Victoria Van Hyning, University of Maryland, College of Information Studies

Crowdsourced Data: Accuracy, Accessibility, Authority (CDAAA)

Assistant Professor Victoria Van Hyning (PI) of the University of Maryland, College of Information Studies, requests $458,151 to pursue *Crowdsourced Data: Accuracy, Accessibility, Authority* (*CDAAA*), a 3-year Early Career Research Development program to investigate the sociotechnical barriers that Libraries, Archives and Museums (LAMs) face in making their crowdsourcing data derived from cultural heritage materials **open** and **accessible** to a broad public. **Accessibility** here refers not only to the ability of fully sighted users to search LAM discovery systems, also known as Content Management Systems (CMSs), but the ability for people who are Blind, have dyslexia, or other print-related disabilities for which they utilize screen reader software to hear digital information read aloud. When LAMs provide well-structured discovery systems and metadata to meet the access needs of people who use screen readers, they also meet the needs of a much wider audience.

Hundreds of crowdsourcing projects engage millions of internet-connected volunteers around the world (Ridge, 2016; Blickhan, Ferriter, Ridge, et al, 2021). Transcriptions are the most common type of crowdsourcing data solicited by LAMs because, when integrated into CMSs, they broaden access to otherwise non-machine readable images of documents, and fill significant gaps in traditional archival description and metadata practices (Lim & Liew, 2011; Van Hyning, 2019). Many LAMs explicitly invite volunteer transcription to make their collections more discoverable and accessible for screen-reader users, and volunteers are passionate about this aspect of crowdsourcing. Despite the popularity of crowdsourcing projects among the public and LAMs, and the extraordinary utility of searchable text, practitioners experience significant and under-researched challenges to the integration of crowdsourced data that merit urgent attention.[1] Challenges include field character length limits, poorly configured metadata fields, or searches that return too many irrelevant results and cause systems to run slowly or crash. These issues stem in large part from CMS system design, but may also be compounded by as yet-unidentified factors that the proposed research will uncover.

At stake are practical concerns about data preservation and integration, as well as more fundamental questions about authority, accessibility, and accuracy. LAMs need detailed information about the barriers to crowdsourced data integration, as well as concrete examples of successes they can adapt for their own organizations and workflows. Volunteers who dedicate time and effort to transcribe materials deserve for their work to be meaningful and sustainable. People who are Blind, have low-vision or dyslexia, or use a screen reader for another reason are currently some of the most excluded members of society when it comes to accessing cultural heritage through LAMs, and they deserve better access.

To fully realize the early promises of crowdsourcing, we must understand the challenges to data integration and accessibility, and identify scalable solutions. PI Van Hyning will lead a blended **qualitative and quantitative** study over three years to address the following **research questions**:

- **RQ1 (Authority):** Are LAM practitioners able to integrate crowdsourced transcriptions and other textual data such as tags or notes into their CMSs (the authoritative record)? If yes, how? If not, what **technical barriers** do they face?
- **RQ2 (Accuracy and Authority):** Previous studies by Jansson (2017) and Liew (2016) reveal numerous social barriers to the integration of crowdsourcing data, most significantly, LAM practitioners' anxieties about data quality and whether volunteers can be trusted to do accurate description and transcription work. Is this anxiety widespread, and does it impact whether or not crowdsourced data are incorporated into CMSs? How do LAM practitioners determine the quality of crowdsourced data?
- **RQ3 (Accessibility):** When transcription data is successfully integrated with CMSs, is it accessible for people who use screen readers, and if not, what is required to make the data legible? What are screen-reader users' experiences of crowdsourced data that is integrated into CMSs?

CDAAA addresses these research questions to advance knowledge and understanding of the sociotechnical barriers to transferring data from crowdsourcing platforms into core LAM CMSs. In doing so, we will help

[1] Van Hyning, Victoria and Mason A. Jones. "Data's Destinations: Three Case Studies in Crowdsourced Transcription Data Management and Dissemination." *Startwords*, no. 2 (December 1, 2021). https://doi.org/10.5281/zenodo.5750691; K. Crowe, K. Fenlon, H Frisch, D Marsh, and Van Hyning. "Inviting and Honoring User-contributed Content" in *The Lighting the Way Handbook: Case Studies, Guidelines, and Emergent Futures for Archival Discovery and Delivery*. Practitioner Handbook. Stanford, October 22, 2021. https://doi.org/10.25740/gg453cv6438; Van Hyning. "Harnessing Crowdsourcing for Scholarly and GLAM Purposes." *Literature Compass* 16, no. 3–4 (2019): e12507. https://doi.org/10.1111/lic3.12507.

LAMs understand how to reap the full benefits of crowdsourcing, including increased collections discovery and utility, as well as an improved understanding of how to better serve the members of society with extremely limited access to cultural heritage (Brilmyer, 2018). The Information Science, Library Science, and Accessibility-focused User-Centered Design approaches we will deploy are the ideal blend of disciplines and methods with which to identify challenges and provide scalable solutions. LAMs' ability to engage in maximally impactful and sustainable crowdsourcing is predicated on the research proposed here.

CDAAA aligns with LB21 Goal 2: Objective 2.3, to "support the research of non tenured tenure-track library and information science faculty, furthering the faculty member's long-term research agenda, career trajectory, and professional development." PI Van Hyning's primary long-term research goals are to 1) ensure that LAMs can sustainably engage in cooperative knowledge creation with their users, and 2) make LAM data discoverable, accessible and usable for the widest range of people. Her research and LAM practice have already shaped several leading crowdsourcing platforms and LAM data management workflows that impact tens of thousands of users and hundreds of LAMs. *CDAAA* will enable the PI to expand her research capacity, deepen and broaden her LAM networks, and gain experience supervising her own research team, consisting of a graduate assistant (GA) in the Recovering and Reusing Archival Data (RRAD) Lab (co-founded by the PI in 2021), and an accessibility specialist (Assistant Research Scientist) at UMD iSchool's Trace Center. The research questions at the heart of *CDAAA* firmly align with the work of both labs. *CDAAA* will also contribute to LB21 Goal 3, to "enhance the training and professional development of the library and archival workforce," by supporting the GA to become an information professional, and by supporting LAM practitioners working in digital curation with concrete, practitioner-oriented recommendations for better data integration.

Project Justification:

In the "Addressing Societal and Scientific Challenges through Citizen Science and Crowdsourcing" memorandum (2015), the Obama administration urged Federal agencies to engage the public in participatory activities ranging from document transcription, to water quality monitoring, to mapping the surface of Mars. Successful crowdsourcing projects at NASA (Clickworkers, est. 2000), the Smithsonian (Smithsonian Transcription Center, est. 2012), the National Archives and Records Administration (NARA, Citizen Archivist, est. 2011), and elsewhere laid the groundwork for this directive. One of the memo's three guiding principles is "Openness" of resulting data: "**Data worth collecting and using also are worth preserving and sharing.** Federal agencies should design projects that generate datasets, code, applications, and technologies that are **transparent, open, and available to the public** [... and] should use machine-readable formats to share data, metadata, and results with project volunteers and the general public" (emphasis added). The memo's language registers that crowdsourcing's end goals include not only public engagement but, equally important, creating and preserving accessible, reusable data. A guiding light for crowdsourcing and LAM practitioners in government, academia, and beyond, the 2015 memorandum spurred significant public and private investment in crowdsourcing platforms and engagement programs across disciplines, as well as annual conferences and an interagency toolkit for government workers to help them implement crowdsourcing.

