
Ethics in Web Archiving

Meghan Dougherty

Loyola University Chicago, School
of Communication
51 E. Pearson
Chicago, IL 60611
mdougherty@luc.edu

Kirsten A. Foot

University of Washington,
Department of Communication
Box 353740
Seattle, WA 98195 USA
kfoot@uw.edu

Steven M. Schneider

SUNYIT, School of Arts & Sciences
P.O. Box 3050
Utica NY 13504 USA
steve@sunyit.edu

Copyright is held by the author/owner(s).
CHI 2009, April 4 – 9, 2009, Boston, MA, USA
ACM 978-1-60558-246-7/09/04.

Abstract

Researchers and archivists worldwide have begun to investigate the scholarly potential of web archiving as a complement to the study of the 'live' web. But the grounding of ethical research decisions in web archiving projects is still murky. Archiving web artifacts is a challenge. The procedures require detailed decisions about inclusion and exclusion, meaning, and access. It forces us confront our definitions of public and private, of published and unpublished, and of producer and audience. This proposal offers an exploration of the ethical questions that arise during the building a web archiving project.

Keywords

Web archiving, research ethics, web research, archival research

ACM Classification Keywords

K.4.1 Public Policy Issues, K.7.4 The Computing Profession

There is a growing recognition within digital scholarship and cultural heritage that digital culture — the cultural artifacts, or material culture, of the web — is heritage worth preserving and understanding. Research methods in web history, namely the practices used to stabilize the volatile material culture of the web and communicate its historical narratives, are shifting as scholars and archivists collaborate.

Developing best practices for archiving material culture of the web is highlighting transformations in scholarship and heritage. New methods for research online are blurring the boundaries between researcher and participant; new modes of collaborative work on the web are blurring the definition between producer and user. Web archiving is broadening the scope of responsibility as scholars collaborate with and even become their own archivists [2, 4, 5, 6, 7].

Our case study identifies questions and tensions that emerge when using web archiving as a scholarly method. Archiving web artifacts is a challenge. Research methods for the web require far more than the simple adaptation of traditional methods to an online environment, and archiving web content in a way that is good for scholarly research requires more than the simple saving of digital cultural artifacts.

Research methods dealing with the interpretation of material culture, documents or other mute evidence are concerned with the ethical issues of inclusion and exclusion, replicability, and interpretation of enduring artifacts that are separate from the author producer or user.

Researchers have begun to investigate the potential of material culture of the web as a resource to complement study of the 'live' web. But, there is little agreement, shared practices, or common purpose from which to begin grounding the ethics of web archiving as a scholarly method. This untethered aspect of the nascent field has benefits – it allows for experimentation in a wide variety of paths resulting in myriad possibilities for future development. But, it also has drawbacks – researchers examining the web have little use for archives that are not stabilized, or

comparable, documented or accessible. Collections of web objects pose interesting ethical problems to archivists and scholars, alike.

Ethical questions for archiving and archival research

There are ethical concerns in using traditional archives in research. Three overarching issues that guide practices are privacy, copyright, and security. Practitioners aim to protect privacy and guard identity; they question the inclusion and exclusion of artifacts in an archive that is *about* something, and they question the resulting meaning made within that archive. The actual collection and use of archived materials is guided by copyright law. The legal rights of researchers to access, use and reproduce archival material are typically covered by fair use guidelines, but the ethical questions of fair use cover a broader scope and may vary depending on institutional policy. Security questions can arise when some artifact becomes archived and aggregated when it should not have been available in the first place. These policies may change when the researcher is also acting as the archivist, and the typical archiving activities of collecting, maintaining and serving an archive are merged with the scholar's typical research activities of interpreting and analyzing. Each of these issues gains new dimensions when moved into an online environment.

Ethical questions for web archiving

Web archiving takes on a new flavor by being a merger of stewardship methods and scholarly methods, and also takes on special characteristics of online communication and collaboration. Scope, anonymity

and reproducibility of take on new meaning in communication and collaboration online [3].

The immediacy, reach and interactivity made possible by material culture existing in an accessible network change power structures — the scope of who and what each web actor can reach is vastly changed.

Anonymity can offer protection in many cases, but create problems for integrity in others. Anonymity can make it "... difficult to develop a reliable history of experiences"[3].

