Property or Privacy? Reconfiguring Ethical Concerns Around Web Archival Research Methods

Meghan Dougherty
Loyola University Chicago
School of Communication
mdougherty@luc.edu

Abstract

People are constantly leaving digital traces of themselves online. These digital traces can be captured and archived to study the evolution of web culture, and changing structure of web objects. As archivists develop practices for digital culture preservation, and scholars build methods for web archival research, they consider the ethical implications of their work. Recently the focus on ethical concerns regarding web archiving has shifted from focusing on property to focusing on privacy. Discourse tracing is used to analyze this focus as is changes over time. This analysis shows how archival researchers and archivists move across and between each other's fields, appropriate, and play with methods, and ultimately construct somewhat limiting frames for understanding archival research ethics.

Keywords

research methods; internet research methods; ethics; web archiving; copyright; privacy

Property or Privacy?

People are constantly leaving digital traces of themselves online in the form of stylesheets, stories, opinions, interpretations, dialog, snapshots, scrapbooks and more. Some of it is public, some private, and some of it is discarded as digital trash, but all of it is the stuff of society. These digital traces can be captured and archived using web archival research methods to study the evolution of web culture, and changing structure of web objects.

To examine these digital traces, researchers have been building idiosyncratic, ad hoc, and large-scale archives of web objects, and conducting studies using those archives as research resources for some time now (Schneider & Foot, 2004; Schneider & Foot, 2005; Brügger, 2009). In that time, conversations about the legal and ethical considerations for archiving someone's web content, storing it, providing some level of access to it, and conducting research on it has shifted from a discussion focusing on copyrights, to a discussion focusing on privacy.

In the early 2000s, when discussing the ethical concerns of archiving Web material for retrospective study, the concern was obtaining copyright permissions from content producers and urging those producers to create content in a format easily archived (Charlesworth, 2003). We built opt-out structures for producers to request their intellectual property be removed from developing archives. The orientation was toward content—intellectual property being copied, stored and made accessible through an archive. Ethical concerns were first and foremost whether or not permission was granted to copy, store, and reproduce content in a new context. Secondarily, but not less importantly, we were concerned with what copied content reproduced in a new archival context might reveal about the people behind the original content (Rauber et al., 2008; Dougherty, et al., 2010). But this was a social science research problem—something to be considered when discussing ethics of archival research—it was not treated as a primary archival or informational concern.

In contrast, the same conversations started to change shape in the 2010s. Ethical concerns about archiving Web material for retrospective study began to place increasing focus on the construction of identity through information behavior and informational traces we leave in online spaces, and so shift to concerns of privacy over property. There is increasing attention paid to how individuals' bits of information may be reinterpreted or misinterpreted as we save it, remix it, and republish it. We question how social network sites pledge to use, or not use information, and balance the expectations

about how that information may be reconfigured beyond our control. Our orientation has shifted from treating that information as property someone might want to protect to treating it as extensions of self.

These are not new concerns. Archivists have addressed ethical considerations in their own practices, have been concerned with the ethics of interpretation of archived material, and have their own code of ethics (Society of American Archivists 2011). Researchers have explored identity construction in online spaces as a function of information behavior, and have repeatedly revisited the ethical concerns in collecting data and studying identity-constructing practices (Buchanan 2006; Buchanan & Ess, 2009; Zimmer, 2010; Papacharisi & Gibson, 2011).

However, as the two begin to experiment across fields (e.g., researchers archiving web content on their own or in collaboration with archivists, and conversely archivists seeking out internet researchers for clues about what they need from web archival resources), each is reaching back into their disciplinary pasts to reconfigure ethical concerns. Archivists focus on copyright and other dimensions of intellectual property they can create clear practices around. Scholars have begun to reintroduce the embedded sense of self that is left in information traces.