Many LAMs have embraced crowdsourcing without knowing how difficult it can be to preserve transcriptions, tags, and other crowdsourced data in most CMSs, which often lack the fundamental structures to store page-level metadata. More recently, a small number of organizations have begun to openly discuss crowdsourcing challenges and solutions, but as yet no systematic study exists of the challenges and successes across different kinds of LAMs and collections. In order to maximize the value of previous investments in crowdsourcing, and to help LAMs achieve their full potential to connect diverse users with LAM content, we must act now to investigate these complex data challenges.

With 8 years of industry and research experience in LAM crowdsourcing, a deep network of LAM collaborator-practitioners and crowdsourcing platform creators, and a strong interdisciplinary culture at UMD's iSchool, PI Van Hyning is uniquely well-placed to lead this project (see Trevor Owens, letter of support). As Humanities PI of Zooniverse from 2015 to 2018, and a Senior Innovation Specialist at the Library of Congress from 2018 to 2020, where she co-created By the People, Van Hyning has experienced LAMs' crowdsourcing data integration challenges firsthand and conducted preliminary research on potential solutions. IMLS funded Zooniverse's "Transforming Libraries and Archives through Crowdsourcing" grant for $1.27 million (2016-19), which supported the creation of 4 bespoke transcription projects for LAMs, allowing Zooniverse to build and systematically test the efficiency and accuracy of new audio and transcription tools for cultural heritage institutions. Successful tools were added to the free Zooniverse Project Builder platform, and have supported more than 50 transcription projects.

Van Hyning was the lead architect of "Transforming's" research questions, and ensured that the project centered the needs of LAM users. Zooniverse accomplished a great deal through the grant effort, including creating simpler data exports and better documentation for those outputs. Our work also revealed that creating usable crowdsourcing data outputs is not enough in itself to solve the complexities of getting crowdsourced data into the authoritative LAM record. The sheer variety of LAM CMSs and data practices, in combination with the variety of crowdsourcing outputs, represents enormous, but not insurmountable, challenges to delivering on the promises of crowdsourcing.

CDAAA will dramatically extend a small but significant body of published research that has sounded the alarm about the gaps between crowdsourcing's ideals and the realities of preserving crowdsourced data. Blaser identified technical limitations, such as the absence of appropriate metadata fields to incorporate Zooniverse *Old Weather* transcriptions, in 2014, and the PI has since witnessed more than a dozen institutions and projects encounter this same problem. Jansson (2017) and Liew (2016) cite widespread anxieties among LAM practitioners and the public about the **accuracy** of crowdsourced data and its place in the **authoritative** record. The literature and the PI's experience suggests that these anxieties are the most common social challenge that keeps crowdsourced data out of the authoritative record. Bowser et al. (2020) expose the absence of norms for data structure, description, and integration across citizen science more broadly, calling attention to the ethical problems when institutions do not address these challenges, release crowdsourced data, or even have a data management plan. One of the first resources for practitioners to mention this issue, Blickhan et al (2021) speaks directly to the ongoing technical challenges facing LAMs, warning that "roundtripping [crowdsourced] data, including data synchronization across collections and databases [...] is still beyond the capabilities of many [CMSs]." While it was beyond the scope of that book to offer detailed solutions, doing so is precisely the ambitious goal of *CDAAA*.

By addressing crowdsourcing data integration challenges, *CDAAA* will contribute to a broader scholarly and practitioner-oriented literature about the challenges of data preservation, reuse and management. This research is exemplified by Borgman's *Big Data, Little Data, No Data* (2015), examining the complex, fragmented academic research and LAM cultures that pose significant sociotechnical problems to data sharing, preservation, publication, and reuse. The work of IMLS Early Career Fellow Dr Katrina Fenlon (co-founder of RRAD Lab at UMD) explores related challenges to the sustainability of digital community collections created outside of academic and cultural heritage institutions. Fenlon, PI Van Hyning, and colleagues draw new connections between the urgent need to address data creation, preservation, and reuse in LAM crowdsourcing, and digital community collections, in our co-authored piece (with Crowe, Frisch, and Marsh, 2021) "Inviting and Honoring User-contributed Content," a practitioner-oriented contribution to the IMLS-funded *Lighting the Way Handbook*. *CDAAA* as a whole will be practitioner-oriented, seeking model solutions from crowdsourced data preservation that can be applied widely within LAM contexts.

The Project Team

Dr. Victoria Van Hyning (PI). The PI has extensive knowledge and experience in the full lifecycle of crowdsourcing, which permeates her research, pedagogy, and service to the LAM field (Van Hyning, 2015; Van Hyning, 2019; Blickhan et al, 2019; Ferriter et al, 2019; Van Hyning et al, 2021; Van Hyning and Jones, 2021). As a postdoctoral fellow and Humanities PI of Zooniverse (2014-2018), Van Hyning developed new methodologies for cultural heritage crowdsourcing. She pioneered text transcription and data aggregation methods for *Shakespeare's World* (Folger Shakespeare Library and *Oxford English Dictionary*), *AnnoTate* (Tate), *Decoding the Civil War* (Huntington), and other LAMs. She coathored the IMLS-funded grant "Transforming" (described above), guided the development of new transcription methods for *Scribes of the Cairo Geniza* and *Anti-Slavery Manuscripts*, and co-conducted the first study to assess the accuracy of Zooniverse transcription methods (Blickhan et al. 2019).

As a Senior Innovation Specialist at the Library of Congress, she co-created *By the People* (2018-20), a highly successful crowdsourcing project in terms of engagement, pages transcribed, and pages integrated into the LOC CMS. Working with technologists, archivists, librarians, scholars, and accessibility specialists, she led the successful initiative to integrate *BTP* crowdsourced transcriptions, and is the lead author of a data paper about the

BTP transcriptions for *The Journal of Open Humanities Data*, which aims to publicize dataset creation and opportunities for reuse.

The PI incorporates crowdsourcing methodologies, project-building (i.e. *David C. Driskell Papers Project*, built on FTP for the Driskell Center, UMD), and modules about data integration into her undergraduate and graduate teaching in the Information Science undergraduate major, Masters of Library and Information Science, Human-Computer Interaction Masters, and Information Science doctoral programs. She is the co-founder of the Recovering and Reusing Archival Data (RRAD) Lab at UMD, where she leads research about LAM data preservation and reuse.

