White Paper

Grant Number HD-248607-16

Black Book Interactive Project

Project Director: Maryemma Graham

University of Kansas

1 December 2018

In "All the Digital Humanists Are White, All the Nerds Are Men, but Some of Us Are Brave," Moya Bailey points to the glaring contradictions in a field that has come to dominate a great deal of our work in the digital humanities. As she argues, in the inaugural issue of the *Journal of Digital Humanities*, "The ways in which identities inform both theory and practice in digital humanities have been largely overlooked." The Project on the History of Black Writing has been cognizant of this contradiction from the onset of our project digital initiative, adopted in 2007. Over the last thirteen years, our work – which not only involves incorporating black writing into DH, but also utilizes DH to allow for a greater representation and consequently, a more holistic examination of black writing - has made substantial progress, making possible our first grant from the Office of Digital Humanities at NEH in 2014. Our guiding motivation has been a need to rethink how we understand African-American literature once it is freed from constricting ideas of 'canonization' and is allowed to include more black writers from the margins within its ranks. As Bailey also points out, such a "move…offers the opportunity to engage new sets of theoretical questions that expose implicit assumptions about what and who counts in digital humanities as well as exposes structural limitations that are the inevitable result of an unexamined identity politics of whiteness, masculinity, and ablebodiness." We share this white paper, therefore, as a public document to contribute to the larger conversation about race and the Digital Humanities. We hope that this model project provides further incentive for new and ongoing initiatives.

The Black Book Interactive Project (BBIP) received funding for the NEH-ODH Grant in 2016 with the primary goal of creating a metadata framework for 75 African American novels. Having identified a serious lacuna in academia that does not account for issues pertaining specifically to race in the digital humanities, BBIP set out to chart categories which would engender new dialogues in academia concerning the same. We sought to achieve this via collaborating with partner Universities, libraries and associations to extend the overall scope, outreach and access to these texts. The benchmarks outlined in our proposal alluded to the following goals: curating other communities to figure out the metadata categories, hiring project managers, planning training sessions, metadata creation for the 75 novels, conducting workshops with students, detailing findings at CLA, and creating our web portal for communication and access.

In the grant narrative, we listed the original benchmarks as follows:

Benchmark I (May-June 2016): Immediately upon notification, we will curate scholarly community for metadata creation, develop project management portal for communication and access; and hire project staff. (1) May, consult with Advanced Research Consortium (Laura Mandell) for development of Resource Descriptive Framework (RDF) datasets representative of the 75 novels.(See Letters of Support, page XX) (2) June, All participants and partners meet in Chicago to set guidelines and timeline for overall project, and for training session for using new Digital Archive in classroom instruction.

Benchmark II (July –December 2016): Orientation and Training for HBW staff and select scholars identified at CLA through online workshop with Erik Radio (IDRD). Complete Metadata Creation in first 35 novels. Identify other suggested venues for information sharing and conference presentations.

Benchmark III (January-March 2017): Metadata creation in final 40 novels. Online workshop for implementing use of digital archive in selected classrooms, using Earhart's model. Continue to develop and modulate partnerships with additional groups.

Benchmark IV (April – September 2017): Compile all information for completion of project and data instruction of 75 novels. Presentation of results at workshop for CLA conference (April). Revise and update web portal with findings. Continue and expand dialogues through online discussions with the following major libraries: Schomburg Library for Research in Black Culture (New York, the Moorland Spingarn Collection (Howard University), Fisk University Library (Nashville, TN), The Woodruff Library (Atlanta University Center), and the Furious Flower Flower Poetry Center (James Madison University, the only print archive dedicated to African American poetry). Included will also be smaller, important collections such as the Margaret Walker Papers (Margaret Walker Research Center, Jackson State University; portions of the which have been digitized); Gwendolyn Brooks Collection (recently obtained by the University of Illinois, Champaign-Urbana), and the Eugene Redmond Collection (Southern Illinois University, Edwardsville). Outline next stage of project for grant proposal.

