Data Quality and Accessibility Issues for Crowdsourced Transcriptions

SAA Research Forum 2024
July 17, 2024

Dr. Victoria Van Hyning (PI) and Mace A. Jones
The University of Maryland College of Information
CDAAA Project Email: cdaaa@umd.edu

This research is supported by the University of Maryland College of Information Studies and funded by the Institute of Museum and Library Services (IMLS) Laura Bush 21st Century Librarian Program Grant#: RE-252344-OLS-22. IRB package 1985210-2

Crowdsourced Data: Accuracy, Accessibility, and Authority (CDAAA)

Crowdsourced Data: Accuracy, Accessibility, and Authority (CDAAA) is a 3-year Institute of Museum and Library Services (IMLS) early career grant project to identify the sociotechnical barriers that Libraries, Archives, and Museums (LAMs) face in making **crowdsourced transcriptions accessible** to sighted users, and print-disabled people who use assistive technology to access digital text.

Assistant Research Scientist and accessibility specialist J. Bern Jordan also serves on our team.

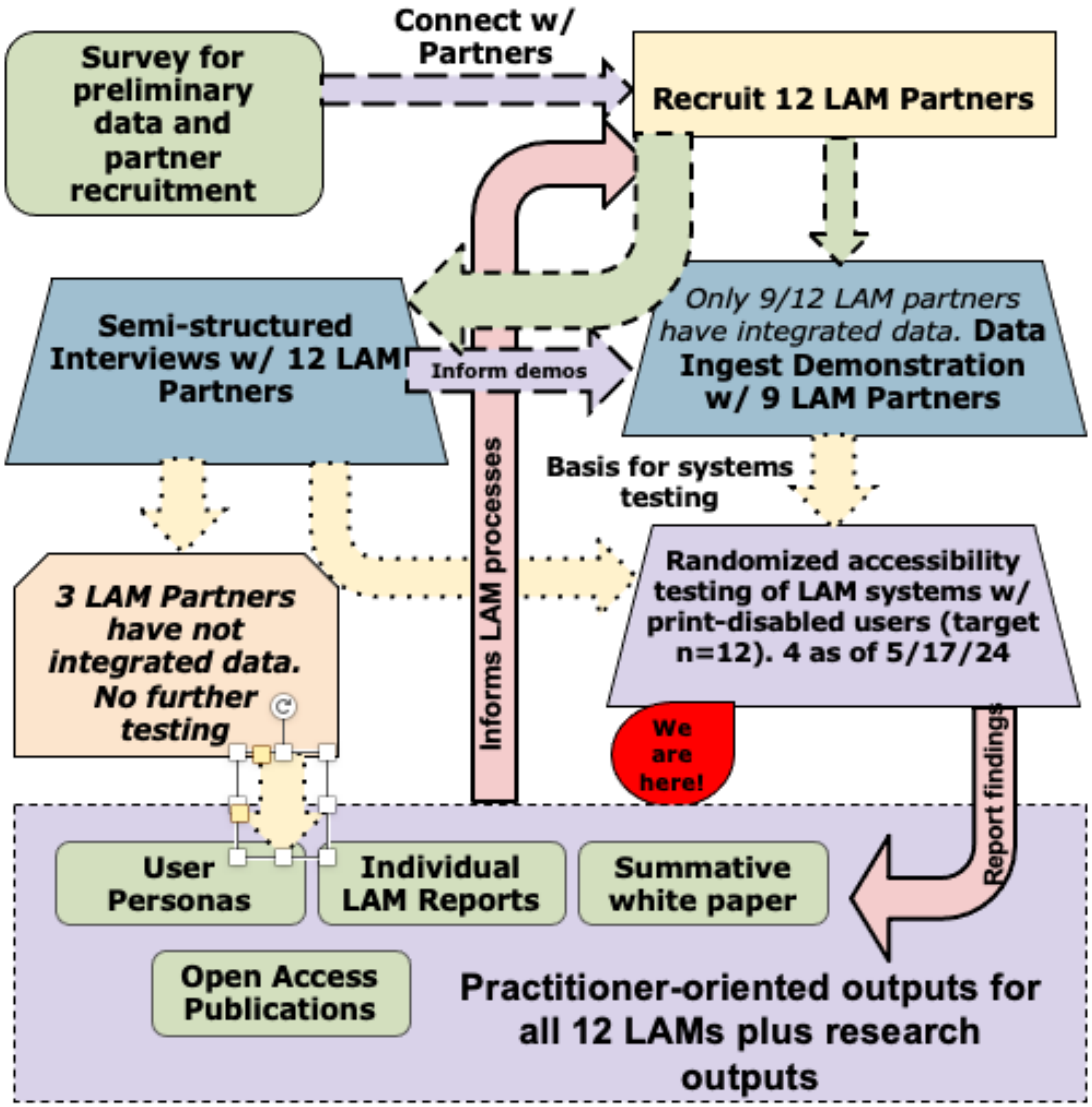
CDAAA GitHub QR Code
https://github.com/VVH/CDAAA

