Peer Review:

Copyright:

Open Access:

Digital Preservation:

OLH
Open Library of Humanities

REMAKING COLLECTION

Out of the Shadows: Bringing African American Digital Collections Together in *Umbra Search African American History*

Cecily Marcus[1] and Sarah Carlson[2]
[1] University of Minnesota, US
[2] University of Texas at Austin, US
Corresponding author: Cecily Marcus (marc0082@umn.edu)

This article explores in four sections the logic and impact of the ways in which all archival collections, but African American collections most poignantly, are incomplete; and how a national search engine for African American history confronts and attempts to address the absence of African American stories, voices, documents, and histories. Following the work of scholars such as Verne Harris, Michelle Caswell, and others, the first section analyzes how and why archives are always necessarily incomplete, as well as the particular reasons behind the bias and erasure of and within African American history and the archives that have come to collect and represent it. The second section discusses how *Umbra Search African American History* (umbrasearch.org) was conceived as a response to the need for a more complete archival record of African American history and culture. Section three presents *Umbra Search* as a case study—what it is, how it has grown, the role of partners, and the challenges it faces. The final section considers the roles of academic and community collections, technology, and collaboration in creating access to a deeper and more fulsome representation of American history and culture.

Introduction

This article explores in four sections the logic and impact of the ways in which all archival collections, and African American collections most poignantly, are incomplete, and how a national search engine for African American history confronts and attempts to address the absence of African American stories, voices, documents, and histories through discovery, digitization, and engagement with audiences within and outside of an academic university context. Following the work of scholars such as Verne Harris, Michelle Caswell, and others, the first section analyzes how and why archives are always necessarily incomplete, as well as the particular reasons behind the bias and erasure of and within African American history and the archives that have come to collect and represent it. The second section discusses how *Umbra Search African American History* (umbrasearch.org) was conceived as a response to the need for a more complete archival record of African American history and culture. Section three presents *Umbra Search* as a case study—what it is, how it has grown, the role of partners, and the challenges it faces. The final section considers the roles of academic and community collections, technology, and collaboration in creating access to a deeper and more fulsome representation of American history and culture.

An Incomplete History of the Incomplete Archive

In the archive, we constantly traffic in collections that are missing, some of which don't exist at all. Of the collections we do hold, there is always something missing *in* them as well. Archives are always and necessarily incomplete, made up only of what was saved or safeguarded, only containing those materials that were not thrown away, the things that were not lost or destroyed, whether purposefully, accidentally, or by the vagaries of time. The incompleteness of archives is a given, rigorously explored, documented, and articulated by archivists, writers, and thinkers. George Orwell (1981 [1946]) wrote: 'When I think of antiquity, the detail that frightens me is that those hundreds of millions of slaves on whose backs civilization rested generation after generation have left behind them no record whatever.' Archivist scholar Randall C. Jimerson (2007: 254), echoing Orwell, writes: 'In looking at the history of archives since ancient times and how they have been used to bolster the

prestige and influence of the powerful elites in societies, I contend that archivists have a moral professional responsibility to balance that support given to the status quo by giving equal voice to those groups that too often have been marginalized and silenced.' Verne Harris (2002: 65), on the astoundingly small amount of history that we are able to document, save, and study, states that:

> The documentary record provides just a sliver of a window into the event. Even if archivists in a particular country were to preserve every record generated throughout the land, they would still have only a sliver of a window into that country's experience. But of course in practice, this record universum is substantially reduced through deliberate and inadvertent destruction by records creators and managers, leaving a sliver of a sliver from which archivists select what they will preserve. And they do not preserve much.

In her article about the founding of the groundbreaking South Asian American Digital Archive (SAADA), archivist educator Michelle Caswell (2016: 27) adopts the concept of symbolic annihilation to describe the absence of materials in American archives that document most aspects of South Asian American histories, stories, and experiences. 'Symbolic annihilation', Caswell writes, is 'a concept first developed by feminist media scholars in the 1970s, [that] describes what happens to members of marginalized groups when they are absent, grossly underrepresented, maligned, or trivialized by mainstream television programming, news outlets, and magazine coverage'. There is important work by archivists and scholars that challenges the idea that archives are more complete than they are representative, that probes the relationship between historical documentation and symbolic annihilation in the archives; and that outlines changes in archival principles and practice that may address the many gaps upon which archives are built.[1] It is not new to acknowledge, as Rodney G. S. Carter writes in his 2006 article 'Of Things Said and Unsaid: Power,

---

[1] See, for example, Cook (2011), work by Michelle Caswell, such as Caswell and Gillibrand (2015), Kate Theimer's Archivesnext blog, Jules (2016), and others.

Archival Silences, and Power in Silence', that archives are spaces of power in which certain voices are heard and others silenced:

> The notion that archives are neutral places with no vested interests has been undermined by current philosophical and theoretical handlings of the concept of the "Archive"; it is now undeniable that archives are spaces of power. Archival power is, in part, the power to allow voices to be heard. It consists of highlighting certain narratives and of including certain types of records created by certain groups. The power of the archive is witnessed in the act of inclusion, but this is only one of its components. The power to exclude is a fundamental aspect of the archive. Inevitably, there are distortions, omissions, erasures, and silences in the archive. Not every story is told. (216)

And yet, the reasons behind why and how archives exclude, silence, and annihilate histories—why we have some collections and not others—are not neutral. They are not equal. Some collections are incomplete merely because they are created by people, and they exist in the world—people are careless or forgetful, orderly in some instances and a mess in others. In the world there are floods or other natural disasters, or the pipes burst, or the boxes were stored too close to the furnace, or some other combination of human folly and an imperfect world. Some collections don't exist because a writer destroyed her drafts and letters, source materials, and notes. Other collections don't exist because it was too dangerous to amass the materials that could betray political convictions and actions. There are more times in history than we can count when private individuals felt compelled to burn their books and destroy evidence of political beliefs, friendship, intellectual kinship. Too many times when offices and hiding places were ransacked or torched by thugs, the police, the military.

Institutional reasons for archival silences and exclusions abound, spanning the seemingly mundane limitation of resources to insidious legacies of white supremacy, discrimination, and racial, sexual, and class prejudice, just to name a few. The founding

and first decades of the Society of American Archivists (SAA) is just one example of how a history of discriminatory practices continues to shape an organization to this day. At its founding in 1936, emerging during the Works Progress Administration when an overabundance of government records needed organization and preservation, the organization was eager to define and professionalize the work of archivists, and did so in part by consolidating 'its strength through its institutionalization', and developed institutional identities associated with local, state, and federal government programs (Brothman, 2011: 422). As such, the scope of archives at public libraries, colleges and universities, and state and federal repositories, often narrowly documented not the diverse constituents these entities represented, but rather the bureaucratic entities themselves. Moreover, in the midst of Jim Crow, institutions with rapidly growing archival and special collections restricted access to African American scholars as 'many tax-supported and philanthropic libraries as well as state and county archives and local historical societies refused service' in segregated reading rooms across the country (Poole, 2014). The Society of American Archivists held more than ten annual meetings—nearly half of the total—in segregated cities between 1937 and 1955 (Poole, 2014: 28). In the 2008 Presidential Address for the SAA, Elizabeth Adkins admitted that diversity had 'been a somewhat uncomfortable topic' for archives as a profession and an institution, citing that government archives split from the SAA in the early 1970s, around the same time a committee on diversity formed to address gender, ethnic, and racial equity (22).

Elisabeth Kaplan's 'We Are What We Keep' soberly reminds us that those who are at the helm of collecting are reflected in the collection. According to Rabia Gibbs (2012), it wasn't until the 1960s that marginalized and 'mainstream' archives begin to converge, with institutional archives actively collecting and documenting things like African American history and the Civil Rights Movement. Moreover, once institutional archives began to collect African American history materials, a tense relationship remained, wherein communities so long barred from the archive regarded 'researchers as poachers and parasites rather than as partners and collaborators', a not uncommon opinion that often creates deep distrust of

institutions (Godfrey, 2016: 166). Communities may thus forgo the potential long-term care that institutional custody can provide, choosing instead to maintain their collections within the communities in which they were created. And then, the pointed absence of such collections in institutions can create the illusion that they don't exist at all, perpetuating cycles of misguided assumptions about what is collected, what should be collected, and by whom.

That which cannot be found in African American collections—whether creators maintain their records in a community outside institutions or because these institutions fail to value and legitimize them in the first place—perpetuates the gaps in our records and understanding of what African American collections are. In African American collections, absence and loss are not only fundamental to the structure and material of the collections, but also to their ethics and scholarly integrity. For more than five centuries, African American history and life have not been valued by most American libraries, archives, and museums, and therefore their records have not been systematically collected. And yet rich collections documenting African American life do exist, even if many of the materials that make up those collections were ill-gotten by institutions that are only now beginning to acknowledge the histories of slavery and oppression on which their libraries and collections are built. Historical materials documenting African American history are scattered all over the country, and yet many of those materials are not labeled in ways that allow us to identify what they are or who they are about.

