

Methods for Capture of Social Media Content for Preservation in Memory Organizations

Robert S. Comer

School of Informatics and Computing
Department of Human Centered Computing
Indiana University at Indianapolis
1-317-278-2938

rscomer@iupui.edu

Andrea J. Copeland

School of Informatics and Computing
Department of Library and Information Science
Indiana University at Indianapolis
1-317-274-0114

ajapzon@iupui.edu

ABSTRACT

The paper explores how a Center for Personal and Community Heritage Informatics could effectively address the problem of preserving valuable content shared through social media sites.

1. PROBLEM

Digital content created by individuals and community groups documenting our collective social history is not being strategically preserved and is dispersed across commercial platforms in myriad formats [1,2]. When not formally connected to institutions (academia, corporations, government), content creators lack access to resources, preservation infrastructure, and the knowledge needed to preserve born digital content and create long-term access to cultural heritage. A vast array of information is shared and stored on social networking sites, blogs, and web pages, from text to images to recordings of music and spoken word to audio-visual works. This content documents our collective social history and is primarily in the hands of corporate entities that could cease to exist tomorrow. For the most part, memory organizations are not involved in collecting and preserving the billions of words, images, sounds, and motion media created and stored digitally by millions of people who create this collective heritage.

2. SOLUTION

A Center for Personal and Community Heritage Informatics would empower individuals and community groups to engage in sustainable information preservation. As it is currently envisioned, the Center will serve as an administrative and collaboration hub, working with individuals, community groups and cultural heritage organizations to support research and development in digital preservation related to these groups. The Center will provide both a scholarly venue and a public forum to bring individuals together who are concerned about public memory and community heritage.

2.1 Process

Customizable collection management tools and educational resources developed by the Center will be disseminated throughout the community via the well-distributed network of local public libraries and other memory organizations. Ultimately, users in the community will be able to access informational programming and platforms, and tools through the public library branches to curate and submit media collections [3]. The Center will also serve to connect individual and community archives to

the appropriate heritage organizations to ensure long-term sustainability to these types of collections.

2.2 Benefits

The Center's prototype archival and analysis system could provide public libraries and other local memory organizations the capacity to build and use archives to the benefit of their communities [4]. Further we can begin to understand how to connect individuals and diverse groups to an infrastructure of support that will ultimately benefit people's personal and community information needs while preserving our collective cultural heritage.

3. AREAS FOR EXPLORATION

In order to deal with issues related to scale, the use of distributed storage and a shared metadata repository is recommended. Further, regarding the sheer volume of born-digital everyday information produced, appraisal and selection will be encouraged along with the development of collections with well-defined parameters relevant to local communities. It would be helpful to:

- * Explore the legal and technical opportunities and challenges related to: capturing content shared via commercial platforms, and exporting this content to the digital collections of relevant memory organizations for long term access [5]. Given that the terms of service for most social media platforms gives the corporation a permanent license to any content contributed to their sites, arrangements could be made with corporations to have that license transferred to the Center.
- * Propose methods to support these processes from the perspectives of the users of these sites, the producers of these sites, and the memory organizations who are likely to have an interest in preserving this content.
- * Explore cases where it may make more sense to capture and preserve personal content and its social context directly from these platforms. Through the development of terms of service for the Center, a registry could be formed that would grant the Center (or participating library) permission to capture content posted by individuals to existing social media platforms.
- * Present specific proposed tool development to meet the digital archiving needs of two distinct communities. We have begun work with two different community groups for the development

of prototype tools to illustrate how the Center would promote digital preservation for the public good and facilitate connections: the Indianapolis bicycling community and the Bethel AME Church of Indianapolis. These groups exemplify where social history will be lost without efforts which make connections between a formal support infrastructure and the knowledge of individual community members.

* Explore frameworks for collaboration between memory organizations and projects developed through the Center. Public libraries become conduits for individuals and community groups to create their own archives. The Center would work with public libraries to develop tools and training to empower community groups to preserve their own histories with the possibility of connecting to a formal preservation infrastructure.

4. CPCHI COMPARISON

Tables 1 and 2 together provide a comparison of CPCHI to other initiatives that currently seek to preserve, organize, and make accessible large-scale digital collections for the public good. The comparison includes the Internet Archive (IA), Digital Public Library of America (DPLA), and International Internet Preservation Consortium (IIPC).

Table 1. Comparison of CPCHI to Similar Initiatives: Scope, Perspective, and Stewardship

	Scope	Perspective	Stewardship
CPCHI	Personal & Community-Based Collections	Bottom-up	Shared; distributed storage; metadata repository
IA	The Internet & Some Collections	Internet Harvesting & Archiving; Mostly top-down except for Archive-It and partners, and the Open Library & Scanning Services Project	Central location: DLT Tape, Hard Drives
DPLA	American's Heritage	Top-down	Shared; distributed storage; metadata repository
IIPC	Internet Content with an International Focus	Top-down	Clearinghouse approach with links to separate digital libraries

Table 2. Comparison of CPCHI to Similar Initiatives: Partners and Purpose

	Partners	Purpose
CPCHI	Potential partners: memory organizations, esp public libraries; American Library Association, Omeka Open Source Community	To connect individuals and community groups interested in preserving their histories, with public libraries and other memory organizations, as well as gain access to the knowledge and tools necessary to create long term access to their own cultural heritage.
IA	Library of Congress, Smithsonian	Offer permanent access for researchers, historians, scholars, people with disabilities, and the general public to historical collections that exist in digital format.
DPLA	Approximately 30 institutions: government agencies, libraries, digital consortia, universities	Aims to expand the realm of openly available materials, and make those riches more easily discovered and more widely usable and used, through its three main elements: a portal, platform, and advocacy for a strong public option
IIPC	Organizations from over 45 countries, including national, university and regional libraries and archives.	Dedicated to improving the tools, standards and best practices of web archiving while promoting international collaboration and the broad access and use of web archives for research and cultural heritage.

5. CONCLUSION

Most formal digital preservation efforts are top-down and do not focus on the preservation needs of individuals and local community groups. Currently, social media platforms offer these groups a means to create and share content that represents individuals and their communities. However, given the commercial nature of these platforms, there is no commitment to the public good or preserving our collective social heritage. Given the improbability of saving all the digital content shared through social media, individuals, community groups, and society at large will have to make value judgements regarding what content is worth the effort of saving. CPCHI would facilitate access to the tools necessary for local communities to build collections and connections to local memory organizations, especially public libraries, to help educate those groups regarding information appraisal, description, organization, and preservation.

6. REFERENCES

- [1] Cloonan, M. V., & Mahard, M. (2014). Digital Preservation: Whose Responsibility? *Annual Review of Cultural Heritage Informatics*: 2012-2013, 3
- [2] Jeffrey, S. (2012). A new Digital Dark Age? Collaborative web tools, social media and long-term preservation. *World Archaeology*, 44(4), 553-570.
- [3] Copeland, A. (2015). Public Library: A place for the digital community archive. *Preservation, Digital Technology & Culture*, 44(1), 12-21.
<https://scholarworks.iupui.edu/handle/1805/6117>
- [4] Copeland, A. & Barreau, D. (2011). Helping people to manage and share their digital information: A role for public libraries. *Library Trends*, 59(4), 637-649.
<http://hdl.handle.net/1805/4555>
- [5] Lipinski, T. & Copeland, A. (2013). Look before you License: The Use of Public Sharing Websites in building Patron Initiated Public Library Repositories. *Preservation, Digital Technology & Culture*, 42(4), 174-198.
<http://hdl.handle.net/1805/4574>