CDAAA Core Research questions

- **RQ1 (Authority):** Are LAMs able to integrate crowdsourced transcriptions into their CMSs (the authoritative record)? If yes, how? If not, what technical barriers do they face?
- **RQ2 (Accuracy and Authority):** What are LAM practitioners’ attitudes towards crowdsourced transcription data quality? Do these attitudes impact whether or not crowdsourced data are incorporated into CMSs? How do LAM practitioners assess the quality of crowdsourced data?
- **RQ3 (Accessibility):** When transcription data is successfully integrated into a CMS or database, is it accessible to people who are print-disabled, e.g. blind or have low vision, and use assistive technology? If not, what is required to make the data legible? What are print-disabled users’ experiences of searching for and reading transcription data?

CDAAA Research Methods

A mixed-methods approach to assessing accessibility of crowdsourced transcriptions and the usability and accessibility of LAM discovery systems.



Survey for preliminary data and partner recruitment

Connect w/ Partners

Recruit 12 LAM Partners

Semi-structured Interviews w/ 12 LAM Partners

Inform demos

Only 9/12 LAM partners have integrated data. **Data Ingest Demonstration w/ 9 LAM Partners**

Informs LAM processes

Basis for systems testing

3 LAM Partners have not integrated data. No further testing

Randomized accessibility testing of LAM systems w/ print-disabled users (target n=12). 4 as of 5/17/24

We are here!

Report findings

User Personas

Individual LAM Reports

Summative white paper

Open Access Publications

Practitioner-oriented outputs for all 12 LAMs plus research outputs

Summary of findings for RQ1 (Authority) RQ2 (Accuracy and Authority):

- 9/12 LAM Partners rate volunteer transcriptions as high enough quality to ingest with varied degrees of text editing and post-processing required, and have successfully integrated or published data online.
- 2 LAM Partners collected crowdsourcing data that is too low quality or difficult to use and have not ingested it.
    - 1 rates the data quality as sufficient for their research, but unlikely to be ingested by the LAM that holds the original collections. This partner has crowdsourced transcriptions for the same dataset twice, because the first effort resulted in poor quality data, due to the crowdsourcing system they used.
    - The other person rates their data as unusable and low quality, despite spending considerable effort and resources trying to clean it.
- 1 LAM Partner has gathered transcriptions but hasn’t looked at them yet due to staff turnover and other constraints. They are keen to integrate the data with minimal vetting and make it part of standard metadata and data management practices. They are unsure what system to use.