But even if every object were identified, and righteous, comprehensive, and deep African American collections abounded, there would still be great loss. For every African American book, work of art, letter, or manuscript that is produced, collected, and preserved in a library, there are millions more that were never created at all, their authors enslaved, legally prohibited from learning to read and write, unable to become the doctors, artists, lawyers, carpenters, poets that they could have been. Perhaps materials were created, but they were never able to be saved and passed down through families given the conditions of life for African Americans over more than five centuries in America. Perhaps the heirlooms and treasured books were created and saved in keepsake boxes or in attics, but were never made publicly

available out of deep distrust of the historically white institutions that now traffic in African American collections. As a colleague recently remarked when reflecting on the experience of being a scholar in residence at the University of Minnesota's Givens Collection of African American Literature, these are collections that weren't supposed to exist at all.

When exclusion is understood as a fundamental, productive aspect of the archive, it becomes necessary to ask what has been excluded, why, by whom, where, and for how long. There are archivists who are undertaking projects now that seek to recover missing voices and excluded histories, from Diversifying the Digital to Murkutu, Documenting Ferguson, the Transgender Oral History Project at the Tretter Collection, and many others. At the same time, and as many of the projects listed above openly acknowledge, it is one thing to recover history. It is another to call into question the very forms that archival collections take, both physically and digitally, that serve to obfuscate rather than call attention to the inevitable losses, failures, and absences that characterize any archival collection. To put it another way, it is one thing to add to a collection dates, names, perspectives, events, peoples, and histories that have been excluded in an effort to make it more 'complete'. It is another thing to rethink the form of the collection, the archive, and history itself. As Laura Helton and her colleagues (2015: 1) remind us, we have to countenance the 'the impossibility of recovery when engaged with archives whose very assembly and organization occlude certain historical subjects'.

Developing *Umbra Search African American History*

It is within this context that *Umbra Search African American History* (umbrasearch. org) lives. Working against centuries of loss and erasure, *Umbra Search* is an effort from the University of Minnesota Libraries' Archie Givens, Sr. Collection of African American Literature to provide access to African American history through multiple means: via a free embeddable widget and online search engine (umbrasearch.org) of African American primary source materials from American archives and libraries; by digitizing nearly half a million African American history materials from across University of Minnesota collections; and by supporting students, scholars, artists,

and the public through residencies, workshops, and events around the country that engage African American history, culture, and scholarship (**Figure 1**). If much of the work of *Umbra Search* has been to develop and hone the technological process of aggregating content (most of which comes from collections at academic institutions), identify and assemble the materials, build a discovery platform for users, and bring primary source materials alive for students, scholars, and the public, then at least an equal if not even greater effort has been made to acknowledge, make plain, and try to address the inevitable exclusions and lapses that mark these collections.

Umbra Search seeks to expand the historical and humanistic record by providing unprecedented access to African American history resources, coalescing centuries of African American history within a national digital context. As of June 2018, *Umbra Search* brings together more than 800,000 items documenting African American history and culture from thousands of US archives, libraries, and cultural heritage repositories, from the Smithsonian Institute, the Library of Congress, New York Public Library, Yale University, Payne Theological Seminary, and many, many more. We developed *Umbra Search* as a response to two main principles. The first has to do with the value and impact of African American history, and the belief that that history is central to any understanding of American history—that American history is African American history. This perspective goes against the idea that there is a fixed
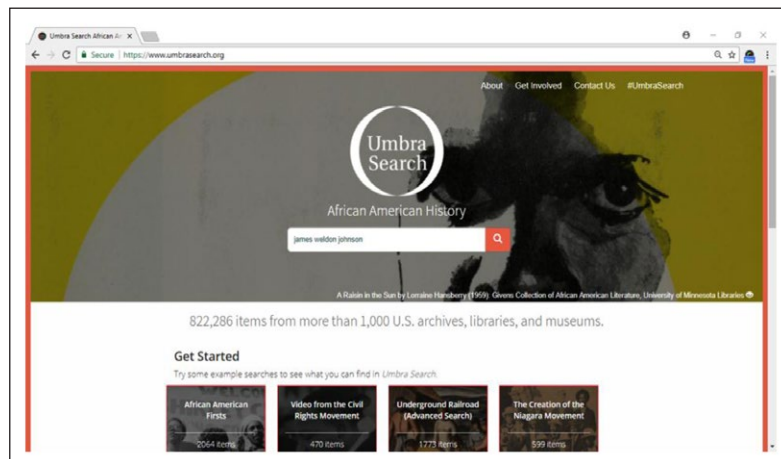

Figure 1: Homepage for *Umbra Search African American History* (umbrasearch.org).

and central historical narrative that implicitly—and at times explicitly—casts white Americans as the actors who make history and relegates the rest (white women, at times, African Americans, Native Americans, LGBTQIA+ people and communities, and many more) to the margins as subordinates who might 'contribute' but cannot create, certainly cannot transform. The insidious fallacy of the 'contribution'—a word that is never attributed to white men—has no place in the logic of *Umbra Search*. African Americans do not contribute to American history. Without their actions and labor (be it forced, free, physical, intellectual, artistic, musical, political), there would be no America, and no American history as we know it.

The second principle is scarcity, and the recognition that much if not most of this American history is not taught in schools, from K-12 to the university, and is not systematically or even adequately collected in our libraries and archives. There is no shared understanding of America's origins or its present, of the history of slavery and its legacies. *Umbra Search* gathers the raw materials of correspondence, manuscripts, notes, ephemera, photographs, and more that go beyond what we find in our history textbooks and that may provide the underpinnings of new works, from History Day projects by middle and high school students, to digital humanities initiatives, to scholarly books, plays, poems, films, dramaturgy, design research, and b-roll.

Where are these archives? For hundreds of years, African American history and culture have largely been left out of centralizing forces of archival collecting and archival principles around collection development, arrangement, and description, leaving us with only a few major African American collections, with the rest scattered across thousands of institutions all over the country and across the world. The Amistad Research Center was established within Fisk University's Race Relations Department to house the historical records of the American Missionary Association (founded in 1846) in 1966; it became affiliated with Tulane University in 1987 after having moved to Dillard University in New Orleans in 1966. What is now the Schomburg Center for Research in Black Culture was founded in 1925 as a special collection of the 135th Street Branch Library in New York City; in 1940 it was formally named the Schomburg Collection of Negro Literature, History and Prints, and was designated part of the New York Public Library with its current name only in 1972. Yale University's James

Weldon Johnson Collection at the Beinecke Rare Book and Manuscript Library was founded in 1941. Howard University's Moorland-Spingarn Research Collection has its origins in the 1914 donation of African American rare books and archives by Jesse E. Moorland, a white minister, YMCA executive, and collector of materials about African American history and culture; in 1946, Arthur B. Spingarn, a Jewish lawyer and NAACP officer, donated his collection of books by Black authors; the Moorland-Spingarn Research Collection was formally named in 1974. The Charles L. Blockson Collection of Afro-American Literature at Temple University was established in 1984 as a result of the donation by historian and bibliophile Dr. Blockson of his vast book collection to the university. The history of Black bibliophiles in America, which goes back to at least the 1830s with the work of writer, publisher, and bookstore owner David Ruggles, and the origins of the African American collections described here, have been movingly if still insufficiently documented by librarians and historians such as Charles Blockson, Dorothy Porter, Arturo Schomburg, Jacqueline Jones, and others. But as Dorothy Berry (2016), 2016–2018 *Umbra Search* Metadata and Digitization Lead, writes, despite the existence of these major collections:

> African American history … is easily perceived as under-collected. While there are online guides to African American archival collections, there is no centralized or authoritative source, especially when it comes to smaller, less researched collections. Stories are spread out across the nation following the trails of academics in Mississippi who collected photos of the Southside of Chicago, music librarians in Durham who collected African American sheet music, and other even more surprising routes.

As libraries and archives have turned to digitization as a means of making collections more accessible to researchers, the availability of African American history materials online has grown dramatically, as have millions of links, sites, online exhibits, and resources that are the results of two decades of investment in digitization and the development of digital collections at libraries across the country. So physical scarcity—bits of histories scattered across many places—also produces digital abundance. *Umbra Search* was developed to facilitate access to this changing digital landscape.

Conversations that led to the development of *Umbra Search* began in 2012, in a very different context and inspired by 'Preserving the Ephemeral: An Archival Program for Theater and the Performing Arts', a project also led by the Givens Collection and the Performing Arts Archives at the University of Minnesota Libraries. As a collaboration between the University of Minnesota Libraries and Penumbra Theatre Company, the largest African American theater in the country, 'Preserving the Ephemeral' assessed the needs of the theater community, and theaters of color in particular, around questions of archives and historical legacy. Over 300 theater representatives responded to a national survey developed in partnership with the American Theatre Archive Project (a then fledgling service organization that pairs archivists with theaters to guide preservation and access efforts) and the Theatre Communications Group (the largest professional service organization for theaters in the United States), resulting in artistic directors and founders from over 60 theaters around the country coming to the University of Minnesota to participate in a two-day forum convened to address the challenges and heightened importance of archives and questions of legacy for culturally specific artists and theaters. It was from these conversations with a group of predominantly African American theater practitioners that we discussed the lack of a common understanding of African American history among theater audiences, and even among the playwrights, costume designers, dramaturgs, and others who are charged with bringing aspects of African American history and culture to life on stage. Where were the primary sources that would inform the set design of a play that takes place in Harlem, NY, in the 1930s? Where would you find the posters that promoted and demonized the Black Panther Party for a play about the Black Arts Movement? In the archives.

