

Scholar-Curated Worksets for Analysis, Reuse & Dissemination (SCWAReD)

Funded 2020-2022 by the Andrew W. Mellon Foundation

The *Scholar-Curated Worksets for Analysis, Reuse & Dissemination* (SCWAReD) project is intended to produce a suite of curated, targeted [HTRC worksets](#) and illustrative, reusable research models (the curated worksets, a scholarly introduction, derived datasets and related documentation, and a research report) that demonstrate the collaborative workset-building, textual analysis, workflow development, and dataset creation activities typically carried out by the Research Center. SCWAReD will result in a number of exemplar worksets and research models related to historically under-resourced and marginalized textual communities that will be developed through a funded round of HTRC's [Advanced Collaborative Support program](#). The goal of these projects will be to explore new methods for creating, analyzing, and reusing curated digital library collections, along with research data derived from these collections. SCWAReD aims to address inequities in both library collections and digital humanities research by identifying gaps within HathiTrust and by using computationally-assisted efforts to recover content that is already part of the HathiTrust Digital Library, but that may be difficult to discover with traditional metadata, in a traditional catalog, from within a massive digital collection.

SCWAReD Flagship Project

The Black Book Interactive Project in HathiTrust

Maryemma Graham, Sarah Arbutnot-Lendt, Jade Harrison, Brendan Williams-Childs (University of Kansas)

HTRC is excited to be partnering with SCWAReD co-PI Dr. Maryemma Graham and her team at the University of Kansas to develop a flagship research model based on the [Project on the History of Black Writing](#) (HBW). HBW was founded in 1983 at the University of Mississippi by Dr. Graham, and has been hosted since 1998 at the University of Kansas. Its principal activities include the creation of the largest known bibliographic database of African-American literary texts in existence; a robust program of summer institutes for professional development and student engagement; community building; and research publishing. HBW's more recent project to create and curate digital full-text versions of works documented in the bibliography are subsumed under the [Black Book Interactive Project](#) (BBIP). The motivating question for BBIP is this: How well is the largest known bibliography of African-American authored texts reflected and represented in the largest known academic digital library? With a combination of advanced metadata matching and methods in textual analysis, we will seek not only to answer this question, but also to remediate some of the gaps that answer will surely reveal, by seeking sources for as many missing texts as possible; developing new ways of incorporating as many of these into the HathiTrust Digital Library as possible; devising new ways to derive data from those texts that cannot be ingested into HathiTrust; and creating SCWAReD worksets and research models from all the diverse sources that we're able to gather.

Projects Selected for SCWAReD-funded Advanced Collaborative Support

Mining the Native American Authored Works in HathiTrust for Insights

Kun Lu, Raina Heaton, and Raymond Orr (University of Oklahoma)

This project seeks to compile a collection of Native American authored works in HathiTrust and apply various text mining methods to the collection to reveal the coverage, subjects, perspectives, and writing styles of Native authors. A list of Native authors and their works will be compiled from an existing database created by a member of the project team and from other online resources. This list will be aligned with the HathiTrust digital library to create a workset of Native American authored works in HathiTrust for further analysis. Then, a variety of text mining methods will be used to analyze the subjects, topics, language use, and writing styles of Native American authors. Comparative analysis will be carried out to understand the characteristics of this textual community. The project is expected to develop a database of Native American authors and the bibliographic information of their works, create a reusable workset of Native American authored works in HathiTrust, identify potential gaps in the HathiTrust corpus on this textual community, and provide insights into the characteristics of the community by text mining their works.

The Black Fantastic: Curated Vocabularies, Artifact Analysis and Identification

Clarissa West-White (Bethune Cookman University) and Seretha Williams (Augusta University)

This project focuses on identifying Black Fantastic texts in the HathiTrust Digital Library. The project proposes that characteristics of the Black Fantastic—the cultural production of African Diasporic artists and creators who engage with the intersections of race and technology in their work—exist in historical and current cultural artifacts, including those created by and about future-forward personalities, such as Dr. Mary McLeod Bethune. It builds on previous and ongoing work to create a bibliography of the Black Fantastic that is featured in [Third Stone Journal](#). Works in HathiTrust will be analyzed along with Black Fantastic artifacts from other collections, such as the Dr. Mary McLeod Bethune collection in the Bethune-Cookman University archives. By working across collections, the project will test methods for locating Black Fantastic texts and lives.