As intimated earlier in our interim report, we have largely succeeded in meeting the criteria detailed as Benchmarks I and II, aside from a few caveats. First, because of the particular structure of KU, the actual grant work could not begin until August, when a project manager could be hired consistent with currently approved contractual arrangements. We have been in continuous collaboration with the Chicago Text Lab (University of Chicago) who recently finished the digitization and the OCR of nearly 450 African American texts that HBW had identified and provided access to in the form of photocopies. Additionally at KU, The Institute for Digital Research in the Humanities and the Ermal Garinger Academic Resource Center have provided metadata generation support, as well as location and technological support.

With regard to project staff, we confirmed the project manager, Daniel Caine. With the departure of Erik Radio from KU Libraries, who served as our metatada consultant, Erin Wolfe, newly appointed metadata librarian at KU Libraries' Institute for Digital Research in the Humantiies, joined our team as our metadata consultant. Beginning Fall 2016, PhD student Arnab Chakraborty also joined the project as a part time GRA.

We also made a concerted effort to supplement the NEH grant funds in order to continue the work at the University of Chicago and to expand our outreach work as the project was unfolding. Our first success was with University of Kansas Hall Center for the Humanities Seed Grant for $15,000 for the Digital Inclusion Group Initiative (DIGIN), a name that specified the project's purpose. This would enable BBIP to foster the kind of scholarly community discussed in Benchmark I. The DIGIN grant narrative is available upon request; its chief purpose is "to create the first, largest, and most diverse digital corpus of novels after 1923—the current copyright threshold for creative works along with a professionally trained community of scholars engaged in digitial projects associated with the archive."

Before the meeting of the project directors at NEH in the fall, we designed the project portal for BBIP, housed on the HBW website: http://bbip.ku.edu. The BBIP link serves as the promised project management portal for communication and access.

In our October 14, 2016 meeting with our team members and 3 advisors, Kenton Rambsy, Amy Earhart, and Howard Rambsy, in Lawrence, Amy Earhart suggested we opt for generating metadata using an XML framework. A revised timeline was drawn up, which included the following guidelines:

- A mid-January XML Header training workshop led by consultant Amy Earhart.
- A winter-to-early-spring metadata generation for 75 novels, assisted by the members of the January training.
- Another meeting of all consultants and HBW's advisory board at April's CLA conference in Columbia, MO.
- A possible June application for Level II NEH Digital Humanities Grant.

Of these, Goals 1 and 2 have been largely met, as will be detailed below. In addition, to increase access to what we saw as a "hidden collection," we submitted a grant application to the Council of Library and Information Resources Hidden Collection (CLIR). Unfortunately, we did not receive the grant because our suggested activity of OCR, TEI and metadata gathering fell outside the specificiations of the grant. In addition, the number of texts remaining to be digitized that we had indicated did not justify the amount we requested in the budget. We found this feedback especially useful as the project moved forward.

Benchmark III has largely been met, and metadata completion for the 75 novels has been completed. Kenton Rambsy and Will Cunningham had shared their data from earlier stages of the pilot project as a guide to generating further metadata for our 75 books. The process of selecting 75 novels for metadata generation was a collaborative one, and slow response to our queries delayed the process further. Also not anticipated was the challenge of compiling PDFs for the 75 novels to ensure correct mining of information for metadata generation. That process of generating metadata for the 75 categories, while completed, requires further refinements and modifications of the schema, based on ongoing feedback from our project consultants. Before leaving the project, Daniel Caine was in touch with our consultants, and the responders got back to us with advice as to how to proceed. The most important comments led to following categories that have been carefully delineated and negotiated:

1. A sense of place (where action takes place in a novel as well as protagonist's geographical location/origins, and other places referenced in the novel)
2. A sensitivity to tone
3. Splitting the category of author location into birth location and writing location
4. Carefully distinguishing between historical novels and neo-slave narratives
5. Representing various kinds of violence as represented in the novels
6. Charting the reception of the novel, especially with regard to popularity, book reviews and inclusion in bestseller lists

Most of these suggestions have been incorporated, with the exception of 'writing location', owing to the obscure nature and publishing history of some of these texts.