Originally conceived as 'The African American Theater History Project', and with funding from the Institute of Museum and Library Services to build a national aggregator for African American history digital archives, the scope of *Umbra Search* immediately went beyond theater and the performing arts. In seeking to bring together the historical artifacts and documents that represent as fully as possible the depth and breadth of African American experiences—its peoples, places, ideas, events, movements, and inspirations—*Umbra Search* was from the start conceived

of as a research tool that would inform and inspire research by students, writers, scholars, dramaturgs, artists, and educators.

These initial forums, from 'Preserving the Ephemeral'; conversations with prospective (and then founding) partners the American Theater Archive Project, the Amistad Research Center, the Apollo Theater, Columbia University, the Digital Public Library of America, the Library of Congress, the New York Public Library for the Performing Arts, the Schomburg Center for Research in Black Culture, the Theatre Library Association, the Theatre Communications Group, and the Smithsonian Institute National Museum of African American History and Culture; and early user-centered design sessions with University of Minnesota faculty and students and Penumbra Theatre staff, led to the development of *Umbra Search*'s widget and online search tool, which began in earnest in early 2014.

If *Umbra Search*'s power lies in its ability to connect and make accessible a vast digital trove of materials that can be used by scholars, students, and the public, then its core challenge goes back to scarcity, form, and the representation of history. How does *Umbra Search* honor the voices that have been lost, the writers who never became writers, doctors, baseball stars, and scientists—people who were never allowed to cross the color line and become who they were, and about whom no articles were written, no awards bestowed, no archives amassed? Does *Umbra Search* challenge the tendency of archives to focus on individuals and events that make history and ignore the everyday lives of individuals, or does it reproduce it? How does *Umbra Search* address the history of American and institutionalized racism that contributes to the many lacunae in our collections, the lack of trust that many African American communities have of the libraries and universities in their areas, and the existence of important, cared-for collections that reside in churches, offices, and private homes, well outside traditional public and academic libraries? How can *Umbra Search* demonstrate and dramatize this scarcity, make plain the reasons for it, and serve as a call to action for the uncovering and inclusion of diverse histories in American libraries, online, and in reading rooms? *Umbra Search* is a product of the shadows of history and of the specter of histories that were never made. We have

attempted to respond to this through a variety of strategies, from the choice of name for the project, to our pursuit of partners, and from our digitization and description efforts, to the programming and education initiatives that engage students, scholars, and the public with African American history.

At every phase of *Umbra Search*, we attempt to call attention to the critical, foundational absences in the corpus we were amassing. The name we took is a part of this effort, with *umbra*—a Latin term that signifies the darkest part of the moon's shadow, its core—a fitting way to gesture towards the marginalization of African American history and archives that *Umbra Search* is trying to address, while also asserting the impossibility of understanding America and American history without understanding African American history. The idea of the shadow can be read as a reference to darkness and the need for illumination, as well as what you see when standing in the sun: your shadow places you in the world, locates you, makes you a part of things. The poetics of *umbra* fit the principles that underlie the work of *Umbra Search*, also tying us to Penumbra Theatre, the country's preeminent African American theater, located in St. Paul, Minnesota, whose archives are part of the Givens Collection of African American Literature, a principal partner of and one of the inspirations behind *Umbra Search*. The name also pays homage to the project's forbears, and evokes the radical poetry collective and Black Arts Movement literary magazine *Umbra*, founded by David Henderson and Calvin C. Hernton in 1963, which featured the work of Ishmael Reed, among many others. We frequently include the publication in classes and during visits to the Givens Collection by students, scholars, and members of the community. The logic of the choice of *umbra* as the name for the aggregator follows Bergis Jules's (2016) apt question:

> If we know that African Americans and other historically victimized and marginalized people in the United States were absolutely essential to building this nation, then why do these silences and erasures continue to exist in our special and distinctive collections, our digital collections, our rare books, our web and social media archives, or our university archives?

Building *Umbra Search*: A Case Study

Throughout its conception and beta phases, and with its public launch in 2017, *Umbra Search* has been embraced by scholars and students (university as well as high school) as a resource for research across the disciplines because of the quality and depth of its content, navigability, and general access. At the same time, though the 800,000+ materials aggregated by *Umbra Search* are rich and surprising (and the number grows with each new data harvest), the corpus is limited by two important factors: 1) *Umbra Search* can only aggregate materials that have been collected and digitized by libraries and other cultural heritage institutions; 2) technologically, *Umbra Search* can only harvest in a sustainable and scalable way digital collections that are easily accessed either through the existing Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) standard or another well-documented and open Application Programming Interface (API). These barriers, though necessary for ensuring a sustainable and scalable search engine, nevertheless limit *Umbra Search*'s corpus to institutions and digital initiatives that have the most financial resources for digitization and infrastructure. As a result, *Umbra Search* passes over many of the institutions whose collections are the most valuable and the most underused, from smaller colleges and universities, many historically Black colleges and universities, and community collections that reside in churches, community centers, or the homes of private individuals. Even the number of materials aggregated by umbrasearch.org is misleading—more than three quarters of a million of anything seems like a lot—and indeed *Umbra Search* encourages the misconception with its homepage highlighting the figure to convey *Umbra Search*'s utility as a centralized portal to many thousands of endpoints, or this essay proclaiming *Umbra Search* as providing unprecedented access to hundreds of years of American history. However, in comparison with the digital aggregation of the Digital Public Library of America (DPLA), which brings together digitized materials on any subject from US libraries, archives, and museums, *Umbra Search*'s 800,000+ objects is just under 3% of DPLA's nearly 22,000,000 materials. As the source of over half of the *Umbra Search* corpus, DPLA only includes 628,254 materials (as of June 2018) that we are able to identify

as documenting some aspect of African American history. The rest of *Umbra Search*'s corpus comes from institutions whose content, for a variety of reasons, is not (or not yet) represented in DPLA, such as Yale University, the Amistad Research Center, and about fifteen others.

In curating a collection of collections, *Umbra Search* strikes a somewhat uneasy balance between relying on and leveraging the methods, practices, and traditions of the archival field on the one hand, and calling them into question on the other. Indeed, *Umbra Search* remakes versions of extant collections first in its initial harvest and then in every subsequent search result delivered to users—a limitation we might first attribute to technical requirements. Technical infrastructure (or lack thereof) alone does not define the boundaries of our attempts to reconceive collections management and access. Innovative work that challenges entrenched practices within libraries, technical and otherwise, requires substantial resources, often requiring grants to fund planning or implementation phases, and then relying on significant institutional commitment to address sustainability needs.[2]

---

[2] Funding for *Umbra Search* comes from several grants, and from the University of Minnesota Libraries. Its planning and implementation phases were funded by the Institute of Museum and Library services: a total of $350,000 over about four years that was used to fund a survey about archives and performing arts organizations, with a focus on African American theaters; national forums with leaders of African American and other theaters on archives, legacy, and history; a half-time project manager that became a full-time position; about 12 months of a developer's time; and graphic design, usability testing, and travel/meetings with partners. Funding for digitization of African American 'hidden' collections comes from the Council on Library and Information Resources (CLIR): nearly $225,000 for a 24-month project, with a full-time project archivist and digitization/metadata lead; many thousands of hours of students' time spent digitizing materials; a small amount of outsourced digitization for audio-visual materials; and some supplies. Community engagement and dissemination work is funded by the Doris Duke Charitable Foundation: $168,000 over the course of 2.5 years that covers some Project Manager time, travel, outreach activities, support for the *Umbra Search* Advisory Council, some *Umbra Search* collateral (stickers, bookmarks, traveling exhibit panels), etc. These figures do not represent the significant cost-share, sometimes as high as 100%, provided by the University of Minnesota: the time of multiple staff, sometimes as much as 50% of a staff member's appointment. and including directors, curators, catalogers, metadata librarians, designers, communication staff, event planning staff, and many more; frequent flier miles; discretionary funds from Libraries administrators; indirect costs (office space, phones, Internet, heat, air conditioning, etc.), and more, all of which are factored into the budget and tracked throughout the grant terms.

A Front End

Throughout every stage of development, we sought feedback and input from potential users. A group of early project stakeholders and partners, including theater directors and dramaturgs, scholars, librarians and archivists, and community members participated in a wireframe workshop to articulate and refine the user interface for early front-end development on top of the Blacklight discovery platform. When the beta version of *Umbra Search* debuted online in 2015, we released an ongoing user experience survey, along with a series of three direct outreach efforts to librarians and archivists, educators and scholars, and artists and creative professionals for feedback.[3] In both our targeted feedback campaigns and in the ongoing form that collected responses over eighteen months, users often expressed what they would like to see that *Umbra Search* did not feature, whether in function or content.