Mason Jones (GA Y1-Y3): Mason Jones will be a second-year Ph.D. student in UMD's iSchool at the beginning of the grant period. He holds an MLIS from The University of Alabama with a concentration in archival studies. His experience includes working in Auburn University's Archives & Special Collections, The University of Alabama's Gorgas Library, and The University of Alabama's College of Communication and Information Sciences' Library Commons. Jones also worked as a Research Librarian - Student Trainee at The U.S. Government Accountability Office (GAO). The GA will work closely with the PI to conduct the proposed surveys, interviews, and demonstration sessions with **LAM Partners** in Y1-Y3; work closely with the PI and ARS to conduct User-Centered Design testing sessions with **Screen Reader Users** in Y2; and collaborate with the PI and ARS to write up the research findings (articles and papers), and the practitioner-oriented findings (GitHub reports and a white paper) for publication in Y3. Grant work will support the GA in the process of researching and writing his doctoral thesis, a adjacent and significant outcome of the proposed work.

Dr. J. Bern Jordan (ARS, Y2) is an Assistant Research Scientist at UMD iSchool's Trace Center. He holds a Master's of Science and a PhD in Biomedical Engineering with a focus in technology, disabilities, aging, and accessibility. He was appointed to the U.S. Federal Communication Commission (FCC) Disability Advisory Committee for 2020-22. Jordan has provided analysis on accessibility regulations, including the Section 508 refresh. The ARS will work closely with the PI and GA to recruit 12 **Screen Reader Users.**

The **Trace Center** is a key collaborator in this effort. Trace was created in 1971 to unlock technology's potential to better serve people experiencing barriers due to disability, aging, or digital literacy, and to prevent emerging technologies from creating new barriers. The Trace Center has built trust with a broad community of people who live with a variety of disabilities, and will offer appropriate guidance and support for *CDAAA* to recruit screen-reader users for the project in Y2.

An **Advisory Board** of five specialists will guide the project. Board members will provide feedback on research protocol design, help disseminate information about the project through their networks, and review drafts of the research products.

Confirmed Advisory Board members include:

1. **Dr Samantha Blickhan** is the Zooniverse Humanities Lead and co-Director of the Adler Planetarium Zooniverse team, leading research and development for Zooniverse Humanities projects and tools (see Letter of Commitment).
2. **Ben Brumfield** is the co-creator of FromThePage, supporting over 110 LAM institutions to create meaningful and sustainable crowdsourcing projects (see Letter of Commitment).
3. **Professor Jonathan Lazar** is the Director of the Trace Center, and is focused on Information and Communications Technology accessibility for people with disabilities, user-centered design methods, assistive technologies, and law and public policy related to accessibility and Human-Computer Interaction.
4. **Zuhair Mahmoud** is a Washington DC-based Blind accessibility professional and usability expert, formerly of the Library of Congress.
5. **M.A. Matienzo**, is the Assistant Director for Digital Strategy and Access at Stanford University Libraries, and an expert in metadata aggregation and the design of archival discovery systems.

Crowdsourcing Platforms and LAM Partners:

When LAMs first began to participate in crowdsourcing in the early 2000s and 2010s, there were few off-the-shelf platforms they could use. Some LAMs, such as NYPL and NARA, created crowdsourcing systems from scratch. Others, such as the Folger, partnered with to create bespoke projects, which informed the development of transcription tools on the free Zooniverse Project Builder (launched in 2016) that enables LAMs to create their own projects from a template. Since 2015, the number of Zooniverse transcription projects has grown to ~70, and registered volunteers have more than doubled to nearly 2.5 million.

FromThePage (FTP) is another significant transcription and indexing platform explicitly designed for cultural heritage materials that has hosted 110 institutions and >25,000 volunteers, who have transcribed >1.3 million pages. Although some LAMs (NARA, Smithsonian, Library of Congress) still use their own bespoke crowdsourcing systems, Zooniverse and FTP host a majority share of all active transcription projects in English. *CDAAA* will benefit from the guidance and expertise of Blickhan (Zooniverse), and Brumfield (FTP), who will serve on the Advisory Board and help disseminate information about the project to their LAM users (see Project Work Plan).

CDAAA will collaborate with **12 LAM Partners**, including four that are confirmed and described below, and eight that will be recruited in Y1. Each **LAM Partner** will agree to take part in the designated research activities, including two surveys, a semi-structured interview, a CMS demonstration, and CMS usability and accessibility testing sessions (described in the Project Work Plan). The following four LAM Partners have been selected to ensure a diverse and representative range of LAMs, crowdsourcing platforms, community stakeholders, CMSs and collection types at all stages of this project. Eight additional LAMs of varying size, mission, and collection focus will enable us to diversify our research even further, and maximize the benefit of the research outcomes to the broadest LAM audience. Partners' current challenges, described briefly below, exemplify the crowdsourcing data complexities this project will address. Partner selection criteria for all 12 LAMs include:

1. The LAM has run one or more crowdsourcing transcription projects within the last five years.
2. The LAM has used Zooniverse, FTP and/or another bespoke or off-the-shelf platform for their project.
3. One or more of these projects were created with the express intention of engaging audiences with collections by authors or about populations who have been historically marginalized on the basis of race, gender and/or religious identity.
4. The LAM has attempted to ingest transcriptions into their CMS or has concluded that their current CMS will not accommodate the data, and is exploring a change of CMS in order to accommodate the data.

Confirmed LAM partners (See Letters of Support):

The David C. Driskell Center (UMD) is a small, highly specialized LAM embedded in a university that created the David C. Driskell Papers on FTP to increase collection accessibility and audience engagement. The Center currently uses the Past Perfect CMS, which works well for art objects but does not support page-level metadata, and consequently cannot support transcription integration. The Driskell Center partnership will help us understand the needs of institutions with diverse object and archival holdings, and the role of existing crowdsourcing data in helping LAMs choose a new CMS.

Special Collections and University Archives (SCUA, UMD) is a medium-sized repository comparable to those of many U.S. colleges and universities. They use ArchiveSpace, one of the most promising CMSs for text integration and a major vendor for the field. SCUA has conducted crowdsourcing transcription projects since 2015 and possesses thousands of transcriptions from a bespoke crowdsourcing platform (now defunct) generated by student and volunteer efforts. In Spring 2022, SCUA is working with the PI and 15 undergraduates to build a new project on FTP featuring 19th-20thC English language manuscripts from a Quaker family archive. Currently, SCUA is migrating all their digitized content and metadata, including transcriptions, from Fedora to ArchiveSpace. The SCUA partnership will help us understand the needs of institutions with multiple sources of crowdsourced transcriptions, and a widely used CMS.