Given our late start and the need to allow for the orientation of a new project manager in fall 2017, we applied for, and received a 3 month extension to the NEH-ODH grant through January 2018 to complete the project in full.

Partnerships are thriving, with expert consultants from Howard University, the University of Richmond, Dillard University, and Texas A&M, Southern Illinois-Edwardsville and the University of Texas-Arlington in addition to IDRH at the University of Kansas Libraries. All of these experts weighed in at all stages of the process, from narrowing the HBW library to 75 texts for metadata generation, to expanding the metadata schema to be more robust and flexible for scholars.

Forming partnerships with all of the libraries and collections listed in the original grant narrative proved quite ambitious, especially since all of them are at different stages on the DH continuum. The Margaret Walker Center, for example, has a digital archive of selected journals by Margaret Walker, but it is not a searchable archive. James Madison University's Furious Flower Poetry Center has digitized 25 years of conferences, poetry readings and interviews, but has not begun the process of providing access. The Eugene Redmond Collection continues to search for funding to digitize the large collections of photographs, books and correspondence by Eugene Redmond now held in the library at Southern Illinois University, but remains a print collection only. One additional partner has been most helpful: the Mellon-funded Publishing without Walls project at the University of Illinois that focuses on African American Materials (AFRO-PWW). This partner helped us to further shape our outreach work and to offer options for digital publications by scholars associated with BBIP. We also strengthened existing partnerships primarily through the previously mentioned metadata schema expansion. Further, in January we hosted an XML workshop and digital humanities roundtable led by Amy Earhart to foster increased connection across the DH spectrum at KU. Our goal was to eliminate the "silo effect" and we made progress towards that. We invited several undergraduate and graduate students, as well as faculty from across KU to participate. Using online resources such as www.tei-c.org (Text Encoding Initiative), the workshop instructedparticipants to encode a piece of text in XML.

Erin Wolfe, the new metadata librarian now associated with the project, assisted us in expanding the metadata schema and ultimately developing a tentative sheet of field descriptions which will better help us refine the metadata schema for XML output.

As planned, we organized a session at the CLA meeting in April, in Columbia, MO, to talk about the project and its findings so far. The decision to make this a joint session by partnering with AFRO-PWW, allowed us to discuss both projects and the status of digital work among younger scholars, while confirming the necessity of giving more attention to outreach, since many who are doing DH work are often isolated on their campuses. We did not meet with our full board at the CLA meeting as planned, due to scheduling conflicts, opting instead for digital communication. However, a brief meeting with two HBW board members, Kenton Rambsy and Ayesha Hardeson, one of whom is on the BBIP advisory team (Rambsy), helped to better focus our next steps for the project. We concluded that we would not be ready to apply for a Level II NEH-ODH grant, and instead to push the project as far as we could, in hopes that other opportunities for funding would arise.

Beginning in July 2017, GRA Arnab Chakraborty took over from Project Manager Daniel Caine and began implementing some of the suggestions received from our board of consultants, as listed at the beginning of the report. Working together with Christopher Peace (PhD student, KU-English) and visiting scholar Lili Wang (English Professor, Harbin Engineering University) , the team extended the initial 15 metadata categories into 51. Having selected a representative set of 75 African American novels from the late 19th to the 20th century, the team finished the metadata for all 51 categories. Some of the specific difficulties the team faced include:

1. Not being able to do close reading of all 75 texts
2. Refining the category of "vernacular"
3. Delineating between genre/theme and style/tone.
4. Ascertaining linguistic features
5. Gauging the popularity of a book in a given decade.
6. Confirming whether the first edition of a given text was in hardback or paperback.
7. Identifying specific rhetorical and linguistic features pertaining to any given text

