for identification. The first is *gap-filling*, which uses the existing Corpus. The second is *active identification*, which uses outside resources.

Gap-Filling

Gap-filling is an entirely manual process. Determining the number of books by Black authors cannot, by default, be a single undertaking. Authors are often prolific, writing multiple novels over the course of their careers, which necessitates constant upkeep and updating of the Corpus. Most authors only write one memoir, but this is not a firm rule, so regular review of the Corpus's author list is necessary. The following locations are utilized to check for titles:

- Author's website, if available

- [The Library of Congress](#) (LOC)
- [WorldCat](#) (a catalog that allows users to search for books in libraries worldwide)
- [Bookshop](#) (an independent bookstore equivalent to Amazon)
- [Thriftbooks](#) (an online organization that sources used books and often sells mass-market paperbacks not found at other locations)
- [Goodreads](#) (a social cataloging website that allows users to search, review, track, and discuss books—used as a starting point before crosschecking with more reliable sources, like LOC)

Active Identification

Active identification is a proactive process. It has both *manual* and *automatic* components.

Manual Identification

Manual identification requires a member of the HBW staff to search for new titles. The question of “where do I find new texts?” is a persistent one, but several resources, many listed below, may help:

- [African American Literature Book Club](#) (AALBC)
- [African American Literature Awards](#)
- [African American Library](#)
- [Mahogany Books](#)
- [African American Review](#)
- [Reparations Club](#)
- [Publisher’s Weekly](#)
- [New York Times](#)
- [Kirkus](#)
- [Amazon](#)
- [African American Book List](#) – a topic-based bibliographic resource guide published by the Detroit Public Library annually

With HBW identifying novel titles for 40+ years, most popular websites for identifying books have been used to their limit. Therefore, it is vital to find smaller, independent sources to find less popular books. To do this, HBW searches the realm of niche bookstores and publishing houses across the U.S. that sell unique titles. The websites for these bookstores and publishers help immensely with finding underrepresented authors of Black writing.

Search Terms:

“Black-owned online bookstores”

“Black-owned bookstores [enter state name]”

“Black women-owned bookstores”
“Christian Black-owned bookstores”
“new Black-owned bookstores”

Memoir and Fiction bookstores/websites to check:

African Book Addict

Black and Bookish

[African American Literature Book Club](#) (AALBC)

[Cultural Front](#)

For memoirs, searching for themes such as sports, romance, cooking, business, politics, poverty, etc. can yield results. Looking up specific topics narrows the search results and focuses on smaller memoir niches. HBW also consulted lists of 20th and early 21st century Black notables and luminaries to identify individuals who may have written memoirs.

Manual identification is not a fast process. To avoid repetition and/or duplication of work, it is important to carefully structure manual identification procedures. Typically, only one HBW staff member at a time adds new titles to the Excel spreadsheet to maintain the Inventory, and there is a clear schedule of when and what resources will be reviewed for adding new titles.

Automatic Identification

Automatic identification is performed in collaboration with a library. A librarian can utilize the library software Gobi, which uses publisher-supplied metadata to identify Black-authored works (both fiction and non-fiction), to produce a spreadsheet that can be checked against the Corpus on a regular basis. Once setup, Gobi updates the Corpus spreadsheet automatically.

Verification

Verification ensures that the HBW Corpora contain only Black authors. This can be a difficult and imperfect process. It is easiest to find authors published by traditional publishing houses within the last fifty years.

Best practices include:

- Using reliable sources during manual identification.
- Verifying the identity of newly identified authors with at least **three** different sources.
 - Example: HBW comes across an unfamiliar book on a Goodreads list. A staff member finds the author’s website. From there, they discover that the author

has been reviewed on AALBC and mentioned in an article in the *Los Angeles Times*. This author would be considered verified.

- Triple-checking the identity of authors published roughly before the 1970s through academic resources, digitized newspapers, and even genealogy databases.
 - These authors routinely have less information available, and so the verification process for them requires different resources. Some examples of these resources include:
 - [JSTOR](#)
 - [EBSCO Information Services](#)
 - [Chronicling America](#)
 - [National Archives, Resources for Genealogists](#)
 - An example: HBW receives a tip via email suggesting works from a deceased author not listed in the Corpus. A simple search turns up a book by the author on a rare-book collector's website; the rare-books site describes it as "a novel by an African American." The publication date allows a staff member to search for the volume more specifically, and they find a review of it on *Kirkus* that additionally confirms the author was Black. The staff member then searches for the novel, complete with author and publication date, and discovers it referenced in several academic papers. This author and title would be considered verified.
- Deferring to the Library of Congress (LOC) and WorldCat as the definitive sources for bibliographic information (except in the case of e-books, which are not reprinted).
- Frequently double-checking the Corpus by using the search function for titles and authors in order to minimize duplicate entries.

Because HBW records many authors who are understudied, there may be little to no information on certain authors and/or texts. Clues are often in the text itself (such as in dedications or the author's statement). Early works, especially understudied ones, can be difficult to verify due to insufficient information. Earlier works that were self-published as well as more recent works, such as digital self-publishing via Amazon, often contain limited information.

Verification is an imperfect process. Occasionally, HBW will discover that an author, who was listed in the Corpus and was historically classified as Black, was, in fact, only writing *about* Black experiences. In cases like this, the author and their titles are removed from the main Corpus. HBW, however, keeps a separate record of this information. Similarly, sometimes HBW learns that a text listed as a novel is actually another genre like a collection of short stories or a memoir. These titles are then recorded elsewhere for potential future projects.

Both the HBW Novel and Memoir Corpora are contained in Excel spreadsheets stored on a private SharePoint, a platform for creating, storing, organizing, sharing, and accessing information. Only HBW administrators and current student staff have access. This ensures the security of the information collected.

