Collecting Practices for, and Perceived Research Value of, Social Media Data

Brian Dietz @diuhtez
NCSU Libraries
Why archive social media data?
Who’s collecting?

• Library of Congress
• British Archive
• State Archives
• Academic Libraries
Research Value

- Epidemiology
- Public Health
- Anthropology
- Psychology
- Linguistics
What about archival researchers?
LSTA EZ Innovation Grant
Surveys

Researchers

• Perceived archival research value of social media data

Cultural Heritage Professionals

• Perceived archival value of social media data
• Collecting practices
Researchers
Sample

• SCRC registered researchers in past 2 years

• Excluded researchers who worked for Libraries, e.g., staff, graduate students who worked in special collections

• Contacted via email

• 308 sent, 28 bounced

• 72 started, 70 finished (25%)
Demographics

- 17% Undergraduates
- 29% Graduates
- 16% Faculty
- 34% Other 24 (Researcher, Friend, Independent, Alumn, Staff, Private citizen, etc)
- 4% Blank
Serious discourse occurs on social media?

- 45% Agreed
- 22% Strongly agreed
Value in using social media data in research?

- 34% Agreed
- 37% Strongly agreed
- 71% Total!
Important platforms for archival researchers

- 63% Blogs
- 58% Facebook
- 45% Twitter
- 31% Instagram
- 21% Google+
- 15% Flickr
- 12% Other (LinkedIn, Reddit)
- 6% Won’t be important
Open Ended Responses
I strongly believe in the relevance of this information because it is the "front lines" of movement development--this is where the important ideas and debates are happening. Traditional academic spaces are usually behind (it takes 1-3 years for articles and books to be published) and, again, they tend to bias in favor of whites, men, and long-standing leaders. Ignoring social media means ignoring marginalized voices and it thus provides an incomplete picture of the movement.
If we are to begin actively archiving and using social media content, plans need to be developed as to what we are saving and who social media portrays and how it portrays individuals and large communities.
I fear the sheer bulk of social media data will be a hindrance to anyone seeking to utilize it as for archival purposes.
Cultural Heritage Professionals
Sample

- Attempted to ID archives, special collections, museums, historic sites in North Carolina
  
- Contacted 45 CHOs via email
  
- 22 started, 18 finished (40%)
Demographics

- 44% Small to medium academic libraries
- 22% Research libraries
- 22% Museums or historic sites
- 6% Public library archives
- 6% No response
Social Media in Collections

- 50% Social media data in collections, but not in significant amounts
- 39% No social media in their collections
Platforms Represented

- 53% Blogs
- 41% Facebook
- 41% None
- 29% Twitter
- 12% Flickr
- 6% Instagram
- 6% Other (Tumblr)
- 0% Google+
Future of collecting

- 38% Very likely
- 13% Likely
- 51% Total!
Open Ended Responses
Along with email, social media will probably provide the main source of information for researchers studying our current time. However, our institution just does not have the resources right now to collect and store the social media of other people or organizations.
There are ethical questions about the participation of others being captured. There's a limit to the extent of the conversation you can capture, so relying on the captured content only does not have the breadth of the entire conversation.
Conclusions

• 71% of surveyed researchers saw future value in using social media as a source for research

• Only 51% of surveyed cultural heritage organizations thought it was likely that their institution would archive social media in the future
Possible Next Steps

• Expand survey

• Revise to look for correlation between researcher and CHO
Conclusions: Restraining Factors

- Ethical and legal ambiguity
- Insufficient technical resources
- Lack of curatorial tools
- Researcher access models
Social Media Toolkit

• Environmental scan

• Legal and ethical analysis

• Documentation of tools and collecting methods

• Uses in research and other areas

• Annotated bibliography
Social Media Combine

- Social Feed Manager Twitter collecting software from GWU Libraries
- Lentil Instagram collecting software from NCSU Libraries
- A full pre-configured server environment to support all included software
- A single web-based configuration interface for all included software
Thanks!

Social Media Archiving Toolkit
http://go.ncsu.edu/smalt

Social Media Combine
http://go.ncsu.edu/smcombine

Brian Dietz
bjdietz@ncsu.edu

Jason Casden
jmcasden@ncsu.edu