## Enhancing the The Diverse BookFinder

**Abstract.** Bates College is seeking a National Leadership grant to address the challenges faced by children's librarians trying to develop picture book collections that reflect the racial and ethnic diversity of their constituencies and the nation. Towards this end Bates College received an IMLS Sparks! Ignition grant (SP-02-16-0006-16) to create a controlled vocabulary to fill the need for a systematic methodology for pinpointing titles featuring people of color. We have solidified a vocabulary and are on schedule to make it public in September, 2017, via what we call the Diverse BookFinder (previously called the Diverse Picture Book Resource Hub), a searchable online database of picture books featuring human characters of color published since 2002 (this tool is linked with The Diverse Book Collection, a circulating body of picture books featuring people of color housed at Bates College Ladd Library). Within the current proposal, co-Principal Investigators Krista Aronson (Associate Professor of Psychology, Bates College) and Christina Bell (Bates College Humanities Librarian), in collaboration with Anne Sibley O'Brien (children's book creator and diversity educator), seek funds to make two Diverse BookFinder enhancements: 1) The creation of an online collection development tool to facilitate the identification of diverse picture book titles for acquisition; 2) Additional Search Engine Optimization (SEO) to enhance the discoverability of website information through the inclusion of colloquial synonyms and phrases in available search mechanisms. Following these enhancements we will build awareness of the Diverse BookFinder among librarians and library users through a Strategic Dissemination plan, partnering with State and national library organizations and members of our Advisory Council, and making use of social media.

The American Library Association recognized the importance of diversity in library collections in 1953 through the Freedom to Read Statement (American Library Association, 1953), and again in 2013 through the Diversity in Collection Development Statement (American Library Association, 2013). But, despite this recognition, there is evidence that more than one third of libraries spending $100,000 annually on book purchases do not achieve adequate representations of racial and cultural diversity within their youth collections (Williams & Deyoe, 2014). The enhanced Diverse BookFinder will address two challenges: identifying titles for purchase and retrieving and utilizing titles once they're in local collections, for patron requests, programming or display. Feedback from librarians about the BookFinder indicates that the proposed enhancements are necessary to fully support collection development and book discovery. If funded, enhancements and dissemination will commence October 1, 2017 and continue for three years.

By assisting with the development of diverse children's collections, the Diverse BookFinder will enhance libraries' position as a community anchor through the creation of a collection that reflects and serves their local communities (IMLS Strategic Goal #2). By facilitating access to diverse books this project will also help librarians support the development of positive self-image and intercultural competence among children, thereby contributing to their ability to thrive in our increasingly diverse world (IMLS Strategic Goal #1). The online format of this project offers the potential to reach a broad audience, including public librarians, school librarians, K-3 teachers, higher education faculty and students, parents and catalogers of children's materials as well as publishers.

# Enhancing the The Diverse BookFinder

**1. Statement of National Need.** Libraries "are at the forefront in the movement to create a nation of learners" and are "stewards of cultural and natural heritage with rich, authentic content...provid[ing] learning experiences for everyone" with the goal of "prepar[ing] people to be full participants in...our global society" (IMLS, 2017, retrieved from https://www.imls.gov/nofo/national-leadership-grants-libraries-fy17-notice-funding-opportunity/ section/a). Through the diversity of their collections, libraries have the opportunity to engage young readers with their own cultural heritages and introduce them to those of the people around them and around the world, helping them to become members of our global society. The inclusion of characters from multiple racial and ethnic backgrounds allows readers of color to see their lives and experiences legitimized (mirrored; Cullinan, 1989), rendering their experiences and behaviors visible and valuable (Baghban, 2007). The inclusion of characters of color also supports literacy development: when children of color see themselves depicted in books they retain and recall more plot and character information because they feel connected (Smith & Lewis, 1985).

Diverse books also allow children to experience the breadth and depth of other experiences and cultures (diverse books provide a window; Cullinan, 1989). This may be particularly important in racially homogeneous communities, where children are unlikely to come into direct contact with those who may differ from themselves (e.g. Maine). Although the importance of diverse books within racially heterogeneous communities may be intuitive, there is evidence that children living in homogeneous envirinments benefit most from exposure to diverse books (Cameron, Rutland, Hossain & Petley, 2011). When the range and breadth of the experiences of children of color are absent from children's literature, stereotypical images -- what author Chimamanda Ngozi Adichie (2009) calls the "single story" -- go unchallenged, limiting all children's opportunities to deepen their understanding of those who may differ from themselves, to see them as more than the "other" (Cameron, Rutland, Brown & Douch, 2006).

Previous research done via the Diverse Book Collection at Bates College confirms that picture book collections in libraries across the country do not accurately reflect the demographics of our country. Using WorldCat data we illuminated 20 national libraries that have the most diverse titles, from an array of areas (urban and rural) with varying collection budgets. **The aggregate percentage of picture books featuring people of color within these top collections was determined to be between two and four percent.** These data, coupled with interview data from the librarians responsible for building these collections, indicate that identifying diverse titles for acquisition is difficult. Many librarians in this group have specialized graduate or ALA training, which is not common nationally (Schliesman, 2017). All "plug away", scouring the School Library Journal, Kirkus Reviews, blog posts and award winner lists, to develop their diverse holdings, which is time consuming. Although they are leading the nation with their efforts, our analysis indicates that these libraries have almost identical diverse titles on their shelves, which is surprising given the array of community compositions (Aronson, Murray, Peralta, Daus-Haberle, Orlando, 2017).