The Folger Shakespeare Library is a small LAM with older multilingual manuscript and printed collections, and an advanced scholarly audience focused in the early modern period (1500-1800). Folger staff advocate for radical changes to traditional archival discovery tools, and coined the term "full-text finding aids" to describe CMSs that incorporate transcriptions to enable full-text search. Folger has one of the most complex histories of crowdsourcing, and consequently complex data needs. Folger uses an in-house transcription tool (Dromio) for teaching purposes, Zooniverse *Shakespeare's World* for broader public engagement (2015-2019), and FTP (2021-present) to edit Zooniverse transcriptions, and for public engagement. Folger will migrate all their data to a new CMS during the course of this grant, and will benefit from accessibility-testing. The Folger partnership will help us understand the needs of institutions with multiple sources of crowdsourced transcriptions, a complex data migration, and how well screen-reader technologies read early modern texts.

The Tennessee Genealogy Indexing Project is a volunteer-run organization that crowdsources genealogical data held by various LAMs. Started in 2021, the project recruits, trains, and empowers citizen researchers to participate in the indexing and transcription of Tennessee-related historical documents on FTP. The project was created by Taneya Y. Koonce, MSLS, MPH, and Billie McNamara, MS, and is conducted in partnership with the Tennessee State Library and Archives (TSLA). All crowdsourced data will be ingested to the TSLA CMS to ensure it is freely available in perpetuity. Koonce and McNamara have decades of experience leading and collaborating in online volunteer-driven genealogical and historical projects, including the USGenWeb Project, and, through their networks, are connecting the indexing project to multiple organizations and societies in Tennessee. McNamara is the current and founding State Coordinator of the TNGenWeb Project, and Koonce is President of the Nashville Chapter of the Afro-American Historical and Genealogical Society. This network of community and LAM partners are vital to the *CDAAA* team's understanding of volunteers' perspectives on the value and utility of crowdsourced data.

Project Work Plan:

From the Recovering and Reusing Archival Data (RRAD) Lab at UMD, Van Hyning will plan, execute, and manage the project, leading the RRAD Lab GA and Trace Center ARS in a blended qualitative and quantitative approach. The PI will manage all aspects of the project using the Agile project management methodology, which she has used professionally for the last 5 years to manage web development, crowdsourcing projects, and (since 2021) the RRAD Lab.

The PI and GA will conduct the majority of the work in Y1-Y3. The ARS will provide two months in Y2 of targeted expertise and support to design the usability-testing protocol, recruit participants, and co-conduct a subset of the usability-testing sessions with the **Screen Reader Users**. The PI has experience developing and managing IRB protocols and has begun to scope the IRB package for the proposed research; in accordance with institutional policy, IRB approval will be pursued at time of the award. In Y1-Y3, we will disseminate our findings through relevant research and practitioner conferences, associated workshops, seminars, and invited talks (the PI and ARS each typically deliver 3-5 invited talks each year).

Methods and theoretical framing: The central goal of *CDAAA* is to devise recommendations for best practices that enable LAMs to deliver on the promises of crowdsourcing and improve **accessibility** for people who use screen readers **(target users)**, and internet-connected LAM users more broadly **(beneficiaries)**. Drawing on the PI's interdisciplinary background, and supported by the rich interdisciplinary environment of the UMD iSchool, *CDAAA* will employ **qualitative** and **quantitative methods** including a **landscape review (Y1), survey (Y1)**, **semi-structured interviews (Y1-Y2)**, **usability testing (Y1-Y2), and a follow-up survey (Y3)** with **two target groups**:

1. **LAM Partners**: We will recruit 8 additional LAM practitioners from among survey participants to join the **4 confirmed LAM Partners** described above. The 12 LAM Partners form a **purposive sample** for the interviews and demonstration sessions.
2. **Screen Reader Users:** We will recruit 12 people who use screen readers to test the accessibility and discoverability of transcriptions in **the LAM Partners'** CMSs by a) attempting to find a document with transcribed text following prompts we provide and b) reading it with their preferred screen reader.

This two-group, blended approach is appropriate because it will:

1. Reveal the full life-cycle of crowdsourcing data ingest and use in LAM CMSs;
2. Enable us to identify successes, pain points, and opportunities for our **target group** of individual LAMs to improve their integration and accessibility of transcription data; and
3. Extrapolate generalized guidance for the wider LAM community **(beneficiaries)**.

Year 1:

A **landscape review** in month 1 of Y1 will be conducted by the PI and GA to capture all recent publications (2021-2022) about LAMs' efforts to ingest crowdsourced data into CMSs. A prior literature review conducted in support of "Inviting and Honoring User-contributed Content" for the *Lighting the Way Handbook* means our initial review period can be short. Using previously successful search phrases such as "ingest user-generated content cultural heritage," we will conduct a systematic review of twenty publications most likely to contain recent articles about crowdsourced data integration. Indicative journals include *Archival Outlook*, *Citizen Science: Theory and Practice*, and the *Journal on Computing and Cultural Heritage*, as well as crowdsourcing and LAM blogs i.e. Old Weather Blog and Transcription Center News | Smithsonian Digital Volunteers.

A **Qualtrics survey** will be created and distributed within the first four months of the project (allowing for IRB review time). Qualitative questions will be designed to elicit respondents' beliefs about the **accuracy**, **authority,** and value of crowdsourced transcriptions; quantitative questions will seek information about each respondent's LAM's crowdsourcing transcription ingest processes. The survey will be used to identify and invite 8 respondents to become **LAM Partners**, and take part in **45-minute semi-structured interviews, 60 minute demonstrations, and a follow-up survey** (described below). The survey and interview responses will provide data for at least one publication (see Project Results).

The **Qualtrics survey** will be sent by Advisory Board members Blickhan (Zooniverse) and Brumfield (FTP) to the LAM practitioners who have created projects on their platforms; and by the project team to prominent LAM and crowdsourcing listservs, i.e. ALA, SAA, JESSE, ALISE, Crowdsourcing. We aim for >60 LAM practitioner respondents (~20% of an estimated 300 existing projects) with direct knowledge of their organization's crowdsourcing projects, transcription data, and CMS ingest efforts (see Appendix A for draft survey questions). Survey questions and methods will draw on the work of Boswer et al. (2020), Jansson (2017), and Liew (2016), and the PI and GA will use SPSS for analysis.

Y1-Y2, each **LAM Partner** will take part in an individual **45-minute semi-structured interview** and a **60-minute demonstration** of the transcription ingest process at their LAM, using real data, their CMS, and any other relevant tools or processes (see Appendices B and C). A central goal of this project is to **increase awareness and transparency** about the difficulties and successes LAMs experience when ingesting crowdsourced data into their CMS (the authoritative record). We therefore seek interviewees who can participate without anonymization or of their LAM, though individual participant names could be anonymized (see Data Management Plan). Interviews will be transcribed through Rev.com, and coded using a qualitative analysis software (NVivo) to enable us to align the coding via inter-rater reliability testing by the PI and GA.

The **demonstration sessions** with **LAM Partners** and the CMS accessibility testing sessions with **Screen Reader Users** (see below) will both deploy **User-Centered Design (UXD) methods**, which are ideal for gathering quantitative data such as whether or not users can complete tasks, and how long tasks take; and qualitative data about users' experiences and mental models of the systems with which they interact. The PI has used these methods at Zooniverse and LOC, and teaches a core undergraduate UXD course at UMD. **Demonstrations** will be formatted as **usability-testing sessions** in which a facilitator (the PI or GA) gives tasks to a participant and observes their behavior, and asks additional clarifying questions as needed. Participants may be encouraged to narrate what they are doing aloud (Moran, 2019).