Once titles have been identified and authors verified, basic bibliographic metadata is added to the spreadsheet. This includes:

- Author name
- Title
- Publication year
- Source (where the title was found, i.e., Amazon)
- Date the title was added to the corpus

Identification and verification are time-consuming processes, and the required attention to detail can be tedious. Other challenges come with the natural turnover that occurs when employing student workers. This makes it even more necessary to have rigorous procedures that maintain continuity, quality control, and a consistent progress schedule. To meet these challenges, HBW instituted one-on-one training of new hires; prepared for succession through a system of shadowing; formalized inventory processes through manuals, including this white paper; and arranged regular consultations with librarians.

Project Outcomes

HBW Corpora

During the NEH HCRR grant term (2021-2023), HBW added 1353 titles to its Novel Corpus. This brings the total contents of the HBW Novel Corpus to about 7000 titles (encompassing publication dates 1853-2023). Additionally, HBW created a Memoir Corpus by collecting 800 titles² (encompassing publication dates 1937-2023) during the NEH DHAG grant term (2022-2024).

Coinciding with its 40th anniversary (1983-2023), HBW decided to pause the inventory process with texts published through 2023. Alongside identification and verification, HBW has also been diligently digitizing Black texts for about 15 years. This time—and labor—intensive process is a major endeavor that demands a significant amount of funding and staff hours. While a lot has been accomplished over the years (about 4500 of 7000 novels digitized and 550 of 800 memoirs digitized), attention to the other key component of this work—metadata—has been uneven. The decision to pause inventory has allowed HBW to focus more time and resources on metadata collection, ultimately resulting in staff members collecting bibliographic metadata on nearly 5000 fiction titles in about a year. Metadata collection on the Memoir Corpus is planned for a later date.

² To avoid duplicating work done by other scholars, HBW chose to focus on Black memoirs that were not slave narratives for the Memoir Corpus.

Access

PhiloLogic

A portion of the HBW Novel Corpus is accessible within PhiloLogic, an online system that provides search and document retrieval functions for searches within multimedia databases. Access to this defined collection of 2172 full-text digitized novels and accompanying basic metadata represents a collaboration between the Textual Optics Lab at the University of Chicago and HBW. Because many of these novels are still under copyright, the majority are only accessible to Black Book Interactive Project (BBIP) Scholar Program participants and select scholars. The full text of 55 titles from the HBW Novel Corpus that are no longer under copyright are available to the public (https://textual-optics-lab.uchicago.edu/black_writing_corpus).

Black Lit Network

In 2021, with funding from the Mellon Foundation, HBW partnered with Howard Ramsby (Southern Illinois University Edwardsville), Kenton Ramsby (Howard University), and Drew Davidson (University of Kansas) to create the Black Lit Network (<https://www.blacklitnetwork.org/>), a new website initiative designed to advance African American literary studies through digital humanities methodologies. The website enhances access to Black literature by leveraging metadata, computational tools, and public-facing scholarship. More specifically, the project makes ideas about Black novels and archival collections available through a podcast, visualizations, briefs, and interactive search engines. One of these search engines, the Novel Generator Machine, features encoded and structured metadata for over 1500 novels in one of the most comprehensive datasets for Black fiction. The Novel Generator Machine utilizes this metadata to generate personalized book recommendations based on themes, literary styles, and historical context, significantly improving accessibility to Black literary works (<https://www.blacklitnetwork.org/novels>).

Training and Professionalization Opportunities for Students

Work on the HBW Corpora provided training and professionalization opportunities for undergraduate and graduate students across multiple disciplines. Students developed their skills in research, communication, outreach, collaboration, and more. University of Kansas graduate students Onat Kolcu and Brendan Williams-Childs as well as undergraduate student Mahala Higginbotham also contributed to the writing of this white paper.

Project Impact, Evaluation, and Continuation

Even in its in-progress state, the HBW Novel Corpus has had a big impact on the participants of HBW's BBIP Scholars Programs, which began in 2019. To date, more than 60 scholars have passed through six cohorts of either HBW's Introduction to Digital Humanities (DH) Program or Digital Publishing Program. As part of these programs, the participants received training on and full access to the restricted HBW Novel Corpus on PhiloLogic. Many have utilized the Corpus in their DH projects. Additionally, HBW regularly receives requests from scholars seeking access to the 2178 copyright-protected novels within PhiloLogic's restricted section. In response to these requests, HBW carefully vets the scholars' intended use and planned project(s), and successful applicants are granted limited-time access.

HBW's "Black Book Interactive Project III" and "Building Literacy and Curating (Critical Cultural) Knowledge in Black Humanities (BLACK DH)" projects are evaluated based on the achievement of major outcomes: nearly 7000 titles in the Novel Corpus (encompassing publication dates 1853-2023) with 1353 of those added during the NEH HCRR grant term (2021-2023). The Memoir Corpus (encompassing publication dates 1937-2023) of 800 titles was built from scratch with NEH DHAG funding (2022-2024).

HBW moved institutions with its faculty director Ayesha Hardison in 2024. Now housed at the University of Indiana Bloomington, HBW resumed digitization and metadata collection in 2025. Multiple and increasingly complex phases of metadata collection are planned to build a metadata schema that pays particular attention to the distinct cultural, historical, and textual features literary scholarship identifies within Black fiction. HBW plans to increase the accuracy and speed of certain components of this work with coding scripts that recognize various textual features. In the coming years, as HBW nears completing digitization and makes significant progress on its metadata collection, HBW's focus will shift to identifying and securing additional funding to design, build, and support a platform to highlight trends curated from said metadata. This platform will increase the accessibility and usability of the Corpora by educators, students, and the public.