Indeed, **no comprehensive source or systematic methodology currently exists for pinpointing titles featuring people of color**. The Cooperative Children's Book Center (CCBC), located in the School of Education at the University of Wisconsin-Madison, houses non-circulating children's and young adult literature, including multicultural titles, donated by publishers (approximately 3500 of an estimated 5000 titles published annually). Their collection of multicultural titles is a subset of all the books they receive; it is not intended to be comprehensive, nor even broadly representative. And, although the staff at CCBC do generate a list representing their multicultural holdings, they publish only data about numbers of books by and about each racial and cultural group, not the lists of actual titles. Consequently, there is no one place where librarians can identify a representative list of books featuring people of color available for purchase.

Once a library identifies a diverse children's book and adds it to their collection, library patrons should be able to locate it. Those who use children's book collections, however, are keenly aware that browsing is inefficient because of the dimensions of the books themselves, the classification schemes often employed (focusing on authors rather than subjects), and the lack of orderliness due to high use. Using the library's catalog offers few additional avenues to discovery due to the limitations inherent in the current Library of Congress subject headings available for juvenile literature, even those utilizing general subject headings (LCSH), instead of juvenile subject headings (LCSHAC). Those that do exist provide limited information about characters, themes or thematic approaches to content.

For instance, although the Bates College Library houses a collection of picture books featuring children of color, a preliminary round of searching in the library catalog reveals several problems in locating diverse children's books using existing subject headings. For example, although the collection includes works with children of Asian descent, searches for fiction with diverse ethnic and racial groups of children (i.e., "Japanese American children--Fiction" or "Chinese American children--Fiction") retrieve no records. Some listings within the Bates catalog include adult versions of the headings instead of generic group designations (i.e., *When I was Eight* includes a heading about "Inuit women" instead of "Inuit children" or "Inuit"). This inconsistency makes comprehensive retrieval impossible. Also, there currently exists no way to find all books including characters from a large ethnic or racial grouping. To find all diverse books featuring Latinx, for example, one would have to search under multiple headings such as "Mexican Americans," "Cuban Americans," "Puerto Ricans," etc. Similarly, a search for all books with African Americans would also need to include headings for "Ghanaian Americans," "Sudanese Americans," and so on.

**Creating a controlled vocabulary to improve discoverability.** Bates College received an IMLS Sparks! Ignition grant (SP-02-16-0006-16) to create a vocabulary to address this challenge. The product makes several enhancements to existing lexicons, including greater specificity regarding the **background of a book character** (*who* is in the book*)*, delineating a set of terms that can be used to describe character heritage at a number of levels, including *racial/cultural* (e.g. Asian/Pacific Islander/Asian American, Bi/Multiracial, Black/African/African American, Central and South American (Hispanic/Latinx), First/Native Nations/American Indian/Indigenous, Middle Eastern/North African/Arab, White/European

American/Caucasian, unable to identify); ***ethnic/national*** (e.g. Afghan, Congolese, with hundreds of possible designations); and ***tribe/tribal nation*** (if applicable; e.g. Shawnee Tribe, with hundreds of possible international designations), as well as terms that describe character gender, immigrant status and religion (where applicable). This vocabulary also includes terms to describe the **book's setting** *where* the story takes place (e.g. United States -- Montana; as well as international locations, e.g. Nigeria, with hundreds of possible setting designations). The identification of a character's race, culture or ethnicity is done by consulting published information about the book (e.g. information included in the narrative itself, author notes, and/or other online information such as that provided through Kirkus Reviews).

In addition, this vocabulary incorporates terms to describe *what* the book is about, leveraging the nine themes identified by Aronson, Callahan and O'Brien (2017) through content analysis: Every Child, Cross Group, Beautiful Life, Concept, Oppression, Folklore, Biography, Incidental, Informational (see Thematic Description Table). We initially envisioned translating our thematic vocabulary into terminology from the primary existing vocabularies used to describe picture books -- Library of Congress subject headings (LCSH) and/or subject headings for children (LCSHAC), BISAC terms, etc. However, we decided that this was not feasible due to insurmountable limitations in the existing vocabulary structures and the rules for applying them in bibliographic description, such as inconsistent application of racial/ethnic identifiers, and the lack of flexibility to incorporate both aggregating *and* specific ethnic/cultural terms to describe diverse picture books, to assist in retrieval for the multiple ways the books could be used. For example, a librarian might be seeking all books about African Americans for a Black History Month display, but might later need books about a more specific subgroup for a lesson on intercultural communication with a local community of Somali Americans.

Particularly challenging is the cataloging convention that limits inclusion of subjects to only what a book is "about," so that for books where characters of color are included but their ethnicity or race is not central to the book's narrative (in our categories of Every Child or Incidental, for example), subjects indicating diversity would be left out. Further, there are manifold LCSH terms to describe conflict and oppression, for example, but painfully few to describe intergroup cooperation and friendship (essential for our Cross-Group category), or a cooperative versus conflicted portrayal of intergroup relations. Searches for materials depicting interactions across racial or ethnic groups are particularly difficult to identify. Within the Bates College catalog, for instance, a search for "Intercultural communication" retrieved no books. Even "Race relations--Juvenile fiction" retrieved only five books. Not included in those retrieved was the book *Desmond and the Very Mean Word*, which deals with cross-group relations. (These issues are not unique to the Bates catalog, but are used as examples to illustrate the challenges associated with LCSH utilized in cataloging juvenile literature.)

Because of these limitations in existing controlled vocabularies, we have created new thematic and racial/ethnic/tribal vocabularies to coexist with traditional bibliographic and subject descriptions of picture books. Following in the footsteps of other projects like the First Nations House of Learning Subject Headings used in the Xwi7xwa Library of the University of British Columbia Library System, our database blends traditional cataloging with our specialized vocabulary. We hope by doing this to build on the strengths of existing bibliographic description

(familiarity for librarians, extensive vocabulary of topical subjects like "Birthdays," pre-existing descriptive cataloging), while utilizing our vocabulary to enhance access and retrieval for diversity-specific picture book purposes.