Year 2:

With the help of the **ARS** and the Trace Center**, 12 Screen Reader Users** will be recruited to test LAM partner CMSs. To ensure continuity of the usability sessions, and maximum utility of the outcomes to our LAM beneficiaries and users with disabilities, the PI, GA, and ARS will co-design the protocol and seek feedback from our Advisory Board, which includes assistive technology, disability, crowdsourcing, and cultural heritage experts. Protocol overview: Each **Accessibility Tester** will be invited to participate in an individual 90-minute usability testing session and asked to locate one transcription in each of 3-4 different **LAM Partner** CMSs.

Participants will use their preferred laptop device, operating system, and screen-reader technology to ensure machine familiarity (Lazar et al 2017). These sessions will employ standard UXD user-testing methods, with particular attention to whether the CMS architecture and transcriptions data adhere to the Web Content Accessibility Standards (WCAG) 2.1 (See Appendix D for draft questions and prompts).

After each interview and demonstration session with **LAM Partners** and usability-testing session with **Screen Reader Users**, we will translate our qualitative data into four standard **User-Centered Design (UXD) outputs**: Empathy Maps, User Personas, Scenarios, and Stories. These outputs enable researchers to distill findings into human-centered recommendations for sociotechnical solutions (Gibbons, 2018). **Empathy Maps** place the participant at the literal center of a visualization split into quadrants labeled "says," "thinks," "does," and "feels." We will populate the map for each participant with findings from the interviews (**LAM Partners**) and usability-testing sessions (**LAM Partners** and **Screen Reader Users**), creating a 1-1 relationship between participant and visualization. Next, we will create **User Personas**, composite, fictionalized users that represent a combination of findings across multiple participants that distill recurring themes. User Personas will enable us to articulate the needs of the **target groups** (**LAM Partners** working with crowdsourced data, and **people who use screen readers**) and connect findings to a wider set of **beneficiaries** (i.e. metadata librarians, community stakeholders, crowdsourcing volunteers, CMS vendors, and people who conduct research using LAM collections). **User Scenarios** and **Stories** help bring Personas to life, by providing a fictional backstory and motivations to explain why they use or are discouraged from using a given tool or system. These methods are used throughout the Design world (including LAMs) to identify opportunities for change and improvement, and can help institutions create a set of design and engagement principles (Library of Congress, Concordia, 2019).

Y2-Y3: In Y2-Y3 we will write **12 Individualized LAM Partner Reports** (6-10 pages each including UXD materials: Empathy Maps, Personas, Scenarios, and Stories). To ensure that the perspectives and contributions of our target groups have been accurately represented, we will invite **LAM Partners** to review a draft of the report for their institution, and **Screen Reader Users** to review a draft of the report for each institution whose CMS they tested. The participant feedback period will last one month. After incorporating participant feedback, we will finalize each of the 12 reports and disseminate these to all participants and the broader public, and write a 30 page **summative white paper** of the findings geared to a LAM practitioner audience (see Projects Results and Schedule of Completion).

Year 3:

We will **measure the success** of our practitioner-oriented research products (individualized LAM Reports and summative white paper) by inviting our 12 LAM Partners to complete a **Qualtrics survey** designed to test whether *CDAAA* has addressed the original Research Questions. The survey will ask how our recommendations have impacted LAM Partners' future plans for crowdsourcing approaches, transcription data ingest models, and/or accessibility improvements to their CMSs.

- Do our LAM Partners believe they have sufficient information to overcome technical challenges to integrating crowdsourced data into their CMS, and/or are they able to share their good practice with other LAMs? (RQ1 - Authority)
- Are practitioners more or less confident about the quality of crowdsourced data than before they took part in the study, and does this affect their plans for data ingest to their CMSs? (RQ2 - Accuracy and Authority)
- Which specific accessibility improvements for transcription data discovery by people who use screen readers will be implemented, and on what timeline? (RQ3 - Accessibility)

We will use the qualitative and quantitative data gathered throughout *CDAAA* to write at least three different articles about our findings and submit these to **peer-reviewed scholarly journals** and **practitioner publications** (all open access).

Diversity plan:

CDAAA's commitment to diversity and inclusion is evident in the project design, methodological approaches, target groups (LAM Partners and Screen Reader Users), advisory board membership, project outputs, and dissemination plan. By building strategic partnerships and communicating project results to researchers, practitioners, and LAM users, *CDAAA* will strengthen the LAM field's commitment to diversity, equity, and inclusion practices. We will measure the success and impact of our diversity and inclusion efforts by the number of LAM Partners and Screen Reader Users we recruit through the grant, using the recruitment criteria outlined above.

CDAAA is strategically designed to evaluate the accessibility of crowdsourced transcriptions in CMSs for people who use screen readers, and will center the perspectives and expertise of 12 people with disabilities who are recruited as Screen Reader Users, and Zuhair Mahmoud, a member of the Advisory Board who brings personal and professional experience of using, improving, and designing systems to be more accessible. The involvement of people with disabilities across *CDAAA* ensures that representations of disability in the findings are directly supported by the lived experiences of people with disabilities.

LAM partners include public and private LAMs of various sizes, missions, and collection focus including papers of BIPOC authors, artists, thinkers, and women writers from the sixteenth century to the present. The involvement of community-LAM partnerships through the Tennessee Genealogy Indexing Project ensures focus on volunteer stakeholder needs and perspectives. By helping LAMs overcome barriers to incorporating volunteers' contributions (transcriptions and tags) into CMSs, *CDAAA* will also diversify the authoritative record, and help to shift the definitions and scope of LAM authority.

All scholarly outputs will be submitted to open access journals, and all practitioner-oriented UXD outputs and reports will be freely available via GitHub and the Digital Repository at the University of Maryland (DRUM) to ensure that cost is not a barrier to access for any potential (English language) audience. All project outcomes will be Web Content Accessibility Standards 2.1 compliant, and available in formats that are amenable to screen-reader technologies.