   As we aggregated more materials from increasingly diverse collections, and the corpus of *Umbra Search* materials grew, user needs changed. What had initially been a small collection that was relatively easy to control and manipulate rapidly expanded with each new ingest. Links might break, thumbnails might be too large or too small, metadata might not exist or might be too excessive to be helpful. Over 18 months of development, we also conducted three usability testing sessions with the University of Minnesota Libraries' Web Presence Management Group (WPMG). By posting fliers throughout the main library on the University of Minnesota campus, WPMG recruited volunteers (mostly undergraduate and graduate students) to participate in 30-minute testing sessions during which users navigated through the *Umbra Search* site, following the prompts of a facilitator, while the rest of the WPMG and *Umbra Search* team observed virtually. Whereas wireframe workshops and ongoing beta tests often informed our team of what users would like to see, in-person usability testing

---

   [3] Early beta test results (planning phase) can be viewed at: drive.google.com/file/d/0B4DlkgKyZjVPb3VRWDRTckJNekU/view?usp=sharing. Results from early beta site user experience feedback can be viewed at: drive.google.com/file/d/0B4DlkgKyZjVPbEdPZVV5RzNYa0U/view?usp=sharing. Beta test on social media use can be viewed at: drive.google.com/file/d/0B7_lwiOQGlbOQXAyQll2VjViZlE/view?usp=sharing. Pre-launch user experience beta test results can be viewed at: drive.google.com/file/d/0B7_lwiOQGlbOYUFhSW0ySDBwNU0/view?usp=sharing.

clearly demonstrated what users actually used and needed, what was distracting or altogether ignored, and what features worked well. For example, while users did not frequently use faceting/refining featured on an early version of the homepage search, an autosuggest search feature that culled keywords from the corpus of materials in *Umbra Search* helped users understand what they might find, and how to search efficiently. Once past the homepage, users utilized faceting to refine their broad search entered on the homepage. While technology largely determined what content was included in *Umbra Search*, user feedback directly informed how it was presented.

In its design and architecture, *Umbra Search* emphasizes the possibility for finding more and more relevant hits through suggestions of additional keywords, related materials, and the like. It does not point toward what could be included but is not. The user interface and experience of *Umbra Search* promote a 'successful search': the homepage, with a large search bar, announces a corpus of hundreds of thousands of materials; canned searches for various material types and subjects; and featured content that highlights blog posts, digital essays, exhibits, and more that incorporate materials found using *Umbra Search*. Faceting features, now ubiquitous in library databases and digital collections and built into Blacklight's front end, allow researchers to limit search results, taking too many hits to a more usable and useful few. Again, the idea of abundance is programmed into the user interface itself. When a search query fails to return any results, *Umbra Search*'s message states: 'No results found for your search. Try modifying your search (Use fewer keywords to start, then refine your search using the links on the left)', suggesting that the lack of materials found is the result of the word choices of the user. *Umbra Search*'s message does not say, '*Umbra Search* does not include the materials you searched for'. There are many reasons, however, that may explain such absence:

a. It's possible that no library, archive, or museum has collected materials in this area.

b. It's possible that relevant materials have been collected, but they are cataloged and described under terms that don't allow *Umbra Search* to

recognize their relevance to African American history and culture.

c. It's possible that relevant materials have been collected and are adequately described, but have not been digitized.

d. It's possible that an institution has collected, adequately described, and digitized relevant materials, but *Umbra Search* is not currently including them for a variety of reasons (lack of awareness, lack of inclusion in search strategy, lack of resources, lack of infrastructure, etc.)

e. All or some of the above, plus more…

Behind Umbrasearch.org

Digital aggregations have been the work of libraries long before the Digital Public Library of America or *Umbra Search* debuted online.[4] From a technical perspective, *Umbra Search* can be viewed as a digital libraries project that consumes and then provides a platform for discovery of extant data and that attempts to be nimble enough to a) account for disparate digital collections practices and digital repository platforms (such as contentDM, Islandora, and others); b) anticipate appropriate descriptive terms in the search strategy; and c) represent dramatically different collections in a single usable interface. Wanting to serve as a model for other thematic digital aggregations, *Umbra Search* was developed openly, and all project documentation is available online at github (github.com/UMNLibraries/umbra.search), in the hope that others may use and improve upon it. The *Umbra Search* website was built by customizing an implementation of Blacklight, an open source discovery platform, as well as by using and extending the University of Minnesota ETL (Extract, Transform, Load) Hub open source platform (**Figure 2**) that was developed for the purpose of serving as a partner hub for the DPLA. The ELT Hub system harvests metadata records from digital repositories and from known large-scale aggregations such as the DPLA, HathiTrust, and the Internet Archive.[5]

---

[4] A particularly helpful guide for navigating digital collections work done thus far, and for practices moving forward, is *DH Curation Guide: A Community Resource Guide to Data Curation in the digital Humanities*, available at: http://guide.dhcuration.org/contents/digital-collections-and-aggregations/.

[5] The Open Archive Initiative Protocol for Harvesting Metadata is a framework for repository
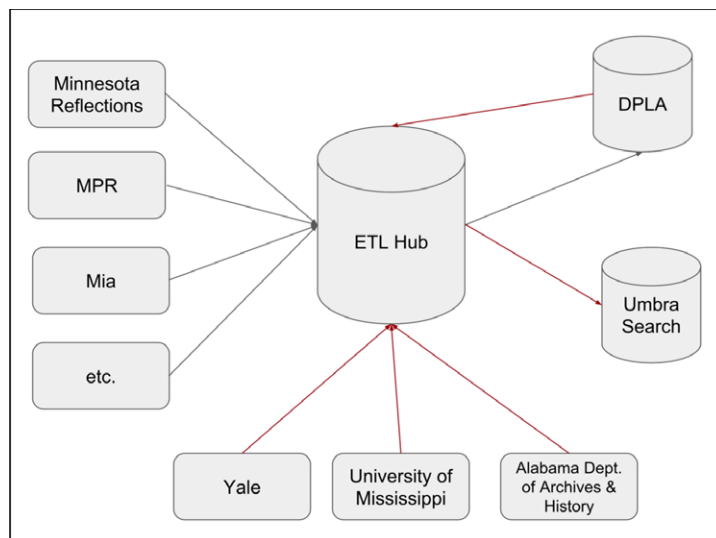
Figure 2: Example of *Umbra Search* metadata aggregation flow.

Content is harvested by comparing a curated list of keywords (the search strategy, or, as *Umbra Search* developer Chad Fennell calls it, 'a big bag of words') against individual records and then including those records that match any terms within the list.[6] Matching records are then normalized in numerous ways: words are formatted to the same case; metadata is enriched with additional keyword terms; and metadata is mapped to the *Umbra Search* metadata schema, which was derived from the DPLA metadata schema. Records are saved within a database and are accessible via the ETL Hub public API. The ETL Hub processes many millions of records; the custom Blacklight implementation then consumes and stores these records in an Apache Solr search index, resulting in the content ultimately found within the *Umbra Search* interface. Not all of these records, however, are accessible to users on the front end. As with any automated aggregation, false positives are inevitable.

The challenge of identifying appropriate materials is primarily due to non-standard and other archival description practices that do not map well to large-scale

---

interoperability. See: https://www.openarchives.org/pmh/.

[6] https://github.com/UMNLibraries/ETLHub.profiles/blob/master/templates/umbra_term_matchers. json. What our developer calls 'a big bag of words', this list of terms constitutes the search strategy, which runs against millions of records to match with content that fits the scope of *Umbra Search*.

aggregation and, as a result, can become lost in aggregation, so to speak. Insufficient descriptions that fail to identify race; a lack of archivists of color; a lack of knowledge on the part of predominantly white archivists working with culturally specific materials; the separation of collections in libraries from the communities that created them and libraries' failure to work or establish trust with these communities; and common arrangement protocols such as 'More Product Less Process' that result in gross box- and folder-level description (a practice that has deep, negative implications for mass digitization initiatives as well as digital aggregation and discovery, which we discuss more fully below) can all account for many of the limitations around metadata that have an impact on an any algorithm being able to systematically identify and cull materials related to African American history. Given all this, we implemented several metadata remediation methods—both manual and algorithmic, that attempted to make the most of the materials we could identify and include in *Umbra Search*.

Remediating Metadata and the Ethics of Naming

In order to increase our chances of successfully and efficiently identifying metadata relevant to *Umbra Search*, we leveraged Apache Solr to create an indexing mechanism and custom search interface for metadata records. *Umbra Search* partners and content providers are also asked to declare their metadata records copyright-free under a Creative Commons (CC0) license, thus granting *Umbra Search* site administrators limited control over the indexing process via a public Graphical User Interface (GUI). This level of access allows *Umbra Search* to normalize, de-duplicate, and publish the metadata records that we have stored in a local database. It also allows site administrators to hide individual records from end users (records that represent false-positive matches within the keyword matching process—such as records about 'bird migration' rather than the 'Great Migration', for example) and to manually enrich records with additional keywords (**Figure 5**). As a general rule, additional keywords followed thematic categories that would include materials in broad searches for common topics, including the Black Arts Movement, Civil Rights, Diaspora, Politics and Government, Music and Theater, and more.