We have been vetting our vocabulary with targeted audiences through conference presentations and consultations with librarians and publishers participating in national research through the Aronson lab. Thus far, feedback regarding the vocabulary has been overwhelmingly positive. We are looking forward to further vetting our vocabulary as a tool to facilitate the discovery of diverse picture books during the soft launch of our website. Consistent with the Sparks! timeline, this soft launch is scheduled for July, 2017. As part of this phase of the project we will circulate a questionnaire to regional and national librarians through the following groups and associations: Association for Library Services to Children (the Children's Collection Management in Public Libraries Discussion Group); ALA's Ethnic and Multicultural Information Exchange Round Table (Guidelines for Selecting Multicultural Materials Taskforce); Maine Youth Librarians; and the Maine Association of School Libraries, soliciting feedback about the comprehensiveness and comprehensibility of our vocabularies with a focus on their utility in aiding discoverability. Using this feedback, we will make refinements to our vocabulary prior to the public launch in September, 2017.

**More about the Diverse BookFinder.** The Bates Sparks! Ignition grant (SP-02-16-0006-16) also funded creation of The Diverse BookFinder, an online searchable tool to facilitate the discoverability of multicultural titles through various mechanisms, including a faceted search designed to allow users to apply multiple filters according to project vocabulary. This faceted search will operate much like any "shopping" feature, whereby a visitor will be able to select from a pre-designated set of categories enabling them to sort any term included in our vocabulary that is relevant to their interest (e.g. racial/cultural, ethnic/national, thematic, publication year, etc.). In line with the Sparks! grant timeline, this component of the website (a faceted search) is currently under development and will be rolled out at the soft launch (July, 2017). Following the completion of short tail SEO work (described in section 2b), visitors will also be able to discover titles by entering commonly used terms synonymous with, but not currently included in our vocabulary (e.g. prejudice, racism) in a simple search bar and arrive at a relevant list of titles (e.g. oppression books). This component of the website (simple search bar) will be available at the time of the public launch in September, 2017. Finally, visitors will also be able to discover books by accessing our "Quick Lists," which we have developed based on requests we have received and topics identified as commonly searched (e.g. skin color, civil rights, immigration, friendship, multiracial families). Clicking on one of these will provide the visitor with an immediate list of titles related to that topic. Finally, the website will contain information about who we are (the project team), our vocabulary, scholarly research on the impact of diverse books, and further reading on the need for diverse titles. Although the search tool and logo portion of the website are currently under development, background information is currently live and can be accessed using the following link (www.bates.edu/diverse-bookfinder).

The Diverse BookFinder is linked with The Diverse Book Collection housed at Ladd Library on the Bates College campus, featuring characters from First/Native Nations, Asian, Black, Latinx, Middle Eastern/Arab, and multiracial backgrounds, in stories set in the United States and abroad.

Inclusion criteria are as follows: trade books (those traditionally acquired by libraries and reviewed in literary journals, not including mass market and media tie-in titles; the body of literature most commonly used by those interested in working with and transforming the lives of young readers) suitable for children in kindergarten through grade three, approximately 32 pages in length, with illustrations that are a key part of understanding the book, and with human characters featured in fiction and narrative non-fiction stories. Books featuring animals as primary characters are not included, with a few exceptions of stories in which animals clearly represent people from a specific cultural context, such as *Pancho Rabbit and the Coyote: A Migrant's Tale* (Abrams Books for Young Readers). Books are identified for inclusion using multiple mechanisms, including consulting The Cooperative Children's Book Collection, regularly searching 111 children's book publisher websites, the "World Full of Color" children's book database and the "I'm Your Neighbor" database, and 18 award sites, including but not limited to the Coretta Scott King Book Award, the Pura Belpré Medal, the Ezra Jack Keats Book Award, the Asian Pacific American Award, and the Caldecott Medal. The team also obtains Baker & Taylor lists using the following profile: ages 0-9, all bindings except board books, English or bilingual, and fiction or nonfiction. Any book with a possibly multicultural Juvenile BISAC (Book Industry Standards and Communications) code (including sub-categories of 'People & Places,' 'Historical' Fiction, 'History,' 'Social Issues,' 'Holidays & Celebrations,' and 'Biography & Autobiography'). **It is important to note that no other existing collection has been so purposefully and exhaustively developed, and that the Diverse Book Collection represents the only circulating collection of its kind.**

The Diverse BookFinder provides a central location where librarians can identify picture book titles depicting racial and ethnic diversity for inclusion in their collections using a common language. By assisting with the development of diverse children's collections, the Diverse BookFinder will enhance libraries' position as a community anchor through the creation of a collection that reflects and serves their local communities (IMLS Strategic Goal #2). By facilitating access to diverse books this project will also help librarians support the development of positive self-image and intercultural competence among children, thereby contributing to their ability to thrive in our increasingly diverse world (IMLS Strategic Goal #1).

**2. Project Design**. Feedback we have received from librarians thus far regarding the Diverse BookFinder has clarified what is needed to make it fully accessible and useful: a collection analysis tool and additional SEO work. We will need three years to develop these enhancements. **Below, we describe the enhancements we are seeking to fund**, followed by a timeline for their completion.

**2.a. Collection Analysis Tool.** Librarians need access to detailed information -- the books' number and nature -- to advance their diverse book holdings. **Although many libraries already receive detailed collection and subject profiles from companies like Baker and Taylor, these products are not relevant for diverse titles.** Further, there is no good way for librarians to track diversity in their collections, once they're ordered and cataloged. As previously described, tracking down books about African Americans through searching subject headings, for instance, may or may not result in accurate information given inconsistencies in term application. Further, there's no way to further break down what books are about (e.g. getting

along across difference; positive or resolving cross-group friendship), making it difficult to find books for patrons or programming. While a "Best of" or "Suggested titles" list can be useful, it does nothing to consider what is already in a library collection or help identify gaps in representation.