Creating Period-Specific Worksets for Latin American Fiction

José Eduardo González (University of Nebraska, Lincoln)

This project seeks to create large datasets to research the history of Latin American fiction and question traditional periodization of this literature by attempting to detect the boundaries between literary periods and subgenre distinctions in Latin American fiction. It will look critically at the techniques for detecting genre distinctions that have developed over the last few years and evaluate how they apply to the particular development of Latin American literary system. While many of the subgenres in the English-speaking literary market such as detective fiction, the Gothic novel, and speculative fiction have followers in Latin America, the genres that have traditionally been considered important for the changes in the literary history of the region are less formulaic and more closely linked to national and regional historical and/or social developments. Instead of attempting to identify canonical documents that typify a genre, this project will examine how documents diverge from a particular canon in order to explore the social and cultural reasons an author might accept or deviate from a dominant style.

The National Negro Health Digital Project: Recovering and Restoring a Black Public Health Corpus

Kim Gallon (Purdue University)

This project draws on HathiTrust's collection of public health documents on Black health to explore how early twentieth Black public health officials communicated and addressed health disparities that impacted African American communities. The major goal of the project is to create a series of worksets and visualizations that scholars and students of African American health and medicine along with public health experts and physicians can use to deepen historical narratives about Black health that might offer insight into the development of contemporary health communications targeted toward African American communities. The project also establishes some of the research for *Technologies of Recovery: Black DH Theory and Praxis*, a book in-progress. Finally, the work will fill a gap in the history of African American public health.

Get involved

[Call for Proposals](#) for SCWAReD ACS program (**applications now closed**)

[SCWAReD Advanced Collaborative Support Application FAQs](#)

Learn more

[Indiana University press release](#)

[U. of Illinois press release](#)

[U. of Kansas press release](#)

Project Leadership

John Walsh, Ph.D.

John A. Walsh is the Director of the [HathiTrust Research Center](#) and Associate Professor of [Information and Library Science](#) in the [Luddy School of Informatics, Computing, and Engineering](#) at Indiana University. He also has an appointment as an adjunct (affiliate) Associate Professor of English at Indiana University. His research involves the application of computational methods to the study of literary and historical documents. Walsh is an editor on a number of digital scholarly editions, including: the [Petrarch](#) (Co-Editor), the [Algernon Charles Swinburne Project](#) (Editor), and the [Chymistry of Isaac Newton](#) (Technical Editor). He has developed the [Comic Book Markup Language](#) (CBML) for scholarly encoding of comics and graphic novels. Walsh is the creator of [TEI Boilerplate](#), a system for publishing documents encoded according to the Text Encoding Initiative (TEI) [Guidelines for Electronic Text Encoding and Interchange](#). He has also served as the Technical Editor for [Digital Humanities Quarterly](#) (DHQ), the online journal of the [Alliance of Digital Humanities Organizations](#), since the journal's founding in 2007, and as Editor-in-Chief of the [Journal of the Text Encoding Initiative](#) (jTEI) from 2014-2019. Walsh's research interests include: computational literary studies; textual studies and bibliography; text technologies; book history; 19th-century British literature, poetry and poetics; and comic books.

J. Stephen Downie, Ph.D.

Stephen Downie is associate dean for research and a professor at the School of Information Sciences, and Co-Director of the HathiTrust Research Center. He has been an active participant and leader in the digital libraries and digital humanities research domains. At Illinois, Downie leads the HTRC's Research Support Services (RSS) unit, which is responsible for providing the staff and technical support for HTRC's ACS program. Downie was PI for the Mellon-funded WCSA and WCSA + DC projects where he led—and now continues to lead—the development of the HTRC workset and the Extracted Features models and their realizations as production products. Similarly, Downie is responsible for the ongoing development and deployment of both the HTRC Workset Builder 2.0 and HTRC Bookworm tools.

Maryemma Graham, Ph.D.

Dr. Maryemma Graham is University Distinguished Professor in the Department of English at the University of Kansas. In 1983 she founded the Project on the History of Black Writing at the University of Mississippi, which has been hosted since 1998 under her leadership at the University of Kansas. Graham has published 10 books and more than 100 essays, book chapters, and creative works. Graham has been a John Hope Franklin Fellow at the National Humanities Center, an American Council of Learned Societies Fellow, a Ford and Mellon Fellow and has received more than 15 grants from the National Endowment for the Humanities.

This project is supported in part by the Andrew W. Mellon Foundation. Any opinions, findings & conclusions or recommendations expressed here are those of the researchers and do not necessarily reflect the views of the Mellon Foundation.