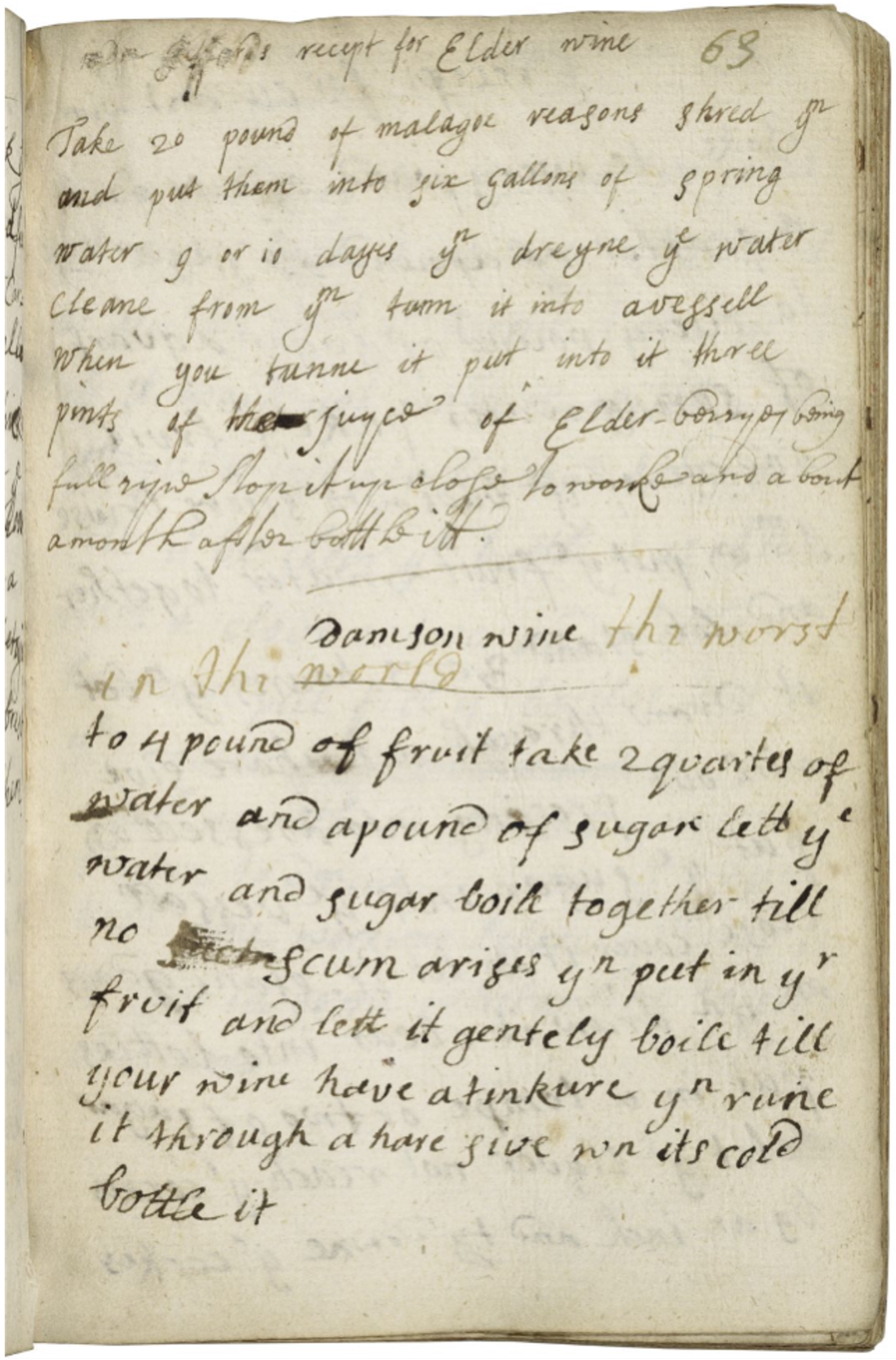
To test RQ3 we conduct accessibility-focused user-testing interviews:

1. ~120 minute interviews with blind or low-vision users.
2. We ask them to navigate 3 different LAM Partner systems in a randomized order (randomized stimuli) that differ for each interviewee
3. Participants are asked to navigate to each LAM repository and narrate each step of their search, and how they know/believe they have arrived at the right point (think-aloud protocol).
4. Participants are asked to search for transcriptions associated with special collection items and read a page with their screen reader, magnifier or other assistive device.
5. Participants are then asked to search for a specific phrase in a crowdsourced transcription that we know is present in the system, and read it with their assistive technology.
6. We ask 2 System Usability Scale questions about each system.
7. Finally, we ask all participants to read a cleaned transcript with consistent headings and structure, and compare this reading experience with the pages they encountered in other systems.

Here's a specific example of what this looks like in practice:

- We ask participants to search for the Folger Shakespeare Library manuscript collections.
- We ask participants to look for transcriptions. Folger is unusual among our LAM Partners in having a ‘transcriptions’ facet to easily locate and access the text.
- We ask participants to read a page if they can find one with a transcription.
- Then we ask them to search within the Folger system for "the worst in the world", and read the results.