**This project will develop a free online collection analysis tool to provide such information in both narrative and graphical forms**. Specifically, a function will be embedded in the Diverse BookFinder to which librarians can upload a file with information about their picture book (or children's book) collection. We are aware that libraries across the country use different data/catalog platforms; all of the platforms we have researched have the capability of exporting information into .csv (excel based) format. The online collection analysis tool will ingest files of these forms through a web based interface. Librarians interested in receiving a detailed report of their diverse books will be asked to create a login, at which point they will share their email address and respond to a set of demographic questions about their library (state, region of the country) to help us with benchmarking (described below). They will then be prompted to upload a .csv file with book title and ISBN information for all of their picture books. Information contained within this data file will then be cross-referenced with master data for books contained within the Bates College Picture Book Collection. This matching process is necessary because existing catalog information related to race, ethnicity and culture is insufficient and applied inconsistently as described in the above narrative. Comparing each library list with the Diverse Book Collection allows us to extract a representative sample of titles from which to create a report. Following the matching process, the percentage of racially diverse picture books for each library will be calculated by dividing the number of matching titles between the library's shelf list and the Bates Diverse Book Collection list by the total number of individual titles in that library's shelf list (calculated using each library's shelf list after the duplicates have been parsed out). The resulting report will convey several pieces of information important to collection development: the number and titles of diverse titles and their percentage of the overall whole; what groups are depicted within these titles and their relative percentage; and, how groups are depicted. Once the report has been generated (anticipated time 5 minutes) the librarian will be sent an email with a link and passcode that will allow them one-time confidential access to their downloadable PDF report (see sample report).

We have piloted this tool with 9 libraries, resulting in 8 confidential reports. **Pilot librarians have been overwhelmingly pleased with their reports (see letter of support)** the content of which led one to realize that only 2.4% of their picture books depicted racial, ethnic or cultural diversity, a particularly shocking realization given that 19% of their constituents were people of color, including people of Asian, Latinx and African (largely immigrant) heritage. It was also interesting for the librarian to learn that 65% of their diverse titles feature characters of African American descent, most often in books classified as oppression (stories of systemic and institutionalized prejudice; slavery, civil rights), and that books classified as cross-group (books portraying character interactions across racial or cultural difference) featured Black and White characters only. Thus, their report revealed important gaps in their collection in relation to their population it serves. After receiving their report librarians will be able to use the Diverse BookFinder to generate targeted lists (e.g. cross-group Asian, or cross-group Latinx) to identify titles for inclusion.

Individual reports will be confidential, but submitted data will be held in a repository and utilized to **benchmark** the number and nature of diverse picture book holdings nationally. This will allow us to generate benchmark documents by national region, service area type, and collection budget size. These publicly available benchmark documents will delineate aggregate data, without making mention of specific libraries, their collections or their collection content. As such, these benchmark documents will provide useful information to librarians as they consider their collections within broader contexts and facilitate granular and gross tracking of growth and change over time.

**2.b. Search Engine Optimization**. Discoverability — of the website itself through a Google search as well as of the titles it contains — is of paramount importance to the success of the Diverse BookFinder. Funds from the Sparks! Grant enabled us to complete a naming process and to begin Search Engine Optimization (SEO). In regards to the latter, we have worked with a firm to begin understanding synonyms or commonly used words related to our vocabulary terms. For example, SEO work has identified commonly used synonyms for our oppression term, including "civil rights", "slavery", "racism" or "prejudice". We are currently in the process of examining synonyms; some will be used to replace existing terms, others will be embedded in the underlying website URL structure so that their entry would result in a listing of what we call oppression books. This work is scheduled for completion prior to the public website launch in September. This type of SEO work is referred to as "short tail" work in the industry and represents a first step towards maximizing discoverability.

**We are requesting funds to continue SEO to uncover and include what are referred to in the industry as "long tail" terms.** Long tail SEO involves the identification of commonly used phrases or common speech sentences a user might employ to find a website or locate content. If short tail key terms for oppression are "civil rights" or "slavery", long tail terms could be "what is the best children's picture book for 7-year-olds about slavery". Long tail terms are more similar to the more conversational nature that online searches are taking and include words like "what", "tell me", "how" and "how to." Long tail SEO represents an industry best practice and is s critical for the long term sustainability of the project as search and how people find information online is substantively changing. This type of site architecture work will insure findability for years to come. **Our terms were created through an academic process, and, although we feel confident in their efficacy in capturing the content of diverse picture books, they may not represent the words commonly used outside of academia. The requested SEO funds will help us further translate our academic work, making it much more relevant for practitioners and library patrons.**

**2.c. Strategic Dissemination**. We have developed a four pronged dissemination strategy for sharing information with the library community about the enhanced Diverse BookFinder and collection tool. We will initiate dissemination with **direct, face-to-face** instruction in Maine in collaboration with the Maine State Library (see letter of support) through a training session to be held at the state's annual [Reading Round Up](#) conference, the largest gathering of children's librarians in the state (April, 2019). Doing so will enable us to pilot and revise training content and handouts while facilitating dissemination to children's librarians statewide. Additional opportunities for direct dissemination will be explored with the Maine State Youth Services Librarian. We will also disseminate directly through presentations and workshops at national

conferences (e.g. The Association of Library Services for Children (September, 2019) and The American Association of School Librarians (November, 2019)). We will also undertake **virtual** dissemination through a Webinar (to be hosted by COSLA, who has submitted a letter of support) in the fall of 2019. This webinar will facilitate national reach as direct work in other states extends beyond the current budget. We will utilize our Advisory Council of librarians, educators and diverse book advocates to assist with dissemination through **influencer marketing** (more information about the council, its members and their responsibilities can be found in their letters of support). Finally, we will also enhance and expand the initial reach of the project with **targeted marketing** to audiences of public and school librarians via one of the most advanced and widely-used networks world-wide: Facebook.

