

# Data Quality and Accessibility Issues for Crowdsourced Transcriptions

MACE JONES AND VICTORIA VAN HYNING

## Abstract:

CDAAA is a 3-year Institute of Museum and Library Services (IMLS) grant-project to investigate sociotechnical barriers that Libraries, Archives, and Museums (LAMs) face in making crowdsourced transcriptions open and accessible to sighted users, and print-disabled people who use assistive technology to access digital text. Crowdsourced transcriptions are the most common data solicited by LAMs because, theoretically, they provide access to images of documents that are otherwise not machine-readable. We seek to understand whether blind and low vision users are aware of these resources and able to find them.

We work with 12 US-based LAM Partners spanning private, public, university, and federal LAM institutions and community-led groups who run crowdsourcing projects. All LAM Partners have participated in a semi-structured interview about their crowdsourcing methods, and Partners who had previously ingested their data participated in an additional technical interview demonstrating their ingest process. We have conducted usability and accessibility tests of 9 LAM partner discovery systems where crowdsourced transcriptions have been incorporated, with 11 people who are blind or have low-vision and use assistive technology to navigate the web.

We will share preliminary findings, and illustrate the usability and accessibility aspects of this study through an example from the Folger Shakespeare Library, which we tested with two blind and one low vision participant. We link our broader findings with new rulemaking on the accessibility of web content and mobile apps under Title II of the Americans with Disabilities Act (ADA) which came into effect June 24, 2024.

## About the authors:

*Victoria Van Hyning* is an Assistant Professor of Library Innovation at the College of Information, University of Maryland, College Park (joined 2020). She has a background in medieval and early modern English literature, digital humanities, citizen science and cultural heritage crowdsourcing, and publishes in these areas. She is a co-founder and [Director of the Center for Archival Futures](#) (CAFe) where she focuses on community cultural heritage, crowdsourcing, and data reuse, and a member of the new [Maryland Initiative for Digital Accessibility](#) (MIDA). She was awarded an Institute of Museum and Library Services Early Career Grant in 2022 for her project “[Crowdsourced Data: Accuracy, Accessibility, Authority \(CDAAA\)](#),” to investigate the sociotechnical barriers that libraries, archives, and museums (LAMs) face in integrating crowdsourced transcriptions into their discovery systems, and whether transcriptions increase accessibility for people who are blind or low vision and use assistive technology to access web content. Victoria previously worked at the Library of Congress on the [By the People](#) project (2018-2020), and held a British Academy Postdoctoral

Fellowship in English literature at Oxford University, where she also served as the Humanities PI of the crowdsourcing platform [Zooniverse.org](https://zooniverse.org) (2015-2018).

***Mace A. Jones*** is a Ph.D. candidate in information at the University of Maryland College of Information. Their work focuses on researching archival descriptive and metadata practices to improve the accessibility of humanistic and scientific archives. Jones holds a Master of Arts in English Literature focusing on medieval and early modern English literature from Auburn University and a Master of Library and Information Studies with an archival concentration from The University of Alabama. They are a student member of the [Center for Archival Futures](#) (CAFe) and the [Maryland Institute for Digital Accessibility](#) (MIDA) at the University of Maryland. They were [2023 LEADING Research Fellow](#), where their work focused on visualizing Linked Open Data networks for the Digital Scriptorium project.