With the capacity for metadata enhancements, we took the opportunity to add as best we could important contextual information that either wasn't there, or that

had been lost in the process of separating individual digital objects from the rest of the materials in the collection from which they came. An important example of the need to add metadata terms, when possible, is in cases when materials represented white supremacist and racist ideologies, practices, groups, etc. Considering that users of *Umbra Search* would range from advanced scholars to middle school students, to include materials created out of prejudice and racism without explicit metadata describing them as such would maintain the illusion that metadata and archives are neutral and go against the very ethics of *Umbra Search*. Some materials may be found deeply offensive (e.g. a racist tract by the White Citizens' Council) or inappropriately out of scope (e.g. an abundance of confederate soldier letters from the Civil War), but rather than hide them, keywords serve to contextualize their existence in the *Umbra Search* corpus. The need to identify racist materials was made abundantly clear from the test search of the *Umbra Search* dataset: The very first record shown (**Figure 3**) in a non-specific search inquiry (just press the search/magnifying glass button) is a pro-segregation newspaper article clipping from a white supremacist group, Mississippi's Citizen's Council. When viewed in the digital collection from the University of Mississippi Libraries Digital Collections (**Figure 4**), this critical contextual information is made clear. In *Umbra Search* search results, less so.

In addition to enhancing metadata, we also built a back-end tool that allows site editors (project staff and students) to 'vote' on whether records are 'appropriate' for *Umbra Search*'s corpus.
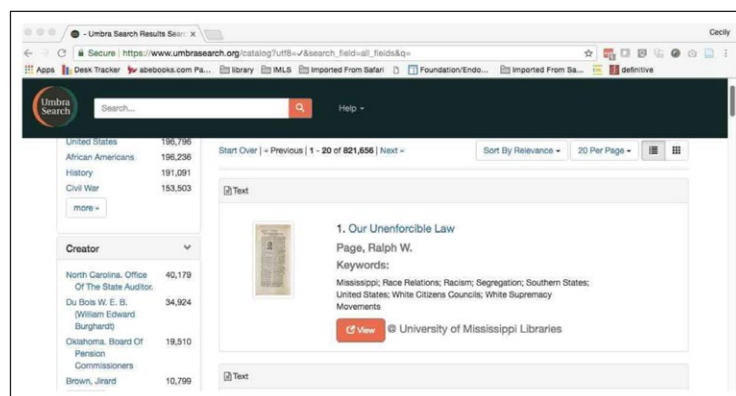

Figure 3: First *Umbra Search* search result. https://www.umbrasearch.org/catalog?utf8=%E2%9C%93&search_field=all_fields&q=.
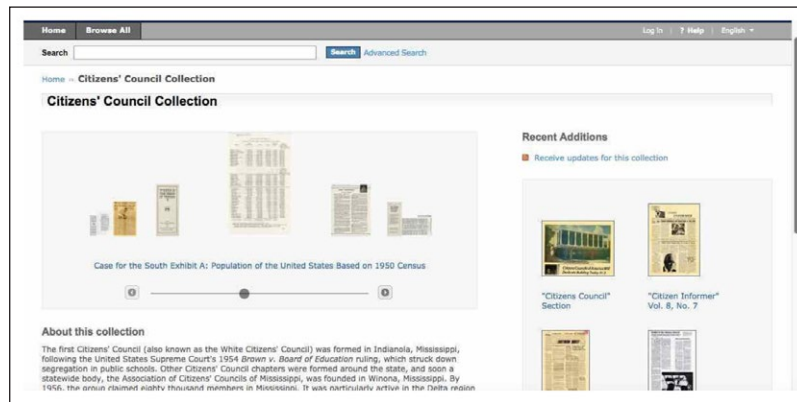
Figure 4: University of Mississippi Libraries Digital Collections. http://clio.lib.ole-miss.edu/cdm/landingpage/collection/citizens.
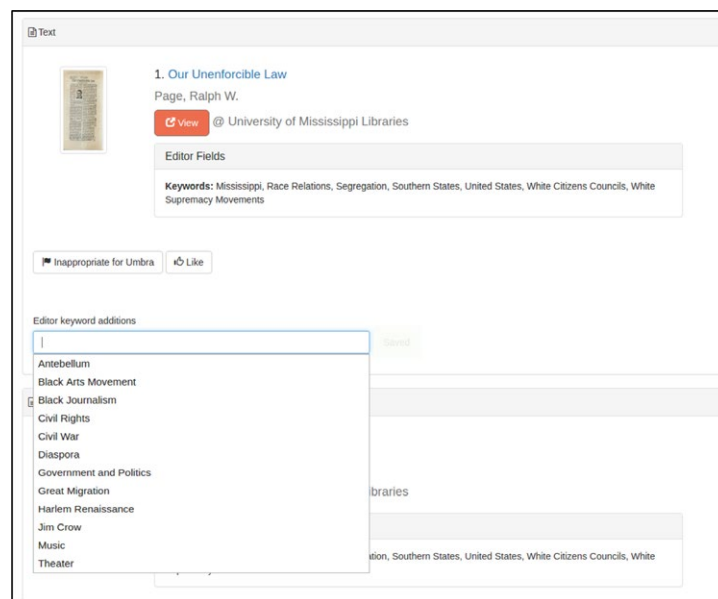

Figure 5: View of editor metadata enhancements.

This mechanism allows us to seed the corpora with a 'ham/spam' system and begin initial experimentation with algorithmically identifying records that are likely to be relevant to *Umbra Search* and those that are not.[7]

---

[7] Initial attempts at automatic classification include 'Hamster' (https://github.com/UMNLibraries/hamster) and 'Gerbil' (https://github.com/UMNLibraries/gerbil), developed by Chad Fennell.

Currently, metadata enhancement and remediation features are not available to end users for a number of reasons that relate to technical and staff capacity. As an openly developed project, we considered the potential of crowdsourcing to assist in tool learning experimentation for record-level and search strategy enhancement. However, we deemed that the technical investment and debt required for this kind of user participation (developing a comment feature, or requiring the creation of individual user accounts, or some other mechanism) was too high to warrant a relatively experimental feature, given our limited budget. Moreover, allowing all end users to rank results for relevance and suggest keywords for records also requires close monitoring to ensure only appropriate and correct suggestions would be used to enhance records and search results. We could guarantee neither the technological nor personnel capacity for crowdsourcing, even with substantial grant funding for the initial tool development, and especially after grant funds were depleted.

Instead, editors manually modify and enhance metadata for the records in *Umbra Search*, and algorithms allow us to identify additional content more effectively. The *Umbra Search* strategy—now over 1,000 terms—first included a non-systematic, non-scientific list of names of individuals, organizations, significant historical events, and places, as well as specific collection or institutional names, and more.[8] While the list continues to grow as we consult partners, archivists, scholars, and end users, it is by no means authoritative or exhaustive and it continues to thwart our attempts to design a more systematic method for building an effective search strategy. The search strategy remains, despite numerous experiments and tests, an art rather than a science.

An example of the limitations of the *Umbra Search* search strategy is a 2016 experiment with partner Howard University and its Portal to the Black Experience project that leverages Howard University's Founders Library's historical cataloging practices of identifying African American authors and authors of African descent. Our partners at Howard shared with us a list of more than 6,000 names culled from

---

[8] The search strategy, along with other technical documentation, is openly available on github (http://github.com/UMNLibraries/umbra.search).

the Moorland-Spingarn card catalog, which has now been digitized, and we added the names to our list. This significant addition, however, yielded small returns in identifying content. Many of the names were not widely known, and, as a result, their works are not well represented in archives and rare book collections, or not identified in the process or arrangement and description if they were there at all.

The experiment demonstrates that commonly used keywords and terms, rather than the number of terms we include in a search strategy, results in a more successful match rate in aggregation and search tools. It shows us how *Umbra Search*'s automated ingestion process privileges content by or about well-known people and events or well-resourced collections and institutions. *Umbra Search* is shaped by the biases, oversights, and archival silences that we find in the thousands of collections that *Umbra Search* makes more broadly accessible. This conclusion, while sobering, had a deep and productive impact on how we continued to shape *Umbra Search* beyond the search tool and widget.

Finding Umbra Search Content

At the start of the *Umbra Search* project, our goal was to build a search widget—a piece of code that can be embedded on any website—that brings together digitized African American materials that were already available, albeit buried in Google searches and dispersed across thousands of institutions with no good way to find related materials even if you made your way to a digital collection in the first place. We assumed, quite wrongly, that given that the materials were already openly available online, we could aggregate and display them without needing to acquire permission from every source institution. Rather, formal and legally binding agreements were necessary. The process of developing and then securing data sharing agreements with institutions was time consuming and labor intensive; that same process became—and remains—a critical part of *Umbra Search*'s institutional success and a driving force behind its capacity for growth, outreach, and engagement.