**Workload Plan. Year 1** -- *(months 1-6)*. SEO work conducted and added to the Diverse BookFinder to offer enhanced discoverability -- longtail work. *(months 6-12)*. Enhancement of book coding procedures and implementation to short and long tail Search Engine Optimization; project coordinator search; initial development of collection analysis tool software. **Year 2 --** *(months 1-3)*. Final collection development tool creation and data repository protocol. *(months 4-7)* beta testing of collection development tool (50 libraries) along with the testing of tools and protocols necessary for data management and benchmarking, including necessary upgrades. *(months 7-12)* Initial dissemination (face-to-face) along with ongoing monitoring and support. **Year 3 --** *(months 1-2)*. Webinar creation and dissemination. *(months 3-12)*. Ongoing dissemination work using four pronged approach. Ongoing data management, benchmarking and dissemination. Our goal is to serve at least 3,000 libraries by the end of the funding period.

**Personnel**. Project Director **Dr. Krista Aronson, Ph.D.** is an Associate Professor of Psychology at Bates College whose work focuses on illuminating how people come to understand complex social constructs like race and ethnicity, including how children process and understand race as well as appropriate, effective and productive ways to discuss this topic with them; specifically, the effective use of picture books to enhance intercultural relationships and self-understanding during childhood. Her own identity as a biracial woman deeply informs her work. The Diverse BookFinder and associated work are the focus of her active research, which represents a substantial portion of the professional effort of faculty at Bates.

**Anne Sibley O'Brien** has served as a long time project consultant and funds are being sought as an honorarium for her continued work in this capacity. Specifically, Annie will serve as an liaison to the Advisory Council and the children's book publishing field, and provide feedback on website development, including SEO products and their website integration. As a co-founder of the Diverse Picture Book Collection, Annie is fully committed to this project and devotes time and energy to the it, on the order of 100-150 hours/year. O'Brien is an award-winning author and illustrator of 36 multicultural children's books whose work has regularly been cited for attention to authentic cultural details. She has also been engaged in education about multicultural literature, diversity and antiracism for more than forty years. Her experience of being raised bilingual and bicultural in South Korea as the daughter of White American medical missionaries was the catalyst for her lifelong focus on human difference.

**Christina Bell**, Humanities Librarian at Bates College's Ladd Library where she performs traditional liaison duties related to reference, instruction, and collection development. She joined the Diverse BookFinder in 2016 as a collection manager and digital project consultant, helping to expand the physical collection of diverse books and to consider how users interact with the technology to find the books. She will play an integral role in all aspects of the proposed work, providing ongoing support for physical collection development, feedback on SEO and collection analysis tool products and their website integration, and targeted marketing support. She will devote approximately 20% of her time to these tasks. Before becoming a librarian, Christina worked in human rights and is deeply committed to information access and diversity. She earned her MLIS from Long Island University and MA in Humanities and Social Thought from New York University.

**Eleven Advisory Council members** — including public and school librarians, children's book authors and scholars, along with one representative from Lee & Low Books (a publisher of multicultural books) — have been invited to serve on an Advisory Council (5 have accepted; see List of Key Project Staff and Consultants for fill list of names; letters of commitment and associated resumes are also included in the application packet). When complete, the Advisory Council will include at least one member each who are: First/Native Nations, African American, Asian American, Latinx, and Muslim. Collectively, the advisors we have invited also represent national organizations and platforms at the center of the discussion on representation in children's books, including We Need Diverse Books, "The Open Book" blog by Lee & Low, and the Reading While White blog, as well as a broad range of positions and activism in critical issues in librarianship.

**Michael Anderson,** website contractor, will develop the collection analysis tool software. Michael has extensive experience building sophisticated, interactive Learning Management Systems, including one work safety training program (simple courses and quizzes) designed to keep 900 employees up to date. Experienced with multiple web based formats, including WordPress and Solar (the formats undergirding the Diverse BookFinder), Michael has worked closely with Bates to insure the accuracy of project conceptualization and associated quote.

**Jane Harrell**, President of 'cause Digital Marketing, Senior Digital Strategist and Consumer Advocate will serve as our SEO specialist. Jane's passion for helping cause-based businesses and focused understanding of the online space has earned her a reputation for innovation, expertise and proven results with organizations like Discovery Communications, Petfinder.com, Purina and IDEXX Laboratories. Jane graduated Cum Laude and was a constant member of the Gold Key Honor Society at Emerson College, where she earned a BA in *Advocacy Writing Across Cultures*, a self-designed major that focused on intercultural communication, advocacy and marketing.

The **Full Time Project Coordinator** (see job posting) will take over the book coding tasks conducted by RAs in year one. Specifically, collection additions need to be coded using our vocabulary, entered into the database and transmitted to Bates ILS for quarterly updates. In addition to this ongoing work, the coordinator will play an important role in carrying out tasks essential to years 2 and 3 of the grant, including: benchmark data repository oversight;

responding to basic questions during beta testing, and conveying any user difficulties to the programmer for revision/upgrades; assistance with the creation and maintenance of Diverse BookFinder social media and the implementation of targeted marketing protocols (FaceBook add purchases and monitoring). In addition, they will manage project communications, including email and phone response to Diverse BookFinder users. Existing requests from librarians and the public have already become challenging for project personnel to manage given their other responsibilities (e.g. teaching commitments and general library reference). As we broaden our discoverability and national reach, the timely communication and responsiveness a coordinator can offer will be key to overall project success.