Damson wine “the worst in the world”

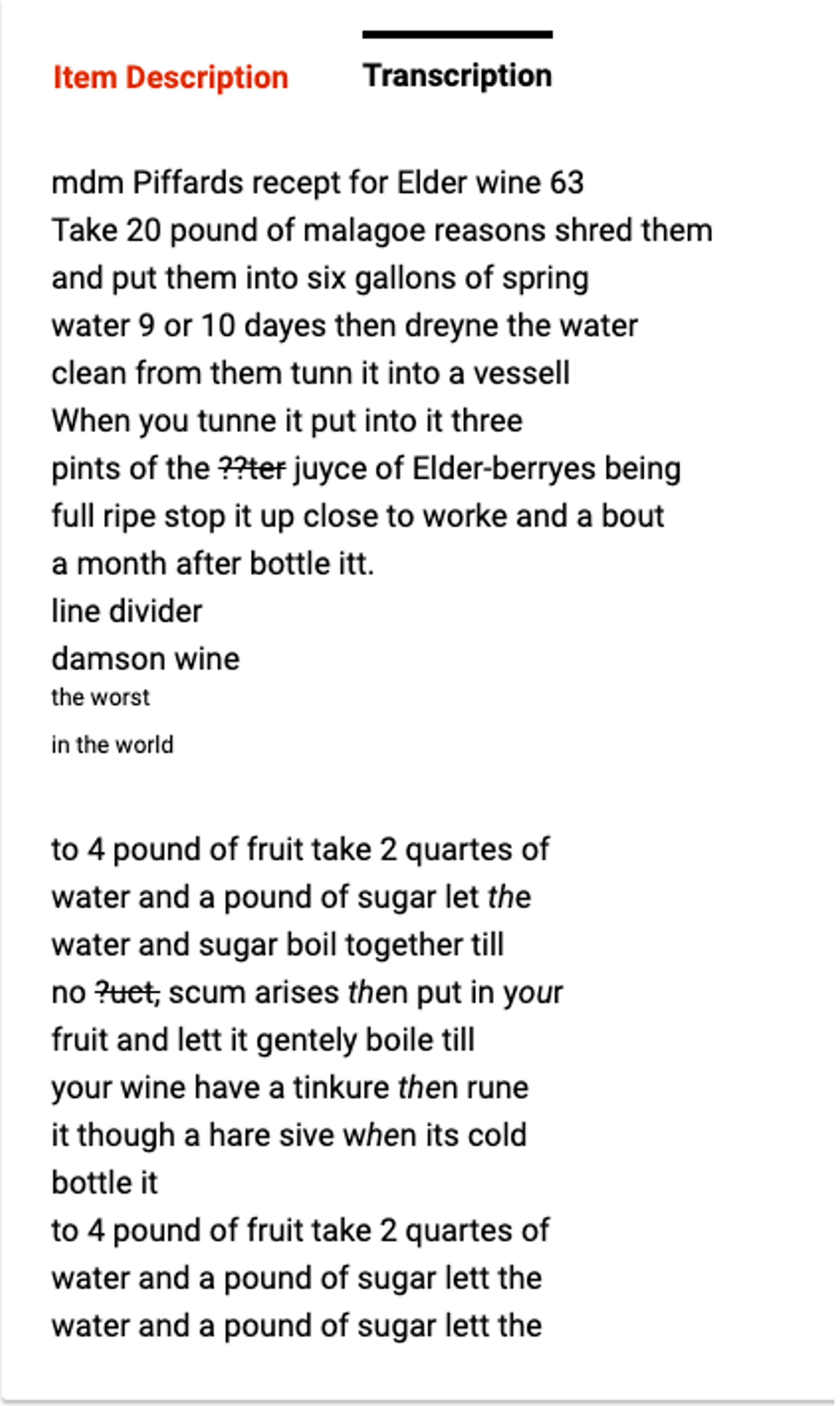


mdm Piffards recept for Elder wine 63

Take 20 pound of malagoe reasons shred y^{m}
and put them into six gallons of spring
water 9 or 10 dayes y^{n} dreyne y^{e} water
cleane from y^{m} tunn it into avessell
When you tunne it put into it three
pints of the juyce of Elder-berryes being
full ripe stop it up close to worke and a bout
a month after bottle itt.