We were already aware in 2013 that a substantial amount of *Umbra Search* content would itself be aggregated from the massive aggregation that is the Digital Public Library of America, which launched on April 18, 2013, after we had

submitted a grant proposal for developing the search tool. At that point, hundreds of institutions, from large research universities to local historical societies, were being gathered together to form the DPLA corpus and could be leveraged for *Umbra Search*, a search tool that would focus results for users and provide deep access to materials that might otherwise be missed in a DPLA search. Many institutions with important collections, however, were not, or not yet, participating in the DPLA initiative, such as Yale University's Beinecke Library, home of the James Weldon Johnson collection, as well as Temple University's Charles Blockson Collection, the Amistad Research Center, the University of Massachusetts Amherst, whose phenomenal and fully digitized W. E. B. DuBois collection contributes more than 100,000 records to *Umbra Search*—the highest number of any institution—and many more. For these institutions, we needed a formal, legally binding agreement that would allow *Umbra Search* to aggregate and make openly available metadata and accompanying thumbnail images. In consultation with University of Minnesota Libraries copyright librarian/lawyer and the University's Office of General Council, we developed a data sharing Partnership Agreement that outlines the terms for participating in *Umbra Search*, including a declaration of a Creative Commons Universal Public Domain Dedication License (CC0 1.0) for metadata, copyright, and the sharing of thumbnail images that would enable *Umbra Search* to aggregate and make available for any transformative use a given institution's digital collections materials. A staff of two— the Principal Investigator, with 10% of her time officially on the project; and a .5FTE project manager—wrote thousands of emails and made as many phone calls in order to introduce the project, negotiate the terms, and secure agreements. The process usually began with outreach to a collection curator, archivist, or digital collections librarian, which then led to conversations with directors of Archives and Special Collections, University Librarians, deans, provosts. We pursued these relationships avidly, and were able to gain important, indeed transformational institutional buy-in as a result. For institutions already participating in the DPLA, we developed and shared a Memorandum of Understanding that notified them of how their already declared CC 0 1.0 records were being used and outlined mutual aspirations to make their African American history materials as accessible as possible. While not legally

binding or technically required, the memorandum also necessitated substantive conversations with multiple representatives across libraries.

Developing these partnerships not only helped us secure the documentation that allowed us to build *Umbra Search*, but also shaped both its content and its curatorial approach. Curators and archivists who knew their collections intimately were able to reveal collections that an ever-changing search strategy alone would have missed, alerting us to collections that were in the digitization queue, and consulting with us about how to make sure the content would be identifiable by *Umbra Search* in order to assure inclusion. Just as impactful was the participation of curators, archivists, and others in the testing and refinement of the *Umbra Search* interface, in the development of #UmbraSearch digital exhibits that featured their collection, and in *Umbra Search* events in their areas that highlighted their collections or shared them with the general public. The Memorandum of Understanding, in particular, served as a tool for establishing relationships because it was about more than just content. It was about collaboration: from sharing *Umbra Search* and its embeddable widget; to a given institution working with its faculty and students to test and use *Umbra Search*; to co-sponsoring public forums, roundtables, and other events that showcase local collections, regardless of what may be available online and according to what a given partner deemed appropriate and feasible. We prioritized contact with institutions not only with premiere African American collections, but also those with rich collections but which may lack (or lacked) the resources or infrastructure to share digital content in an automated way, and with organizations that don't have digital content but whose missions complement *Umbra Search*'s, allowing collaboration to mean much more than merely sharing metadata and thumbnail images.

When Partners Partner

With a robust roster of partner institutions with which we had secured signed agreements, and even with some with which we had not but had nevertheless formed collaborative relationships, *Umbra Search* was extremely well positioned to make the most of the opportunity to build on synergistic content and missions by engaging in substantive outreach efforts with partners all over the country. We designed a robust two-year community engagement and outreach plan that would allow *Umbra Search*

and its partners to promote and share the search tool, the collections on which it was built, and all types of work engaging with archives, African American history and culture, and more. Support from the Doris Duke Charitable Foundation allowed us to think broadly about archives—the materials that can be the most rarefied and intimidating, the least accessible, hardest to find, hard to visit, locked in storage vaults, and accessible only with appointments in reading rooms by those who brave multiple barriers (distance, cost to travel, the need to think ahead, come to campus, find parking, find the reading room, store pens and bags and coats in lockers, register and provide a form of identification, talk with curators, and more)—and what it means to take archives out of the archives. What does it mean to make archival materials and primary sources more discoverable online? What does it mean to engage students, educators, and scholars intentionally and over a sustained period with *Umbra Search*? What does it mean for an archive to reach beyond the core audiences of an academic research library to work with artists, activists, K-12 teachers and students, coders, and more?

For more than two years *Umbra Search* dedicated itself to exploring these questions by presenting or organizing panels at conferences, but also partnering with instructors and faculty to use *Umbra Search* in their classes and have their students use it for digital projects; collaborating with National History Day and its local programming in Minnesota to reach K-12 students and educators; designing teacher training about how to use digital primary resources in the classroom; creating #UmbraSearch, a blog to feature these projects and digital exhibits by guest artists, curators, and archivists; sponsoring artist residencies; launching #UmbraSearch365, a Twitter campaign that pushed out *Umbra Search* content every day during our launch year, 2017; and sponsoring events locally and around the country with *Umbra Search* partners and users.

While we promoted *Umbra Search* nationally as a tool that could be used to access collections from anywhere, local educational partnerships also provided the important opportunity to meet students and teachers where they already were: in the classroom. Students from Gordon Parks High School in St. Paul, MN, visited the Givens Collection of African American Literature to experience working with physical

collections materials. Identifying an object of interest from the Givens Collection, students then went online, using *Umbra Search* to broaden their research, find contextual and supporting materials, and learn how to navigate different collections. We developed this exercise and created a learning experience with the College of Liberal Arts at the University of Minnesota to introduce undergraduate students to archival research, which was presented in a series of skills for academic success. With both high school and undergraduate students, the fundamental objective was to develop information literacy skills, namely being able to identify and analyze a primary source. These activities successfully guided students through working with primary source materials, and also demonstrated a perhaps more critical information literacy need: differentiating types of online search tools. The traditional information literacy model created by librarians was educationally fundamental in our sessions, and yet as archivist Peter Carini (2009: 47) notes, it 'misses one of the most important concepts that students must understand when using primary sources: historical context', which across online search tools is often further muddled.

As a response, *Umbra Search* Education and Outreach lead Jennifer Hootman developed and piloted a straightforward and highly effective exercise that walks students through the difference between Google and *Umbra Search* that has been piloted in multiple University of Minnesota undergraduate courses from a range of disciplines (History, African American Studies, Communications, English). A two-page handout, 'Discovering Digitized Primary Sources Google Search & Umbra Search' (**Figure 6**) provides an overview and brief discussion points about the difference between primary and secondary sources and the Internet.

Students are asked to work in groups of two to three and complete a series of questions based on the use of the same keywords in searches in Google and in *Umbra Search*—compare and contrast how many results a search contained, sources of search results (Wikipedia, history.com, etc.), and a group of questions about a single hit (author, type of source (news article, blog, etc.), intended audience), and more. In a UMN History course about the 1960s, searches on 'freedom riders' and 'Fred Hampton', for example, produced clear differences that were immediately understandable by the students: Wikipedia entries vs. FBI reports; recent articles

from Time Magazine vs. hand-drawn Black Panther propaganda posters; BlackPast.
org summaries vs. a 1961 letter from United States senator Albert Gore, Sr. Multiple
viewpoints (including racist tracts and white supremacist rhetoric) and a variety
of sources and formats immediately brought historical figures to life for 18-year-
old college students and sparked lively conversations about the range of research
inquiries that could be supported by such materials, and how search results shape

Figure 6: Primary Source Worksheet, developed by Jennifer Hootman, Cecily Marcus,
and Ben Wiggins.

one's imagination of what's possible, and of what happened. The difference between the two search engines couldn't have been clearer, and the exercise is easily adaptable for other library databases and online collections (e.g. the DPLA, or any university online digital collection or institutional repository) and for a range of grade levels and disciplines.

Working directly with students also put us in conversation with educators and allowed us to better understand their needs and how *Umbra Search* could be a helpful resource for them and their students. Presenting *Umbra Search* in workshops to teachers at History Day MN, Saint Paul Public Schools, and other educator training sessions created a forum of reciprocal feedback about the tool's efficacy and potential for our core users. Local partnerships with History Day MN facilitated collaborative opportunities with National History Day, which started with National History Day sharing *Umbra Search* in their listserv to teachers nationwide, and came to include the development of a video for History Day students and teachers and promoting the resource at the 2017 national competition by sharing stickers and bookmarks in competitor packets.

Public events locally and nationally were not only welcome opportunities to substantively collaborate with our partners, but also brought people together to have important conversations that might otherwise have been difficult to find, start, or sustain online. With the University of Pennsylvania *Umbra Search* hosted 'An Honest Reckoning: Building Access to Black Collections', with participation from curators and librarians from the Philadelphia area who work with African American collections and materials. With Virginia Commonwealth University we sponsored 'Making the Invisible Visible: Activating Black History Through Digital Storytelling', a panel discussion with institutions and individuals working on digital projects. We had *Umbra Search* events with Howard University, Emory University, a pop-up event and a working session with scholars at UCLA, a book launch for *Hidden Human Computers* by Macalester College professor and author Duchess Harris at the Science Museum in St. Paul, a community hackathon for social justice with open data, and roundtables on archives and Black arts in the Twin Cities. Support from the Doris Duke Charitable Foundation allowed us the flexibility to pursue new and creative

partnerships with a range of organizations, from traditional libraries to Free Black Dirt, an artist collective formed by Minneapolis-based collaborators Junauda Petrus and Erin Sharkey.