**Resources and sustainability.** The Diverse BookFinder and the Diverse Book Collection are Bates-supported services with a high level of institutional commitment. The Diverse Book Collection has dedicated space on the first floor of Ladd Library on the Bates College campus and all books are available for circulation via direct lending or ILL. The Collection has an independent funding line for the annual addition of books. Likewise, the Diverse BookFinder has secure funding to support website maintenance, including quarterly updating, for the foreseeable future (See Bates College Letter of Support and Digital Product Form for more information).

**3. National Impact.** A National Leadership Grant will enable us to add crucial features to the Diverse BookFinder, and engage librarians in ongoing collaboration in a nationally important topic. The movement to increase numbers of and access to books featuring people of color has gained significant momentum in recent years, engaging librarians, authors, publishers, parents, and child readers (e.g. We need Diverse Books; Reading While White). Take for instance Marley Dias, the 11-year-old elementary school student in Germantown, New Jersey and founder of the 1000 Black Girl Book Project. Discouraged by the absence of books with characters who looked like her available in her school library and within her school's reading curriculum, she started a drive to collect books featuring characters of African descent, a movement which went viral and for which Dias has received multiple honors and awards. The online resource hub and the Bates College library catalog will provide a list of picture books for girls like Dias, her parents, publishers, authors and librarians, and a system by which it can be effectively navigated.

The Diverse BookFinder is designed to support librarians as they strive to build diverse collections that reflect their communities and the nation. The proposed Collection Analysis tool will allow librarians to understand the number and nature of diverse picture books in their collections. Then, the enhanced Diverse BookFinder will allow librarians to search our collection to identify targeted acquisitions, **supporting libraries in their roles as community anchors, helping them to more completely connect with and engage as they "facilitate discovery of knowledge and cultural heritage"** (IMLS, 2017).

The enhanced BookFinder has the potential to change how children's books are identified, described and utilized. The online format offers the potential to reach a national audience, including public librarians, school librarians, K-12 teachers, higher education faculty and students, catalogers of children's materials along with children and parents themselves providing access to the physical books if they are not available locally.

**SCHEDULE OF COMPLETION**

| | Oct-Dec 2017 | Jan-Mar 2018 | Apr-Jun 2018 | Jul-Sep 2018 | Oct-Dec 2018 | Jan-Mar 2019 | Apr-Jun 2019 | Jul-Sep 2019 | Oct-Dec 2019 | Jan-Mar 2020 | Apr-Jun 2020 | Jul-Sep 2020 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| SEO long-tail work | ■ | ■ | | | | | | | | | | |
| Enhancement of book coding procedures and implementation to short and long tail SEO | | | ■ | ■ | | | | | | | | |
| Project coordinator search | | | ■ | ■ | | | | | | | | |
| Initial development of collection analysis tool software | | | ■ | ■ | | | | | | | | |
| Final collection development tool creation and data repository protocol | | | | | ■ | | | | | | | |
| Collection development tool beta testing, testing of tools and protocols for data management and benchmarking, necessary upgrades | | | | | | ■ | ■ | | | | | |
| Initial dissemination (face-to-face) along with ongoing monitoring and support | | | | | | | ■ | ■ | | | | |
| Webinar creation and initial presentations | | | | | | | | | ■ | | | |
| Ongoing dissemination work using four pronged approach | | | | | | | | | | ■ | ■ | ■ |
| Ongoing data management, benchmarking and dissemination | | | | | | | | | | ■ | ■ | ■ |

**DIGITAL PRODUCT FORM**

**Introduction**
The Institute of Museum and Library Services (IMLS) is committed to expanding public access to federally funded digital products (i.e., digital content, resources, assets, software, and datasets). The products you create with IMLS funding require careful stewardship to protect and enhance their value, and they should be freely and readily available for use and re-use by libraries, archives, museums, and the public. However, applying these principles to the development and management of digital products can be challenging. Because technology is dynamic and because we do not want to inhibit innovation, we do not want to prescribe set standards and practices that could become quickly outdated. Instead, we ask that you answer questions that address specific aspects of creating and managing digital products. Like all components of your IMLS application, your answers will be used by IMLS staff and by expert peer reviewers to evaluate your application, and they will be important in determining whether your project will be funded.

**Instructions**
You must provide answers to the questions in Part I. In addition, you must also complete at least one of the subsequent sections. If you intend to create or collect digital content, resources, or assets, complete Part II. If you intend to develop software, complete Part III. If you intend to create a dataset, complete Part IV.

## PART I: Intellectual Property Rights and Permissions

**A.1** What will be the intellectual property status of the digital products (content, resources, assets, software, or datasets) you intend to create? Who will hold the copyright(s)? How will you explain property rights and permissions to potential users (for example, by assigning a non-restrictive license such as BSD, GNU, MIT, or Creative Commons to the product)? Explain and justify your licensing selections.

Bates College will hold the copyright. In line with the existing database, we will assign further work a Creative Commons Attribution- NonCommercial 4.0 International license.

**A.2** What ownership rights will your organization assert over the new digital products and what conditions will you impose on access and use? Explain and justify any terms of access and conditions of use and detail how you will notify potential users about relevant terms or conditions.

We will use the CC indicator on the website to notify users of the copyright status.

**A.3** If you will create any products that may involve privacy concerns, require obtaining permissions or rights, or raise any cultural sensitivities, describe the issues and how you plan to address them.

There are no concerns for the existing website and database. Users of the Collection Analysis Tool will create a login to access detailed reports, at which point they will share their email address and respond to a set of demographic questions about their library (state, region of the country, type, size, budget). No other personal information will be required. All of this information will be confidential and stored in Bates College network file storage, and will be stripped from any publicly available data.