Project Results:

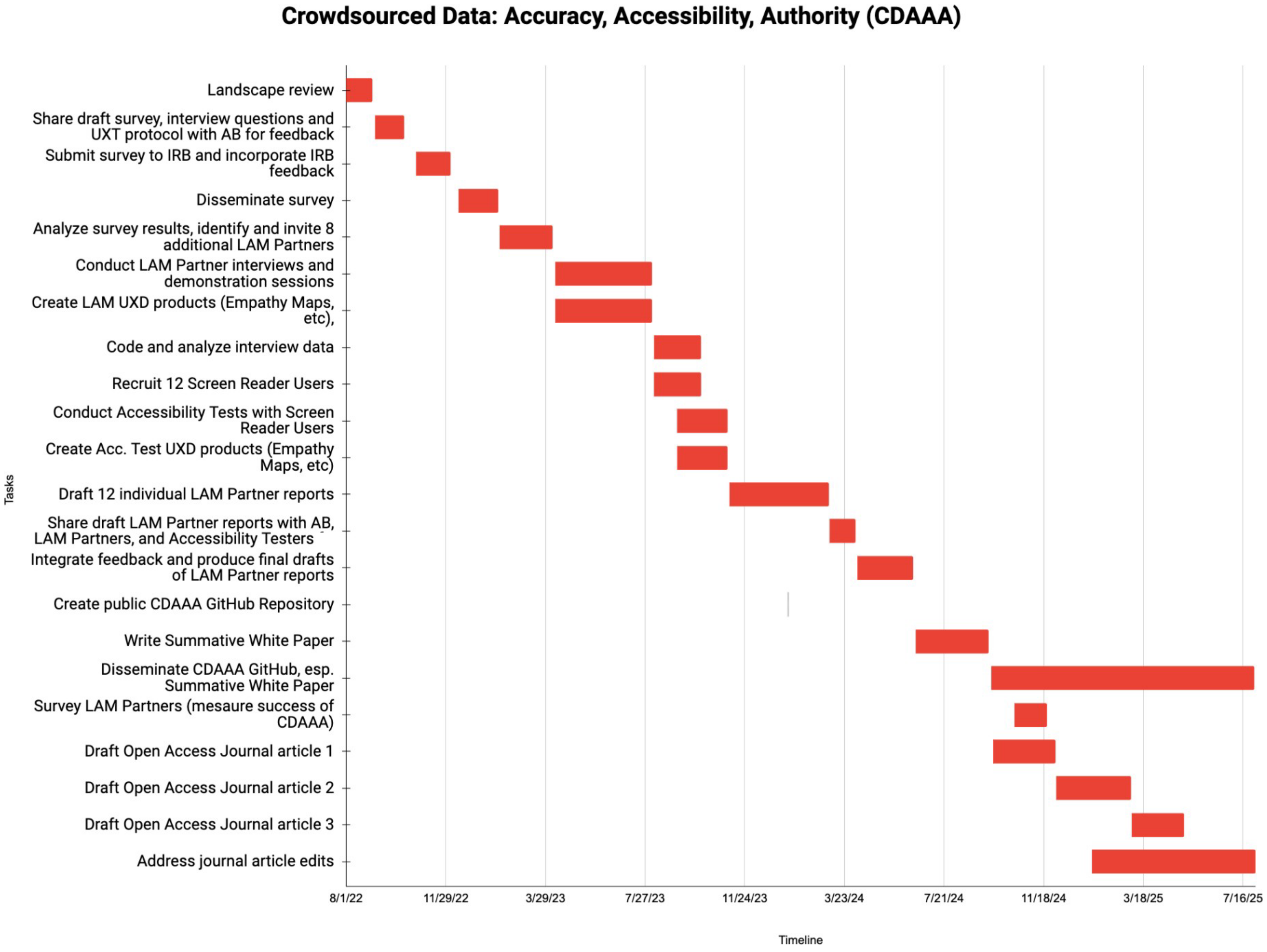
"**Data worth collecting and using also are worth preserving and sharing**" (Memorandum, 2015). This is true of crowdsourced data, and all of the project results of *CDAAA*. This project's findings will advance the LAM field's understanding of the value and current precarity of much crowdsourced data, and the missed opportunities when data is not ingested into LAM CMSs–most notably that LAM collections remain inaccessible to people who use screen readers, and that volunteer effort is wasted. *CDAAA* will provide tangible resources to aid LAMs now and in the future in understanding 1) the technical challenges of ingesting crowdsourced data and solutions to these challenges (**RQ1**), 2) LAMs' assessment of the quality of crowdsourced data and its place in the authoritative record **(RQ2)** and 3) the society-wide benefits to making their collections more accessible **(RQ3)**.

CDAAA findings will be disseminated in ways that are designed to meet the needs of different target audiences and beneficiaries. All project outcomes will be Web Content Accessibility Standards 2.1 compliant, and available in formats that are amenable to screen-reader technologies.

Output name	**Description**	**Audience (Targets and Beneficiaries)**	**Dissemination method**
12 Individualized LAM Partner Reports and UXD materials	A report of 6-10 pages per LAM, describing existing barriers or successes to crowdsourcing data ingest; accessibility to screen-reader technology (if data is in the CMS); UXD outputs (Empathy Maps, Personas, Stories, Scenarios); and recommendations for next steps.	12 LAM Partners (Target); other LAM practitioners, esp. data curation specialists, accessibility specialists, and crowdsourcing project managers (beneficiaries).	Share directly with LAM Partners; Publish on *CDAAA* GitHub; Digital Repository at the University of Maryland (DRUM)
Summative White Paper	A 30-page summary of project findings, including excerpted case studies from individual LAM Partners. This report will be designed to help other LAMs diagnose, test, and address	12 LAM Partners (Target); LAMs with similar crowdsourcing platform or project set-ups, CMSs, metadata schemas and workflows as the 12 LAM	Share directly with LAM Partners; Publish on *CDAAA* GitHub; DRUM

	crowdsourcing data ingest challenges at their own institutions. The accessibility implications for each barrier will be described, and scalable solutions and recommendations provided.	Partners. LAMs that are considering, but have not embarked on crowdsourcing, and are trying to plan the full life-cycle of a future project. CMS vendors (beneficiaries).	
***CDAAA* GitHub Repository**	A GitHub repository dedicated to the *CDAAA* project will be created and maintained for at least 5 years beginning in Y2 of the project (2023-2024).	GitHub is an ideal place to disseminate materials for LAM practitioners in digital content management, accessibility testing, and UXD roles (**target and beneficiaries)** as well as scholars and CMS vendors (beneficiaries).	A final accessible .pdf version of the UX products, 12 individual LAM Partner reports, and summative white paper will be disseminated under a CC BY-NC-SA 4.0 license on GitHub.
At least three publications in peer-reviewed scholarly and practitioner journals (Open Access)	Target journals: ● *Citizen Science Theory and Practice*, ● *Journal of Open Humanities Data*, ● *Journal on Computing and Cultural Heritage*, ● *JASIS&T* ● *Archival Outlook*	Researchers and Practitioners of: Information Science, Library Science, Crowdsourcing, Accessibility and Disability Studies, User-Centered Design, Digital Humanities	Relevant Open Access journal website/repository; DRUM; all Zooniverse-related publications will post on Zooniverse Publications page
Practitioner-oriented Conferences	● SAA ● ALA ● AERI ● ALISE ● Bridging the Spectrum	Practitioners and Researchers of: Information Science, Library Science, Crowdsourcing, User-Centered Design, Accessibility and Disability services; CMS vendors.	Copies of select conference slides, papers, videos, and posters will be shared via *CDAAA* GitHub, RRAD Twitter, and DRUM
Scholar-oriented Conferences	● ASIS&T ● Citizen Science Association ● iConference	Researchers and Practitioners of: Information Science, Library Science, Crowdsourcing, Accessibility and Disability Studies, User-Centered Design, Digital Humanities	Copies of select conference slides, papers, videos, and posters will be shared via *CDAAA* GitHub, RRAD Twitter, and DRUM
GA Doctoral Thesis	The GA will make significant progress towards his ~100,000 word thesis on a topic relating to *CDAAA*. The thesis will not be completed or defended by the end of the grant period.	Researchers of: Information Science, Library Science, Crowdsourcing, Accessibility and Disability Studies, User-Centered Design, Digital Humanities	DRUM with option to publish individual chapters as journal articles, or a revised version of the thesis as a monograph

Schedule of Completion



Crowdsourced Data: Accuracy, Accessibility, Authority (CDAAA)

Digital Products Plan

This plan describes how *Crowdsourced Data: Accuracy, Accessibility, Authority (CDAAA)* will create and manage digital products, and the practices and standards appropriate for this project.