Much of this work was part of *Umbra Search*'s public launch, a year-long effort throughout 2017 that aimed to introduce *Umbra Search* and its main preoccupations to broad audiences, from the general public to university students to artists and archivists. Consisting of a widely-circulated press release in January 2017 that yielded dozens of newspaper articles around the country, as well as local radio and TV coverage, and podcast interviews; the #UmbraSearch365 Twitter campaign that pushes out *Umbra Search* content every day of the year, making the argument that Black history is not confined to Black History Month; and a series of events, the launch has grown the umbrasearch.org user base and exposed new areas for future investment. It culminated with two Minneapolis events featuring the work of Dr. Amma Y. Ghartey-Tagoe Kootin, University of Georgia professor and the creative force behind *At Buffalo*, a new musical in development about the 1901 World's Fair in Buffalo, NY, that included three conflicting views of Black identity: an 'Old Plantation' exhibit featuring formerly enslaved people, 'Darkest Africa', in which hundreds of West Africans of all ages performed 'African' rituals and daily life, and W. E. B. Du Bois's 'Negro Exhibit' that celebrated the intellectual, economic, and cultural achievements of African Americans. With a livestreamed roundtable discussion with Dr. Ghartey-Tagoe Kootin on African American theater and archives with poet and scholar Dr. Alexis Pauline Gumbs and playwright and director Talvin Wilks, and a performative lecture about the creation of *At Buffalo*, Dr. Ghartey-Tagoe Kootin articulated the power of the archive to hold us accountable for the past and our present, and the power of art to articulate what the archive cannot—the silences, gaps, and profound losses that form the historical record.

Building the Corpus Through Systematic Digitization

As a relatively small collection of about 10,000 rare books and archival collections, the Givens Collection of African American Literature has many, many gaps, and it necessarily draws strength not only from the scholarly and creative works it has

yielded, but also from its connections with sister African American collections across the country. *Umbra Search* became the digital network that allows the Givens Collection—and all the other collections around the country—to work in concert to demonstrate their impact, and build a collective power they may not have alone. As a result of this work, we can now connect the Givens Collection's letters from Countee Cullen to a childhood friend, written in English, French, and Latin and containing manuscript poems, some of which have never been published, to Cullen's correspondence with W. E. B. Du Bois held at the University of Massachusetts Amherst, and to his archives in New Orleans at the Amistad Research Center. The efforts of developing *Umbra Search*, and of engaging in conversations with partners all over the country about how they could maximize the impact of their own African American collections, compelled us to look closely at the materials in *Umbra Search* that come from Givens Collection, and the University of Minnesota's archives and special collections more broadly. We looked differently at our own collections at the University of Minnesota, and not only the Givens Collection of African American Literature, asking the same questions of our own materials and collections as we did of partners': Where are our African American archives? What is missing? What can we digitize next? In talking with colleagues around the country, and in making the argument that African American history and cultural practices are central to our understanding of our collective past, embedded in every thread of the fabric of American and global histories, it became clear to us that we needed to survey the hundreds of collections across all 17 collecting areas at the University of Minnesota's department of Archives and Special Collections—from Children's Literature to Social Welfare History to Performing Arts to the history of the University of Minnesota—to find materials documenting African American history that are not in 'African American' collections.

Starting in 2015, we began that process, finding that there was not a single collecting unit that did not hold significant caches of materials documenting different aspects of African American history. Through keyword searches, conversations with our archivist and curator colleagues, and by pulling collections to look inside boxes and folders, we identified materials from 74 collections—magazines, manuscripts,

pamphlets, ephemera, organizational files, correspondence, and illustrations, some of which date back to the sixteenth century, as well as video and sound recordings—that now make up the basis for a two-year mass digitization effort to make African American history materials from University of Minnesota collections accessible online, through a Hidden Collections digitization grant from the Council on Library and Information Resources. When the project is completed in 2018, we will have digitized materials from at least 162 collections rather than the 74 we initially identified, adding richer metadata to make the materials more discoverable, and adding nearly 500,000 scans to the University of Minnesota's digital collections. Those materials will be aggregated by DPLA and they will become part of *Umbra Search*.

Searching for our own 'hidden' African American collections, as well as analyzing the *Umbra Search* aggregation for what was and was not included, was not just a question of identifying content. It is an ongoing process that requires us to identify and interrogate the practices of description, categorization, and physical and intellectual arrangement that contribute to obfuscating African American lives and stories, rather than illuminating them, within our own collections and across the archival field generally. Historically, the descriptive archival practices that have been established by a predominantly white field of archivists, curators, and librarians has failed to capture diverse identities, obscuring the impact and meaning of cultural difference and asserting whiteness as the dominant force within history.

Naming Practices, Reconsidered

In the process of working through our own systematic digitization of materials that documented different aspects of African American history but that had not been labeled or identified as such, we also confronted how our own institutional history and practices shape what our collections are (and are not), who created them (and who did not), what they contain (and what they leave out), and the ways we have arranged and described them that privilege some kinds of information and obscure others. Such erasures and identifiers may be predictable given the historical and current predominance of white people working in the Libraries and its archives and special collections; a historical lack of sustained engagement with the communities

whose collections are housed by the institution; and commonly used arrangement and description practices like 'More Product Less Process' that enable archives to address collection backlogs but require generalized descriptions to cover hundreds if not thousands of discrete objects in folders, boxes, or collections.

Even before the digitization effort began, we were aware that in the increasingly highly distributed environments of digital collections, the ubiquitous challenges around description, discovery, and access are necessarily amplified, but we had not looked closely and systematically at how our own practices contributed to the predicament. We knew that in large-scale aggregations like DPLA and *Umbra Search African American History*, materials that are inadequately described may be lost in aggregation within the much larger corpus, but we had not confronted how and why our own discrete records become unconstrained by and unmoored from their home digital repositories in these aggregate environments, shedding critical intellectual context in the process. If the descriptive metadata is scant or overly general in a local digital collection, the legibility and meaning of materials in a massive aggregate only compound the limitations of standard practices. Without some meaningful interventions, we may be hiding materials more deeply, even in the name of making them more discoverable.

The process of digitizing hundreds of thousands of African American history materials compelled us to look at our own practices, and to take on a more active role in discussions about inclusive descriptive practices in the field of archives and special collections, such as those being led by the Amistad Research Center's 'Diversifying the Digital Historical Record' initiative, with partners the Shorefront Legacy Center, the South Asian American Digital Archive, Murkurtu, and the Inland Empire Memories Project at the University of California-Riverside. It was the subject of a 2016 workshop hosted by the Association for Library Collections and Technical Services (ALCTS), a division of the American Library Association. It was the subject of a working session at 2017 DPLAFest, led by *Umbra Search*. The questions of standards, cultural diversity, automation, scalable practices, and institutional cultures and capacities are heady. Locally, we add enhanced metadata to the collections, boxes, and folders that have

been digitized, which is visible both in the digital collections/aggregate interfaces and in finding aids. We have convened a group to make recommendations about how better to implement inclusive description practices. Any serious progress, both locally and field-wide, towards addressing diverse and inclusive description practices requires further investment—significant time and funding, collaboration, leadership, and sustained institutional commitments. This work continues.

Governance and Sustainability

As the original 'African American Theater History Project' became the *Umbra Search* search tool and widget, and then took on new dimensions around community engagement, outreach, education, and digitization, what had started as a project became a program. As such, we needed to concretize how *Umbra Search*, the Givens Collection, and the University of Minnesota engage meaningfully, reciprocally, and respectfully with the local African American community; how we build trust; and how we sustain and grow not only umbrasearch.org, but also all the dimensions of the *Umbra Search* program. These range from how *Umbra Search* participates in national discussions around Black digital humanities, building an inclusive and diverse field of archives, special collections, and libraries, by addressing staffing and staff retention, collecting priorities, collaborations, and description, and by promoting digital literacy and the integration of African American primary sources across educational, creative, and scholarly contexts.

This work necessitates and has been strengthened by an active Advisory Council, one that *Umbra Search*'s principal investigator and project manager developed in collaboration with Free Black Dirt artists Erin Sharkey and Junauda Petrus as consultants. The charge was to identify and engage a diverse group of scholars and educators, artists and activists, and archivists and curators,[9] who provide broad expertise and insight to guide a number of aspects of *Umbra Search*:

---

[9] Advisory Council members include: Dorothy Berry, Houghton Library, Harvard Univesrsity; Janet Bishop, Associate University Librarian, University of Minnesota Libraries; Valerie Caesar, Black Seed Photography; Sarah Carlson, University of Texas at Austin; Lynée Denise, Los Angeles; Jennifer Gunn,

· Intellectual and programmatic integrity;

· Priority setting and long-term sustainability planning;

· Identifying and implementing strategic partnerships and collaborations;

· Engagement with diverse communities and users.

In its design and member makeup, the council aims to reflect the very ways our encounters with historical silences in the archive take shape—in teaching and research, cultural production, and administration and curation—so that it can better interrogate the state of African American archival collections and the work of *Umbra Search* as one intervention for the systemic gaps that are created within our collections. The Council has a somewhat unusual makeup for a program of an academic library: in addition to several library and special collections leaders and University of Minnesota faculty, there have been four artists from Minnesota, New York, and Los Angeles, two of whom also teach in university settings; an educator from the Minneapolis Public Schools; and an information technology specialist who also founded Blacks in Technology, a Minneapolis-based group, and Code Switch, a community-based hackathon for social change. With an in-person meeting over two days in Minneapolis in 2016, and over quarterly WebEx video conference calls, the Advisory Council is a critical aspect of *Umbra Search*'s ongoing development, growth, and sustainability.