## Part II: Projects Creating or Collecting Digital Content, Resources, or Assets

### A. Creating or Collecting New Digital Content, Resources, or Assets

**A.1** Describe the digital content, resources, or assets you will create or collect, the quantities of each type, and format you will use.

We will expand the existing website and database to create an analytical tool for librarians to identify

diverse titles in their collections and gaps in representation. The public-facing website will be on the WordPress platform, the underlying data are collected in Excel and imported into the CMS platform.

**A.2** List the equipment, software, and supplies that you will use to create the content, resources, or assets, or the name of the service provider that will perform the work.

Computer hardware and software (WordPress, Excel, Solr) will be provided through Bates College's Information and Library Services (ILS) department. At present services related to the database and website will be created and managed through Bates, and the SEO work will be created and managed by the existing project contractor 'Cause Digital Marketing. Additional programming work will be done by programmer Michael Andrews in consultation with Bates IT services.

**A.3** List all the digital file formats (e.g., XML, TIFF, MPEG) you plan to use, along with the relevant information about the appropriate quality standards (e.g., resolution, sampling rate, or pixel dimensions).

Collection data provided by Libraries will be provided in .csv format. Transfer of that csv data will occur via HTTP form post along with other information necessary for processing. Reports generated will be in a common format supported by web browsers such as .pdf, HTML, .png, etc.

## B. Workflow and Asset Maintenance/Preservation

**B.1** Describe your quality control plan (i.e., how you will monitor and evaluate your workflow and products).

The lead investigator, co- principal investigator and project consultants with review and monitor content and workflow prior to public availability. The full time Project Coordinator will manage the products and inquiries beginning in Year 2 of the work under the direction of the lead investigator.

**B.2** Describe your plan for preserving and maintaining digital assets during and after the award period of performance. Your plan may address storage systems, shared repositories, technical documentation, migration planning, and commitment of organizational funding for these purposes. Please note: You may charge the federal award before closeout for the costs of publication or sharing of research results if the costs are not incurred during the period of performance of the federal award (see 2 C.F.R. § 200.461).

Backup and working data files are managed by co-PI Krista Aronson. Storage of publicly available content will occur on servers owned or licensed by Bates College and file maintenance will be managed and/ or facilitated by Bates ILS staff.

## C. Metadata

**C.1** Describe how you will produce any and all technical, descriptive, administrative, or preservation metadata. Specify which standards you will use for the metadata structure (e.g., MARC, Dublin Core, Encoded Archival Description, PBCore, PREMIS) and metadata content (e.g., thesauri).

No new metadata will be created, as the results of the project rely upon existing metadata content and structures (in MARC and XML).

**C.2** Explain your strategy for preserving and maintaining metadata created or collected during and after the award period of performance.

N/A

**C.3** Explain what metadata sharing and/or other strategies you will use to facilitate widespread discovery and use of the digital content, resources, or assets created during your project (e.g., an API [Application Programming Interface], contributions to a digital platform, or other ways you might enable batch queries and retrieval of metadata).

N/A

**D. Access and Use**

**D.1** Describe how you will make the digital content, resources, or assets available to the public. Include details such as the delivery strategy (e.g., openly available online, available to specified audiences) and underlying hardware/software platforms and infrastructure (e.g., specific digital repository software or leased services, accessibility via standard web browsers, requirements for special software tools in order to use the content).

The website that hosts the book collection and search will be freely available through the Internet, accessible using standard web browsers. Access to the analytical tool will be available by request and via login embedded in the WordPress site.

**D.2** Provide the name(s) and URL(s) (Uniform Resource Locator) for any examples of previous digital content, resources, or assets your organization has created.

The Bates College Digital Campus group (co-chaired by Christina Bell) consults on campus digital projects and lists projects here: http://www.bates.edu/digital-campus/projects/

## Part III. Projects Developing Software

### A. General Information

**A.1** Describe the software you intend to create, including a summary of the major functions it will perform and the intended primary audience(s) it will serve.

This project will develop a free online collection analysis tool to provide such information in narrative, graphical and tabular form. Librarians will upload a .csv file with ISBN information about their picture book (or in some cases their children's book) collection. The online collection analysis tool will ingest files of these forms through a web- based interface. Librarians interested in receiving a detailed report of their diverse books will be asked to create a login, at which point they will share their email address and respond to a set of demographic questions about their library (state, region of the country). Information contained within this data file will then be cross-referenced with master data for books contained within the Bates College Diverse Book Collection. This matching process is necessary because existing catalog information related to race, ethnicity and culture is insufficient and applied inconsistently as described in the above narrative. Comparing each library list with the Diverse Book Collection allows us to extract a representative sample of titles from which to create a report. Following the matching process, the percentage of racially diverse picture books for each library will be calculated by dividing the number of matching titles between the library's shelf list and the Bates Diverse Book Collection list by the total number of individual titles in that library's shelf list. As many of the datasets contributed by libraries will contain multiple duplicates of the same title, the percentage will calculated using each library's shelf list after the duplicates have been parsed out. The resulting report will convey several pieces of information important to collection development: the number and

titles of diverse titles and their percentage of the overall whole; what groups are depicted within these titles and their relative percentage; and, how groups are depicted. Once the report has been generated the librarian will be provided access to a  report.


**A.2** List other existing software that wholly or partially performs the same functions, and explain how the software you intend to create is different, and justify why those differences are significant and necessary.

While there are web-based software tools that serve analytical functions and curated lists of diverse books are readily available, there is no tool freely available that allows librarians to analyze an existing collection for diverse titles. We will adapt open source tools to build out a unique functionality.