Damson wine the worst
in the world

to 4 pound of fruit take 2 quartes of
water and apound of sugar lett y^{e}
water and sugar boile together till
no scum ariſes y^{n} put in y^{r}
fruit and lett it gentely boile till
your wine have a tinkure y^{n} rune
it through a hare sive w^{n} its cold
bottle it



mdm Piffards recept for Elder wine 63
Take 20 pound of malagoe reasons shred them
and put them into six gallons of spring
water 9 or 10 dayes then dreyne the water
clean from them tunn it into a vessell
When you tunne it put into it three
pints of the ~~??ter~~ jyuce of Elder-berryes being
full ripe stop it up close to worke and a bout
a month after bottle itt.
line divider
damson wine
the worst
in the world

to 4 pound of fruit take 2 quartes of
water and a pound of sugar let *the*
water and sugar boil together till
no ~~?uct,~~ scum arises *then* put in *your*
fruit and lett it gentely boile till
your wine have a tinkure *then* rune
it though a hare sive *when* its cold
bottle it
to 4 pound of fruit take 2 quartes of
water and a pound of sugar lett the
water and a pound of sugar lett the

Citation: Carr, Lady Anne. “Choyce Receits Collected out of the Book of Receits, of the Lady Vere Wilkinson Begun to Be Written by the Right Honble the Lady Anne Carr, Jan. 28 1673/4.” Manuscript. Washington D.C., Folger Shakespeare Library. https://digitalcollections.folger.edu/bib243782-308828

FOLGER SHAKESPEARE LIBRARY
Digital Collections

Accessibility testing results

ACC Tester ID	Blind or low vision	T1: Can you navigate to the Folger Shakespeare Library's website?	T2: Search for digitized materials i.e. images of manuscripts with transcriptions	T3: Search for "The worst in the world"	Rate statement: 'This discovery system website is easy to use' (strongly agree, agree, neither agree nor disagree, disagree, strongly disagree)	Rate statement: 'This website meets my needs as a user.' (strongly agree, agree, neither agree nor disagree, disagree, strongly disagree)	LAM CMS test position	CMS*
ACC-3	B	Complete Success	Failure	Success with minor issues	Agree	Agree	1	LUNA and Hamnet
ACC-7	LV	Complete Success	Failure	Success with minor issues	Agree	Agree	2	Islandora
ACC-9	B	Complete Success	Success with major issues	Success with minor issues	Neither agree nor disagree	Disagree	3	Islandora

Table: Results of Folger site testing. *Folger migrated from LUNA to Islandora during our testing.

Broader (preliminary!) findings for RQ3 (Accessibility) for all 11 testers

- Most of our participants have or are working towards Masters or PhD degrees, and some use primary sources in their work, yet no one was aware of the availability of crowdsourced transcription resources, and most were unfamiliar with searching for this content in LAM CMSs or databases.
- Users often found existing pathways through CMSs frustrating and difficult to navigate.
- Most users benchmark their expectations for crowdsourced transcriptions against low-quality OCR text.
- All users were unfamiliar with transcription conventions and scholarly editing practices, such as representing original spelling and deletions in encoded text and likened these to “tracked changes.” The ”worst in the world” is a good example.

Recommendations for LAMs and CMS creators

- LAMs can do more outreach to print-disabled users about the availability of free crowdsourced transcription data
- Make transcription conventions available with the transcriptions. I.e. how are deletions represented? Were original spelling and punctuation persevered? What markup is used to indicate transcriber uncertainty?
- CMS and database creators could deploy standard fields for transcription data and improve discovery pathways for all users, not only print-disabled users.
- Transcriptions, OCR, and HTR are important tools for expanding accessibility and meeting the new rules 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

Thank you! Any questions?

- Contact: cdaaa@umd.edu
- Grant Page: https://www.imls.gov/sites/default/files/project-proposals/RE-252344-OLS-22-full-proposal.pdf
- Project Info: https://mida.umd.edu/crowdsourced-data-accuracy-accessibility-and-authority/
- CDAAA GitHub: https://github.com/VVH/CDAAA