

RESPONSIBLE PRACTICES FOR CROWDSOURCED TRANSCRIPTION AND ACCESSIBILITY FOR WEB-BASED LIBRARY, ARCHIVE, AND MUSEUM CONTENT

Introduction

Crowdsourced Data: Accuracy, Accessibility, Authority (CDAAA) investigates whether and how LAMs integrate crowdsourced transcriptions into their discovery systems and whether the results are accessible for people who use assistive technology to navigate the web.

Purpose of Study

Our research provides timely insights and guidance for LAMs who are creating and managing crowdsourced data in their digital repositories or content management systems to better provide their users with accessible cultural heritage data.

Research Questions

- RQ1 (Authority): Are LAMs able to integrate crowdsourced transcriptions into a Content Management System (CMS) or other repository (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 with CMSs, is it accessible to print-disabled people? What are print-disabled users' experiences of searching for and reading transcription data?



Ada Blackjack with the expedition's cat, Vic, aboard the rescue ship, 1923, from Dartmouth Libraries

AUTHORS

Victoria Van Hying (PI), J. Bern Jordan, Mace A. Jones
The University of Maryland College of Information
College Park, MD

Ada Blackjack Diary

4 accessibility testers who tested the Dartmouth Library system encountered a unique object: the diary of Ada Blackjack (Iñupiat, 1898-1983), the sole survivor of the Wrangel Island expedition. 11 of our 12 testers also tested an HTML version of a page from this document.

28.

I Caught a fox today the fox has trap
in his foot it has caught trap
by trap that funny ny. I saw one
cut of wood and chop it. And I
finish my black belt.

29th.

My eye
swollen up and one side of my face
ach, it stard yesterday and today it
very much ach. I caught one male fox
it pretty good meat, and I saw five six,
cut of wood and I chop only three.

Image of the Ada Blackjack Diary and its accompanying encoded text transcript

Quantitative results from user testing of the Blackjack Diary

ACC Tester ID	Blind or low vision	T1: Can you navigate to the Dartmouth College Libraries' website?	T2: Search for digitized materials i.e. images of manuscripts with transcriptions	T3: Search for phrase: "three white foxess that makes seven foxess altogather"	Rate statement: 'This discovery system website is easy to use' (strongly agree, agree, neither agree nor disagree, somewhat disagree, strongly disagree)	Rate statement: 'This website meets my needs as a user.' (strongly agree, agree, neither agree nor disagree, somewhat disagree, strongly disagree)	LAM CMS test position
ACC-1	B	Complete Success	Failure	Failure	Agree	Agree	3
ACC-4	B	Complete Success	Failure	Failure	Agree	Neither agree nor disagree	1
ACC-9	B	Complete Success	Success with minor difficulties	Failure	Agree	Agree	2
ACC-11	B	Complete Success	Failure	Failure	Agree	Agree	3

Original HTML	Change	Resulting HTML
<h2>28.</h2>	No change. Retained headings, which screen reader users liked.	<h2>28.</h2>
I Caught a fox today the fox has trap 	Replaced CSS line break spans with semantic line breaks . This prevented the problem where lines were strung together.	I Caught a fox today the fox has trap
five	Replaced CSS deleted sections with semantic deleted tags .	five
six	Replaced CSS added sections with semantic inserted tags <ins>.	<ins>six</ins>

Simplified HTML for additional user testing.

Conclusion

CMS and database creators could better serve LAM practitioners and their users by deploying standard fields for accessible transcription data. Such fields could improve discovery pathways for all users, not just blind users. LAMs could enhance accessibility of transcriptions by including information about transcription conventions, providing texts with minimal markup, and/or modernized language.