However, we made considerable progress in extending the nature of the categories. The difficulty lay in striking a balance between subjectivity and fact, with regard to metadata generation. Some of the fields we extended and enhanced include nature and presence of music, violence, supernatural and speculative elements; racial emphasis; profession of protagonists; book reviews; and presence of autobiographical elements. The ultimate intention was not to be an exhaustive source of information, but to point to important informational "signposts" that would facilitate better research. The the final list of the categories follows:

Author Header Schema Categories and Data Values	**Publisher Header Schema Categories and Data Values** (all MLA citation info)	**Book/Textual Header Schema Categories and Data Values**
Gender: (M, F, Gender Fluid)	Publisher Location: (City, State, Country)	Word Count: (number)
Ethnicity: (Self-Identified)	Press Type: (self, commercial, trade, vanity, academic)	Illustrations/Photographs: (yes/no)
Education: (None, self-taught, some high school, high school, college, public, private, HBCU, MA/PhD/MFA)	Date published: (day, month, year)	Name of Illustrator: (word)

Author Pseudonym		Presence of Preface/Introduction: (yes/no)
Nationality: (Self-Identified)		Author of Preface/Introduction: (name of author)
Age: (birth, death)		Presence/Nature of Vernacular: (one exchange in vernacular, referenced, none)
Author location when written (city, state, country)		Presence/Nature of music: (lyrics, instrument, club, singing, sound projection, artist reference, song title)
Author's Birthplace:		Presence/Nature of violence: (fighting, war, murder, kidnapping, riots, bullying, suicide, slave hunt, lynching, shooting, massacres, assassination, rape, military violence, sexual violence, mob violence, racial, communal, accidental death, facial mutilation, psychological, gun violence, domestic violence, alcoholism, industrial accident, police violence, whipping etc.)
Age of Author (at time of publication):		Presence/Nature of speculative/supernatural Elements: (Religious, utopian narrative, ghosts, voodoo, magic realism, space opera, uplift narrative, telepathy, biblical etc.)
Alternative Career of Author:		Tone/Style: (Poetic, satiric, dystopic, comic, naturalistic, realistic, melodramatic, sentimental, epic, omniscient narration, ironic, subversive, third person limited narrative, expressive, descriptive, humorous, social commentary, experimental, impressionistic, noir, terse, point of view narration, mythic, traditional third person narration, sensationalist, biographical, pulp, religious, verbose etc.)

		Genre: (Coming of age, speculative fiction, slave narrative, neo slave narrative, urban fiction, religious fiction, romance, life writing, historical fiction, biblical fiction, allegory, romance, bildungsroman, episodic novel, passing, posthumanist, murder-mystery, thriller, biography, epistolary novel, social critique, drama, divorce fiction, racial tension, political novel, utopian novel, activist literature, tragic mulatto, courtroom noir, crime fiction, southern romance, surrealist fiction, fantasy, drama, interracial relationships, christian piety, racial intrigue, picaresque, pan-Africanism, children's literature, autobiographical fiction, sentimental novel, domestic fiction, anti-pastoral, proletariat novel, military fiction, time travel narrative, gay literature, segregation literature, revolution etc.)
		Location of action in book: (name of city, state)
		When action occurs/ era in novel (date range, unspecified, mythical):
		Presence of Folklore: (folklore character, story, practice)
		Literary Movement/Era: (Slavery, Reconstruction, Segregration, Jim Crow, civil rights, Harlem Renaissance, Black Power Movement, Pre-civil rights, Post-civil rights etc.)
		Narrative Voice: (first person, third person etc.)
		Rhetorical/Linguistic Features:

		Racial Emphasis: (yes/no)
		Profession of Protagonist: (name of profession)
		Protagonist Ethnicity:
		Protagonist Sexual Orientation:
		Class of Protagonist:
		Awards won by Novel: (name and year of award(s) won)
		Awards won by Author: (name and year of award(s) won)
		Title. Year and Genre of Previous/Succeeding Novel by Author:
		Literary Predecessors/Successors: (Name of author/Name of book, and author)
		Book Review: (MLA citation of name, issue and volume of journal, periodical or newspaper in which review was published)
		Pioneering Theme/Contribution to Black Writing:

		Popularity in Decade of Publication: (popular, unpopular, critically acclaimed, critically neglected)
		Rediscovered Novel: (yes/no)
		Serialized Novel: (yes/no)
		Periodical Where Novel was Serialized: (name, issue, volume of periodical)
		Dedication:

Additional Activity

As reported in our interim report, the DIGIN grant from Hall Center, amounting to $15,000, helped us complete the scanning of nearly 816 texts. The Chicago Text Lab successfully completed the OCR of approximately 448 texts on their end. The total number of scanned texts now amount to approximately 1264 scans and 415 OCR'S. Our next prerogative is to transfer the files that we possess to Chicago Text Lab led by Hoyt Long, and finish the OCR of the same.

The Hall Center grant enabled the Text Lab in Chicago to finish a prototype of a searchable user interface based on the scanned and ocr-ed texts. The search interface has not been made public as it is still in development. It is based on Philologic, a data retrieval software toolset developed by the ARTFL Project at the University of Chicago in collaboration with The University of Chicago Library.

Just as we had hoped, we identified an ACLS Digital Extension Grant Program, which would allow us to bring our research and outreach components together in moving to the next level of the project. In January 2018, shortly before the end of the No Cost Extension, we submitted the grant. Notification in April confirmed that we were one of five grants awarded. The ACLS grant for the "Black Book Interactive Project: Extending the Reach has the following goals: 1) "Add additional content to the existing digital archive (approximately 500 novels), complete the conversion to machine readable files and enrich these files with descriptive metadata"; 2) "Adopt and test a user interface to enable greater access and discoverability"; 3) "Provide a series of intensive workshops to make BBIP-ER available to a larger group of scholars, educators, and

students currently outside of the digital community". For this grant we also identified major institutional partners, The College Language Association and the HBCU Library Alliance, while maintaining our relationship with KU Institute for Digital Research in the Humanities, The Chicago Analytics Lab, and AFRO-PWW.

Finally, in terms of promotion, distribution and dissemination, we had an opportunity to present our findings, discuss the BBIP goals, and get feedback from an overflow crowd at the Digital Humanities Seminar sponsored by the Hall Center for the Humanities on April 16th, 2018. We titled our seminar "Embracing the Digital Humanities: Digital Divide or Digital Inclusion" Organized into three components, we gave detailed presentations on BBIP's origins, lineage and legacy; the goals, accomplishments and lessons of the project; and showed examples from the interface. The presentations were made Dr. Maryemma Graham (project director), Arnab Chakraborty (project manager), and Christopher Peace (graduate research assistant).

The funding for the ACLS grant begins in July 2018, which ensures that this project will continue to evolve without any significant interruption.

Scope for Improvement

Throughout the process of collecting and scanning texts, collaborating with our partners and creating metadata, we encountered various logistical and theoretical problems. Negotiating these hurdles has greatly informed the way we presently approach our project. Some of the valuable lessons we learned through our experience in working with the NEH-ODH Grant are as follows:

1. Interdisciplinarity proved to be a must, especially since so much of the information we gathered is informed by a wide range of areas, including linguistics, digital humanities, publishing history, fields that are not easily negotiable by students of literature alone.
2. On the part of the student research assistants, some basic knowledge and training of Digital Humanities and coding might be useful, although not crucial, to the smooth transition and performance of the project.
3. Metadata extraction is not an exact science, is time intensive and dependent on how obscure or popular a text is or might have been.
4. There was no exact method to gathering information for any given text. Aside from a few fixed academic resources that we referred to, each text brought with it its own set of questions and required different research strategies. This is especially true for texts which have long since been out of publication and are in poor condition. Among the many helpful insights that we received at our Hall Center seminar presentation, one proved to be particularly invaluable: the suggestion that we increase our focus on a controlled vocabulary.
5. Close reading and a sensitivity to local and global detail in a text often proved crucial to the process of extracting metadata.
6. Ultimately, much more exposure to digital humanities, its history and its practice, as well as training, are required if we are to maximize the use of our archive.