Type

The following born-digital products will be created by the *CDAAA* team (PI, GA, ARS). These are the final products of the grant. For more details on the data collection and management practices underlying these products, see the **Data Management Plan**.

- ***CDAAA* GitHub repository** (Y2): A public and freely available repository will be created in 2023 and maintained for at least 8 years. By the submission of the final grant report, the repository will contain all final products of the research, and contact details for the PI (Data Manager).
- **LAM Partner reports and UX products** (Y3): Final versions of the LAM practitioner-oriented UX products and individual LAM Partner reports will be shared with all LAM Partners, Advisory Board members, and accessibility testers via email. 24 final User-Centered Design Empathy Maps (1 per participant), and ~10 Personas, Stories, and Scenarios will be created by the PI and GA (based on industry standard templates), and populated with data by the PI, GA, and ARS. The number of Personas is typically less than half of the number of Empathy Maps.
- **Summative white paper** (Y3): A final summative white paper will be shared with all LAM Partners, Advisory Board members, and accessibility testers via email, and disseminated more broadly to the public via the GitHub repository dedicated to the *CDAAA* project.
- **Open-access journal articles** (Y3): At least three scholarly journal articles will be submitted to open-access journals to ensure maximum access and reuse potential of *CDAAA* findings.

Availability

CDAAA will make works produced through IMLS support widely available and share work products whenever possible through free and open-access journals, repositories, and websites, including GitHub, DRUM and, for any finalized journal publications that relate to Zooniverse, the Zooniverse Publications page. Final versions of the LAM practitioner-oriented UX products, individual LAM Partner reports, and summative white paper will be shared with all LAM Partners, Advisory Board members, and accessibility testers via email, and disseminated more broadly to the public via a GitHub repository dedicated to the *CDAAA* project. This GitHub repository will be created in Y2 of the project (2023-2024) and maintained for at least 8 years thereafter.

The repository will also contain a description of the data collected through the survey, interviews, and UXD demonstration and accessibility testing sessions. GitHub is an ideal place to disseminate materials to a broader audience of LAM practitioners whose duties include incorporating crowdsourced transcriptions into their CMSs and ensuring their accessibility. It is also an ideal forum for sharing project outcomes with crowdsourcing project and platform creators (Zooniverse, From the Page, and others), as well as CMS vendors, many of whom use GitHub for their own codebases.

Access

All project products released on the *CDAAA* GitHub repository will be available under a CC BY-NC-SA 4.0 license. Scholarly outputs will be submitted to open-access journals to ensure maximum access and reuse potential. The grant budget includes funds to support open-access publication for this reason. The PI and any co-authors of published scholarly articles will assert copyright.

A central goal of this project is to **increase awareness and transparency** about the difficulties and successes LAMs experience when ingesting crowdsourced data into their CMSs (the authoritative record). We therefore seek LAM Partners who can participate without anonymization or pseudonymization of their LAM, because by sharing their challenges and protocols openly, we hope LAM organizations can build a strong community of practice in which they can share solutions during the course of the grant and in the future. The proposed research could go forward without anonymization of the LAM organizations. In any case, we will address privacy concerns at the individual level by anonymizing participants' names and other personal information collected during the grant period. **Accessibility Testers'** identities will be anonymized to protect their privacy. We will work closely with our Advisory Board, IRB, and senior colleagues at UMD to ensure we structure the interview protocols appropriately to achieve sound data outcomes while protecting participants, understanding that this standard of openness (naming LAMs) may not be attainable.

Sustainability

Digital products including the LAM Partner reports and summative white paper created by *CDAAA* will be freely and readily available for use and reuse by libraries, archives, museums, and the public to the maximum extent possible. In addition to publishing project results via GitHub and open-access journals, copies of all scholarly and practitioner-oriented project products will be deposited in the Digital Repository at the University of Maryland (DRUM) for long-term preservation. This includes the doctoral dissertation work of the GA, which will be shaped by his work on *CDAAA*. All project outcomes will be Web Content Accessibility Standards 2.1 compliant, and available in formats that are amenable to screen-reader technologies.

Data Management Plan

Data Type and Purpose

The following born-digital data will be captured with a variety of tools and stored in a secure shared Google Drive only accessible to the *CDAAA* team (PI, GA, ARS).

- **Project management** (Y1-Y3): The PI, GA and ARS will utilize a Trello Kanban board to track project tasks, Google Docs for weekly meeting notes, and Slack for ephemeral messaging to organize their two-week Agile Sprints.
- **Landscape review** (Y1-Y3): The Recovering and Reusing Archival Data (RRAD) Lab Zotero Library will house metadata for all primary and secondary publications gathered in the Landscape Review (Y1) and throughout the project. The Library may be made available upon request.
- **Communications** (Y1): Drafts of survey participant recruitment emails and social media communications will be created in a private Shared Google Drive for *CDAAA* and final copies will be shared with the IRB (for approval), Advisory Board and LAM Partners (to disseminate).
- **Survey** (Y1 and Y3): LAM practitioner survey results will be captured through Qualtrix and analyzed with SPSS. We anticipate 60 LAM practitioner respondents in Y1, and 12 LAM Partners respondents in Y3.
- **Interviews** (Y1): 12 LAM Partner Interviews will be audio recorded if in person, and audio and video recorded if conducted via Zoom. Recordings will be mp3 and/or mp4 formats. Interviews will be transcribed through Rev.com and uploaded to NVivo.
- **Qualitative coding** (Y2): 12 LAM Partner Qualitative coding files of the interviews will be created by the GA and PI in NVivo, saved in .nvpx, downloaded and saved to the shared drive.
- **Demonstrations** (Y1-Y2): 12 recorded demonstration sessions with the LAM Partners, and 12 CMS accessibility testing sessions with Accessibility Testers will be audio and video recorded via Zoom whether in person or virtual, to allow for screen recording. Recordings will be mp3 and mp4 formats, downloaded and saved to the shared Google Drive.
- **Interview notes** (Y1-Y2): 24 sets of interview session notes will be created by the PI and GA (2 total for each of the 12 interviews) to supplement the transcripts and facilitate coding, using a Google Docs template created by the PI.
- **Protocol sheets** (Y2): 24 Demonstration and User-Centered Design testing session protocol spreadsheets (Google Sheets, one sheet per participant) will be used by the PI, GA, and ARS to capture prompts, questions, and participant responses/behaviors. Any additional notes created by the *CDAAA* team during these sessions will be captured in Google Docs.
- **Repository** (Y2): A public and freely available *CDAAA* GitHub repository containing all final products of the research, and contact details for the PI will be created in Y2 of the project (2023-2024) and maintained for at least 8 years (see **Digital Products Plan** for details).
- **Publications and reports** (Y1-Y3): Drafts of all publications, conference papers, LAM Practitioner reports, and the summative white paper will be created as Google Docs in the shared drive to facilitate timely collaboration and easy version control. All drafts will be named in a standard way, i.e., "YYYY-MM-DD-DATA/SESSION-TYPE-Draft#".