**B. Technical Information**

**B.1** List the programming languages, platforms, software, or other applications you will use to create your software and explain why you chose them.

Work to date has been written in PHP utilizing WordPress as the CMS/Programming framework and Apache Solr for search functionality. MARC data from the Bates Library catalog and data from the Google books API was included to expand the data we store and display about each book. All web hosting is on a Linux LAMP stack. The analytical tool will build on this existing work by using R to generate reports.

**B.2** Describe how the software you intend to create will extend or interoperate with relevant existing software.

The analytical tool will build upon work done to date in utilizing the same software while expanding the work with R for reports. We are developing a WordPress plugin to expand the functionality of the CMS to store data about the collection. This plugin will also interface with Apache Solr to increase search functionality.

**B.3** Describe any underlying additional software or system dependencies necessary to run the software you intend to create.

Currently WordPress, Solr, and R. The in development work uses CentOS 7 as the OS, but we're not intentionally building this to be Linux distribution specific.

**B.4** Describe the processes you will use for development, documentation, and for maintaining and updating documentation for users of the software.

We will develop with GIT and allow for a README in the git file for users of the software along with inline comments for different portions of the code.  For end users we will supply descriptions on the screen they are using to complete the task.  For the scope of this project, we believe a custom post type with inline descriptions and possibly a link to a longer description could be built into the code.

**B.5** Provide the name(s) and URL(s) for examples of any previous software your organization has created.

Bates has used both WordPress and Solr as platforms for development, such as at
http://www.bates.edu/faculty-expertise/

**C. Access and Use**

**C.1** We expect applicants seeking federal funds for software to develop and release these products under open-source licenses to maximize access and promote reuse. What ownership rights will your organization assert over the software you intend to create, and what conditions will you impose on its access and use? Identify and explain the license under which you will release source code for the software you develop (e.g., BSD, GNU, or MIT software licenses). Explain and justify any prohibitive terms or conditions of use or access and detail how you will notify potential users about relevant terms and conditions.

We plan to release any new products created under open source licenses, as Bates has done with past projects under either MIT or GNU license on Github. All relevant terms and conditions will be made available to users when they request a login to the analytical tool and at the point that they input data files.

**C.2** Describe how you will make the software and source code available to the public and/or its intended users.

We will make any new software and code available on the Github repository under an MIT or GNU license.

**C.3** Identify where you will deposit the source code for the software you intend to develop:

Name of publicly accessible source code repository:

URL: https://github.com/

**Part IV: Projects Creating Datasets**
**A.1** Identify the type of data you plan to collect or generate, and the purpose or intended use to which you expect it to be put. Describe the method(s) you will use and the approximate dates or intervals at which you will collect or generate it.

After confidential report generation, data submitted by librarians through the Collection Analysis Tool will be anonymized and held in a repository and utilized to benchmark the number and nature of diverse picture book holdings nationally. This will allow us to generate benchmark documents by national region, service area type, and collection budget size. These publicly available benchmark documents will delineate aggregate data, without making mention of specific libraries, their collections or their collection content. Reports will be generated and shared once during year three of the funding period.

**A.2** Does the proposed data collection or research activity require approval by any internal review panel or institutional review board (IRB)? If so, has the proposed research activity been approved? If not, what is your plan for securing approval?

N/A

**A.3** Will you collect any personally identifiable information (PII), confidential information (e.g., trade secrets), or proprietary information? If so, detail the specific steps you will take to protect such information while you prepare the data files for public release (e.g., data anonymization, data suppression PII, or synthetic data).

Data collected by the analytical tool login process will include: name of librarian, email address, type of library, location/ region, size of library, collections budget. This data will be stored on secure Bates servers accessible only to the co-PIs. All of this information will be confidential and stored in Bates

College network file storage, and we will take reasonable steps to protect data provided by libraries. Only aggregate data will be shared publically. No one individual library or library collection will be identifiable.

**A.4** If you will collect additional documentation, such as consent agreements, along with the data, describe plans for preserving the documentation and ensuring that its relationship to the collected data is maintained.

N/A

**A.5** What methods will you use to collect or generate the data? Provide details about any technical requirements or dependencies that would be necessary for understanding, retrieving, displaying, or processing the dataset(s).

There are two points of collection for data: selection of library demographic data at initial login, and the data generated about individual library collections. Both will occur behind a login on the WordPress site and will only be available to the user requesting a report and the co-PI's. The reports are for individual library information only. We will only use aggregate data for benchmarking.

**A.6** What documentation (e.g., data documentation, codebooks) will you capture or create along with the dataset(s)? Where will the documentation be stored and in what format(s)? How will you permanently associate and manage the documentation with the dataset(s) it describes?

The only other documentation generated will be the reports made available in. pdf format, the data of which will be stored in .csv files as part of the dataset. The GIT repository will include an "example" section with a .csv file and an R response with the codebook to be created by the programmer Michael Anderson.

**A.7** What is your plan for archiving, managing, and disseminating data after the completion of the award-funded project?

The book collection and diversity attribute data will be available via the Diverse BookFinder database and website. The aggregate data collected about library collections will be used for the creation of benchmark reports (previously described) Continued hosting and maintenance will be negotiated with Bates College, which will keep the website, database, and analytical tool publicly active. Bates Library has a commitment to maintaining and growing the book collection that drives the research behind the database. Unless a more topic-specific platform is identified, the data will be disseminated via Bates' institutional repository.

**A.8** Identify where you will deposit the dataset(s):

Name of repository: Scholarly Communication And Research At Bates (SCARAB)

URL: http://scarab.bates.edu/

**A.9** When and how frequently will you review this data management plan? How will the implementation be monitored?

Quarterly review of the book data will keep the database and analytical tool current. Broader review of the collected library data will occur annually with the project co-PIs and Bates ILS.