One of the ongoing and impactful roles of the Advisory Council is to help navigate myriad questions related to the long-term sustainability of *Umbra Search*, and to help guide the inevitable transitions that result from project-based, short-term grant funding. Grant funds are transformative and heartbreaking. Grants allow us to pursue work that far exceeds most institutions' regular capacity and scope,

---

Institute of Advanced Study, University of Minnesota; Ezra Hyland, College of Education and Human Development, University of Minnesota; Athena Jackson, Head of Special Collections Library, Penn State University Libraries; Sharon Kennedy Vickers, IT Management Consulting; Kara Olidge, Executive Director, Amistad Research Center; Cristina Pattuelli, School of Information, Pratt Institute; Junauda Petrus, Performance Artist, Writer, Free Black Dirt; Erin Sharkey, Writer, Free Black Dirt; Catherine Squires, Professor of Communication Studies and Director of Race, Indigeneity, Gender & Sexuality Studies Initiative (RIGS), University of Minnesota; John S. Wright, Morse-Amoco Distinguished Teaching Professor, Departments of African & African American Studies and English, University of Minnesota.

and that pushes against systemic problems in our collections, but they rarely result in radical and systemic change. A fundamental limitation of grant-funded projects, regardless of their innovation or commitment to dismantling hegemonic collecting practices, is their inability to promise major improvements or transformations after the grant period ends. In most cases, the end of the grant spells the end of the effort entirely—resources created may be sustained online or otherwise, knowledge will be shared, but more work will not take place. With every grant that comes to a close, critical knowledge and efforts are lost. Talented staff in whom you have invested leave. When projects such as *Umbra Search* talk about sustainability, they are often talking about maintenance, technical life support, and the persistent availability of the project resources that were created (a website, recommendations, a white paper, survey findings, survey tool, etc.). Sustainability for *Umbra Search* means the lights will be kept on, but no active development or outreach or research will take place without new funding. *Umbra Search* will continue to be available, content will be regularly harvested, bugs will be fixed, and software upgrades will be managed. It will be there, but it will not fundamentally change. It won't get better.

All of the objectives outlined in multiple grants and phases of the *Umbra Search* program have been met. We built an aggregator of African American primary source material from institutions all over the country, developed productive and wide-ranging partnerships, engaged diverse audiences, and added important content. *Umbra Search* is viewed as model for how to 'remake collections' around a topic, discipline, or field, and we have consulted with the Chicano Studies and the American Studies departments at the University of Minnesota about how to build an *Umbra Search* for their interests, raising the question of whether 'Umbra Search' is a technology model that can become *Umbra Search Chicano Art* or *Umbra Search Anti-Semitism at the University*; or whether *Umbra Search* is distinct from its technology and driven more by needs and questions that are specific to African American history, culture, and collections. Is it both?

At the same time, there are ways in which *Umbra Search* has yet to fulfill its own potential, and we continue to draw energy and inspiration from related projects—the University of Delaware's extraordinary Colored Conventions effort, the African

American History, Culture, and Digital Humanities initiative at the University of Maryland, The HistoryMakers, and others.

Steps Forward

The future of *Umbra Search* lies not just in its institutional/financial/technological life at the University of Minnesota. Sustainability for *Umbra Search* includes but goes beyond the server it lives on, and the programmer who harvests new content four times a year and fixes bugs or brings it back online when it goes off. It is more than what content we will still add, either through DPLA or independently with organizations like Weeksville Heritage Center, a non-profit in Brooklyn, NY, or the Schlesinger Library at Harvard University. *Umbra Search*'s future is very much tied to the efficacy and impact of the work of our partners, from AADHUM to HistoryMakers to the Amistad Research Center, and how we continue to work together to build curriculum and engage students; systematically assess the state of Black collections in the United States through surveys and other tools; engage in serious efforts to develop collaboratively improved practices around inclusive archival description and collection development; rethink dependence on processing procedures such as 'More Product Less Process' in terms of impact on mass digitization as well as how they obscure cultural difference and re-inscribe whiteness as dominant culture; and share resources for archival arrangement, digitization, and digital collection building and hosting of non-custodial content in order to remake our collecting, as well as our collections. In many ways, the incompleteness of the archive that *Umbra Search* initially sought to address continues to drive its work. We have witnessed what can happen when we have at our fingertips an incredible trove of materials tracing African American history and memory. We have seen how much more material there is to be uncovered, in and outside of our libraries and archives. We know how much work—physical, digital, ethical, and political—there is yet to do to radically remake African American collections in a way that will make them more complete and inclusive, *and* more transparent about how and why they are still and will always be incomplete.

Competing Interests

The authors have no competing interests to declare.

References

Berry, D 2016 Umbra Search African American History: Aggregating African American Digital Archives. *Parameters.* Available at: http://parameters.ssrc.org/2016/12/umbra-search-african-american-history-aggregating-african-american-digital-archives/ (Last accessed 28 August 2018).

Brothman, B 2011 The Society of American Archivists at Seventy-Five: Contexts of Continuity and Crisis, A Personal Reflection. *The American Archivist,* 74(2): 387–427. DOI: https://doi.org/10.17723/aarc.74.2.3853871526353023

Carini, P 2009 Archivists as Educators: Integrating Primary Sources into the Classroom. *Journal of Archival Organization,* 7: 41–50. DOI: https://doi.org/10.1080/15332740902892619

Carter, R 2006 Of Things Said and Unsaid: Power, Archival Silences, and Power in Silence. *Archivaria,* 61: 215–33.

Caswell, M L 2016 Seeing Yourself in History: Community Archives in the Fight Against Symbolic Annihilation. *UCLA.* Available at: https://escholarship.org/uc/item/9gc14537 (Last accessed 28 August 2018).

Caswell, M L and **Gillibrand, A** 2015 False Promise and New Hope: Dead Perpetrators, Imagined Documents and Emergent Archival Evidence. *The International Journal of Human Rights,* 19(5): 615–27. DOI: https://doi.org/10.1080/13642987.2015.1032263

Cook, T 2011 'We Are What We Keep; We Keep What We Are': Archival Appraisal Past, Present and Future. *Journal of the Society of Archivists,* 32(2): 173–89. DOI: https://doi.org/10.1080/00379816.2011.619688

Diversifying the Digital http://diversifyingthedigital.org/index.html (Last accessed 28 August 2018).

Documenting Ferguson http://digital.wustl.edu/ferguson/ (Last accessed 28 August 2018).

Fenlon, K, Lett, J and **Palmer, C L** 2017 Digital Collections and Aggregations. *DH Curation Guide: A Community Resource Guide to Data Curation in the Digital Humanities.* Available at: http://guide.dhcuration.org/contents/digital-collections-and-aggregations/ (Last accessed 28 August 2018).

Gibbs, R 2012 The Heart of the Matter: The Development History of African American Archives. *The American Archivist*, 75(1): 195–204. DOI: https://doi.org/10.17723/aarc.75.1.n1612w0214242080

Godfrey, M 2016 Making African American History in the Classroom: The Pedagogy of Processing Undervalued Archives. *Pedagogy*, 16(1): 165–77. DOI: https://doi.org/10.1215/15314200-3158733

Harris, V 2002 The Archival Sliver: Power, Memory, and Archives in South Africa. *Archival Science*, 2: 63–86. DOI: https://doi.org/10.1007/BF02435631

Helton, L, et al. 2015 The Question of Recovery: An Introduction. *Social Text 125*, 33(4): 1–18.

Jimerson, R C 2007 Archives for All: Professional Responsibility and Social Justice. *The American Archivist*, 70(2): 252–81. DOI: https://doi.org/10.17723/aarc.70.2.5n20760751v643m7

Jules, B 2016 Confronting Our Failure of Care Around the Legacies of Marginalized People in Archives. *Medium*. Available at: https://medium.com/on-archivy/confronting-our-failure-of-care-around-the-legacies-of-marginalized-people-in-the-archives-dc4180397280 (Last accessed 28 August 2018).

Kaplan, E 2000 We Are What We Collect, We Collect What We Are: Archives and the Construction of Identity. *The American Archivist*, 63(1): 126–51. DOI: https://doi.org/10.17723/aarc.63.1.h554377531233l05

Murkutu http://mukurtu.org/ (Last accessed 28 August 2018).

Orwell, G 1981 [1946] Looking Back on the Spanish Civil War. In: *A Collection of Essays*. Orlando: Harcourt.

Poole, A H 2014 The Strange Career of Jim Crow Archives: Race, Space, and History in the Mid-Twentieth-Century American South. *The American Archivist*, 77(1): 23–63. DOI: https://doi.org/10.17723/aarc.77.1.g621m3701g821442

Theimer, K 2012 Two Meanings of 'Archival Silences' and their Implications. *Archivesnext*. October 2017. Available at: http://archivesnext.com/?p=2653 (Last accessed 28 August 2018).

Transgender Oral History Project at the Tretter Collection https://www.lib.umn.edu/tretter/transgender-oral-history-project (Last accessed 28 August 2018).

Umbra Search African American History https://www.umbrasearch.org/ (Last accessed 28 August 2018).