Data Protection

CDAAA is committed to protecting the rights and privacy of project participants, including through protecting confidentiality and personal privacy. *CDAAA* will collect PII including names, institutional affiliation, role, emails, demographic information (race, gender, age range, disabilities that necessitate the use of a screen-reader), and participants' assessments of LAM CMSs, for research purposes. Data made available for broader use will be free of any information that could lead to the disclosure of the identity of individual participants.

A central goal of this project is to **increase awareness and transparency** about the difficulties and successes LAMs experience when ingesting crowdsourced data into their CMS (the authoritative record). We therefore seek

LAM Partners who can participate without anonymization of their LAM, though individual participant names could be anonymized. We will work closely with our Advisory Board, IRB, and senior colleagues at UMD to ensure we structure the survey, interview, demonstration, and usability testing protocols appropriately to achieve sound data outcomes while protecting individual participants, understanding that the proposed standard of openness around LAM identities may not be attainable. This would not negatively impact the results of *CDAAA*, but may limit the development of a strong LAM community of practice that can continue to address the challenges of crowdsourced data management after the course of the grant. Accessibility testers' identities will be anonymized, but their specific reason for using a screen reader (e.g. Blindness) will be made available in any resulting data and research products.

All students, faculty and staff of UMD complete mandatory physical and virtual IT security training annually. The PI, GA, and ARS all utilize password management software (e.g. LastPass), in addition to UMD-mandated multi-factor authentication for all UMD web-properties, devices (e.g. phones and laptops), Google products, email, and other licensed products such as NVivo.

Data Access

CDAAA will deposit all Web Content Accessibility Standards 2.1 compliant finished products resulting from this research in a broadly accessible repository that allows public access without charge, no later than the submission of the final performance report to IMLS. Data will be deposited in a machine-readable, non-proprietary digital format to maximize search, retrieval, and analysis. Final performance reports will identify the data depository locations for access by the public: a dedicated *CDAAA* GitHub repository, the Digital Repository at the University of Maryland (DRUM), and any open-access journal websites. These repositories are freely available through any modern browser (Chrome, Safari, FireFox, etc). All published open-access journal articles resulting from the research will be deposited in DRUM, and provide information about how to contact the PI for access to anonymized raw data (i.e. survey results, interview transcripts).

Documentation

Interviews, demonstrations, and usability testing sessions may be conducted in person (following UMD- or LAM-prescribed COVID precautions) or via video conferencing software with video, audio, and screen sharing. Interviews will be recorded with participants' consent; participants will read and sign a consent form before the interviews, and each interview will begin with reading the consent form to confirm interviewees understand the process and can ask any questions. (See Appendix A and B for draft survey and interview questions). Interviews will be transcribed through Rev.com, and coded independently by the PI and GA using NVivo, a qualitative analysis software, to enable us to align the coding via inter-rater reliability testing.

All raw data including Qualtrix survey results; consent forms; recorded demonstration and testing sessions, and interviews (mp3 and mp4); NVivo codebooks .nvpx outputs; UXD draft products, UXD testing session spreadsheets; Google Doc drafts of all 12 Individual LAM Partner reports, drafts of the summative white paper, drafts of submitted open-access journal articles; and any notes by the PI, GA and ARS produced will be saved in a private Shared Google Drive for the project during the course of the grant, and then hosted for 7 years in UMD secure networked storage space after the end of the grant period.

Dissemination and Preservation

Anonymized portions of raw data will be made available to other researchers and practitioners upon request, per IRB review. PI Van Hyning will be the official Data Manager for the project for 10 years (8/2022-8/2032), and her up-to-date contact information will be included on the *CDAAA* GitHub, the Digital Repository at the University of Maryland (DRUM), and her professional website. All finished products of the grant, such as journal articles and the white paper, made include links to GitHub and DRUM, to facilitate communication with the PI.

Plan Implementation and Review

The PI will review the Data Management Plan quarterly for the duration of the grant, conducting a thorough audit of all research data and products with the GA and ARS to ensure all data are maintained in accordance with the highest standards of data privacy.

Organizational Profile

University of Maryland

Mission:

The University of Maryland College Park is a public research university, the flagship campus of the University System of Maryland, and the original 1856 land-grant institution in Maryland. The University of Maryland is dedicated to achieving excellence as the State's primary center of research and graduate education and the institution of choice for undergraduate students of exceptional ability and promise. With a commitment to diversity of faculty, students and staff, the University advances knowledge, provides outstanding and innovative instruction, and nourishes a climate of intellectual growth in a broad range of academic disciplines and interdisciplinary fields for the benefit of the economy and culture of the State, the region, the nation and beyond.

Service Area:

The Fall 2021 enrollment was 41,271 a total of graduate and undergraduate students. 42% of the population are minority students. 34.991% of students come from out-of-state and 65.009%% are Maryland residents. The University serves the state of Maryland as a premier research institution and reaches national distinction as ranking among the very best of public research universities in the United States.

College of Information Studies
Maryland's *i*School in the Information Capital

Mission:

The College of Information Studies, Maryland's *i*School, engages in collaborative, interdisciplinary, and innovative research, teaching, and service. We educate information professionals and scholars, and we create knowledge, systems, and processes.

Service Area:

The *i*School offers Master's degrees in Library Science (MLS), Information Management (MIM), Human Computer Interaction (HCIM) and a doctorate degree in Information Studies. Per most recent admission data, 355 students are enrolled in the MLS program, 43 enrolled in the MIM program, 123 enrolled in HCIM program and 76 enrolled in the doctoral program. *i*School also offers Bachelor's degree in Information Science (BSIS) and the current number of enrollees is 1479. Approximately 42.39% of the total student body is female and 26.65% are underrepresented students. The *i*School has 45 Tenured and Tenured-track faculty, 35 Professional-track faculty, 54 staff and 113 adjunct faculty representing diverse subject areas in information studies. The *i*School serves the mid-Atlantic region.