

CROWDSOURCING: HOW TO KEEP USERS ENGAGED

Alexandra Lightle

April 26, 2018

INTRODUCTION

- ❑ Crowdsourcing is a tool used by many archival collections, that utilizes volunteer labor to identify aspects of documents and photos that Optical Character Recognition software cannot identify, such as handwritten documents, specialized fonts, and photographs.
- ❑ Without volunteers, a crowdsourcing endeavor is moot. So when it comes to crowdsourcing data transcription, how can archives keep volunteers coming back for more?

WHO BENEFITS FROM CROWDSOURCING?

- ❑ Archivists with large data sets that need to be transcribed.
 - ❑ Particularly those collections who have small staff sizes and constrained by limited budgets.
- ❑ Researchers who have previously had limited access to data.
 - ❑ i.e. travel or access restraints.

LITERATURE REVIEW

- ❑ Main topics of literature consulted:
 - ❑ Expansion of research capability
 - ❑ Engagement with the public
 - ❑ Use of volunteer labor
 - ❑ Reliability of transcriptions
 - ❑ Crowdsourcing projects that also serve as social platforms

RESEARCH EXPANSION

- ❑ Matt Enis “Wisdom of the Crowd;As Libraries Turn to Crowdsourcing for Assistance with Large Digitization Projects, a Consensus on the Best Practices is Beginning to Emerge.”
 - ❑ “This goes to show how special collections can prove useful to researchers in varied and unexpected ways. Here, a novelist peruses NYPL’s unique menu collection to cook delectable dinner details into a historical mystery. There, a paleo oceanographer expands his field’s understanding of how seafood consumption trends of the 19th and 20th centuries impacted the sustainability of modern fisheries.”

ENGAGEMENT WITH THE PUBLIC

- ❑ Jan Zastrow, “Crowdsourcing Cultural Heritage: 'Citizen Archivists' for the Future”
 - ❑ “Crowdsourcing projects provide tools for volunteers to participate and engage. The public is hungry to help, whether as “citizen archivists,” do-it-yourself historians, or amateur info pros. [...] Crowdsourcing can serve as a kind of preservation technique for archival materials; it provides great access to a digital stand-in, thereby greatly reducing the need to see the original.”

USE OF VOLUNTEER LABOR

- ❑ Trevor Owens, “Digital Cultural Heritage and the Crowd.”
 - ❑ “The term “crowd” is somewhat misleading, since most successful crowdsourcing projects do not rely on large, anonymous masses of people. These projects succeed by inviting participation from engaged members of the public. The success built upon long-standing tradition of volunteerism and involvement of citizens in the creation and development of public good.”

RELIABILITY OF TRANSCRIPTIONS

- ❑ Peter Shillingsburg. “Reliable Social Scholarly Editing.”
 - ❑ “Users of scholarly editions are seldom on the look-out for errors; they need something that can be relied on to be accurate and comprehensive. It is not that the crowd cannot a good reliable edition. Like Project Gutenberg, also crowd sourced, the results can be either good or bad, but without a serious vetting process, the user cannot know which is which.”

PROJECTS AS SOCIAL PLATFORMS

- ❑ Ruth Mostern and Marieka Arksey. “Don’t Just Build it, They Probably Won’t come: Data Sharing and the Social Life of Data in the Historical Quantitative Social Sciences.”
 - ❑ [J.P] Birnholtz and [M.J.] Bietz have suggested that repository developers capitalize on the fact that data has a ‘social life’, a concept which argues that documents are a means to make and maintain social groups and not just the means to deliver information, to establish a community of data sharing.
 - ❑ When raw data publishing in a repository is seen as a form of publication, personal rewards include profession recognition as a participant in open access culture, public recognition that the published data are trustworthy (in that they are open to verification and ‘peer review’), and research is made available to a free verification system through the network of peers who are able to duplicate experiments.

METHOD: FOCUS GROUP

❑ Participants:

- ❑ Undergraduate and Graduate students

❑ Recruitment:

- ❑ Advertising study on campus, professor participation in recruiting.

❑ Participant Compensation:

- ❑ Agree with cooperating professors to offer their students extra credit for participation.
- ❑ Coupons donated from campus stores and dining halls.

❑ Research Team:

- ❑ Moderator, Note-Taker, Observer; plus video recording of each session

METHOD: FOCUS GROUP

❑ Focus Group Questions

- ❑ Do you think crowdsourcing is a worthwhile endeavor?
- ❑ How likely are you to participate in a crowdsourcing project on a purely volunteer basis?
- ❑ In your opinion, should participants receive academic or professional credit for this type of work?
- ❑ Would you be more likely to participate at a higher rate if the work was “gamified,” i.e. the user completes certain challenges and then levels up, earning points along the way?
- ❑ How likely are you to participate in an online forum relating to your research?

ANALYSIS PLAN

- ❑ Each team member will be provided with a complete transcript of the session.
 - ❑ They will code the data individually, then compare coding schema among the group, creating an agreed upon and nuanced scheme.
 - ❑ Implement the coder reliability model to map out their data and generate a word cloud to allow them to visualize patterns, upon which they will base their content analysis.