Information and cultural artifacts can be reproduced and stabilized online without loss, and while having not been removed from evolving "live" web circulation [1]. This kind of reproducibility changes expectations of permanence online — especially for users who use the web as temporary storage, or a space for works in progress.

These three overarching ethical considerations confront the archivist-scholar again and again through the process of web archiving. Each consideration shifts in these three special dimensions of online environments when we develop, use and exhibit web archives. The once polarized qualities of anonymous v. identified, published v. unpublished, and public v. private are blurred when we consider how to study the artifacts people build online.

Developing and Documenting Techniques and Best Practices

The basic web archiving operations include collecting, cataloging, and display. For each web archiving operation, there are ethical questions.

The collection stage involves the creation of the archive itself. Rather than being a research byproduct — a collection of data that according to IRB standards will be disposed of or properly stored after research is complete — web archivists may aim to create a reusable resource, an artifact in and of itself.

The creative collection process includes procedures for notification, decisions of inclusion and exclusion and decisions about robot behavior or machine-generated data. Once collection is underway, cataloging begins and poses questions of interpretation and more challenges of interpreting machine-generated data. Display of the resulting archive can vary depending on the goals of the scholar-archivist. The web archivist may want to display portions of archived artifacts to illustrate research, but might also consider wholesale access to and reuse of the archive itself as a function of the research project.

Informing Policy and Practices With the Broader Research Community

As scholars and archivists collaborate to coproduce these dynamic archives, responsibilities will shift. Principal investigators value intellectual freedom, bear a measure of social responsibility, and in some circumstances bear personal liability. IRBs maintain standards for social responsibility and manage institutional liability. Collection commissioners, agents, exhibitors, and site producers have a responsibility to protect intellectual property and prevent harm to users. Law enforcement bodies may become involved with

security concerns. Who should be charged with making what decisions and on what bases? All of these actors will need to work together to navigate the dynamism of web archiving and maintain its potential to offer rich research and heritage resources.

Responsibility in the loss and reconstruction of cultural heritage

Archiving web artifacts is a challenge on many levels. The procedures require researchers to make detailed decisions about inclusion and exclusion, meaning, and access. It forces us confront our definitions of public and private, of published and unpublished, and of producer and audience. The web archivist is also faced with the ethical conundrum of having produced, in addition to scholarship, a research resource — an archive that they may choose to maintain and serve to

References

- [1] Brügger, N. Archiving Websites. The Centre for Internet Research, Århus, Denmark, 2005.
- [2] Dougherty, M. Archiving the Web: Collection, Documentation, Display and Shifting Knowledge Production Paradigms. University of Washington Doctoral Dissertation, 2007.
- [3] Johnson, D. Ethics On-Line. In R. Spinello & H. Tavani (Eds.) *Readings in Cyberethics*. Pp. 30-39. Jones and Bartlett Publishers, Sudbury, MA, 2004.
- [4] Schneider, S.M., Foot, K.A., Kimpton, M. & Jones, G. Building Thematic Web Collections: Challenges and Experiences from the September 11 Web Archive and

others, or not. With each of these web archiving choices comes some level of responsibility that is altered slightly from traditional archiving practices

Archives are becoming dynamic part of web research, and so are shifting primacy from classification, organization, and search to networked modes of remix and discovery. Issues of importance in this shift are the politics of cultural heritage, the democratization of information, and access. The pros and cons hinge on the understanding the how the goals of research and stewardship according to IRBs can merge in web studies. This evolution in research and heritage methods asks us to reflect on ways to analyze, preserve, and understand new media in a manner that is ethically sensitive to the past and to future needs of historical research.

the Election 2002 Web Archive. Paper presented at the European Conference on Digital Libraries Workshop on Web Archives, Trondheim, Norway, August 21, 2003.

[5] Schneider, S.M. & Foot, K.A. The Web as an Object of Study. *New Media & Society* 6(1), Sage Publications (2004), 114-122.

[6] Schneider S.M., Foot, K.A., & Wouters, P. Taking Web Archiving Seriously. In Jankowski, N. (Ed.), *e-Research: Transformation in Scholarly Practice*. Routledge, New York, NY, 2009 (forthcoming).

[7] Schneider S.M., Foot, K.A. Object-Oriented Web Historiography In Brügger, N. (Ed.) *Web History*. Peter Lang Publishing, New York, NY, 2010 (forthcoming).