This study traces the discourse around the ethics of web archiving as focus shifts between property and privacy. Discourse tracing is used to analyze this focus as is changes over time (LeGreco & Tracy, 2009). Field observations, interviews, best practice guides, and published scholarly work describing archival web research methods—show change over time as well as how these different levels of discourse influence each other. This analysis shows how archival researchers and archivists move across and between each other's fields, appropriate, and play with methods, and ultimately construct somewhat limiting frames for understanding archival research ethics.

This shift has an impact on methods for web archiving as a scholarly research method. It is certainly a problem *for* archiving, but more importantly, it is a problem *of* information. We are quickly reconfiguring the dimensions and definitions of information and content based on social norms in digital networked culture—how information is used and what information represents in different contexts. These seemingly minor shifts will have implications for research, personal histories, and institutional policies, and so the kinds of retrospective web culture studies we may conduct in the future.

Because of the inherently participatory and sometimes personal nature of these informational objects that can be captured in web archives despite having been discarded by the creator, archivists and archival researchers are highly attuned to hidden inclusions and exclusions created by dynamism of information online. Some of these ethical concerns present substantial obstacles in building web archives for research. The digital traces researchers have to work with are changing, so we have to interpret the traces that are left to us differently. We also must reconsider the ethical implications of the methods we use to gather and keep those traces. More than likely, future research will prompt a different kind of engagement with people, and with the data they produce.

Scholars have begun to debate the right to forget as essential in a networked age. But while we protect the right to forget, we must also bear in mind the value of remembering and the opportunities that networked technology offers to help us do so. This simple shift is already influencing our ethical debates around Internet research. It is likely to also influence the structures we build to preserve digital culture. This study suggests that we should reconfigure copyright and privacy in such a way that we can more proactively take seriously the social responsibility of web cultural heritage stewardship as an important part of the making and maintaining of a digital historical record that is accessible for scholarly analysis in a variety of methodological modes.

The debate is not simply a matter of whether or not it is ethical to preserve what some users consider to be ephemeral artifacts in permanent and accessible storage. The debate is far more complicated, involving various information behaviors, conflicting expectations, and different interpretations of how our information online represents our most intimate selves. What's more interesting is how this complicated debate has evolved, and the kinds of changes this evolution calls for.

References

Buchanan, E. (2006). Internet Research Ethics at a Critical Juncture. *Journal of Information Ethics*, 15 (2), 14-17.

Buchanan, E. and Ess, C. (2009). Internet Research Ethics and the Institutional Review Board: Current Practices and Issues. *Computers and Society, 39* (3) 43-49.

Brügger, N. (2009). Website History and the Website as an Object of Study. *New Media & Society, 11*(1-2), 115-132.

Charlesworth, A. (2003). Legal Issues Relating to the Archiving of Internet Resources in the UK, EU, USA and Australia: A study undertaken for the JISC and Wellcome Trust. Version 1.0 – 25. Retrieved from http://library.wellcome.ac.uk/projects/archiving_legal.pdf.

Dougherty, M., Foot, K.A., & Schneider, S.M. (2010). Ethics of Web Archiving. Paper presented at the Computer Supported Cooperative Work pre-conference workshop Revisiting Research Ethics in the Facebook Era: Challenges in Emerging CSCW Research, Savannah, GA.

LeGreco, M. & Tracy, S.J., (2009). Discourse Tracing as a Qualitative Practice. *Qualitative Inquiry, 15*, 1516-1543.

Papacharissi, Z. & Gibson, P. (2011). 15 Minutes of Privacy: Privacy, Sociality an Publicity on Social Network Sites. In L. Reinecke & S. Tepte, (Eds.), *Privacy Online: Theoretical Approaches and Research Perspectives on the Role of Privacy in the Social Web* (pp. 75-89). New York: Springer.

Rauber, A., Wachter, B., & Kaiser, M. (2008). Ethical Issues in Web Archive Creation and Usage: Toward a Research Agenda. Paper presented at the International Web Archiving Workshop, Vienna.

Schneider, S.M., & Foot, K.A. (2004). The Web as an Object of Study. *New Media & Society 6*(1), 94-102.

Schneider, S.M., & Foot K.A. (2005). Web Sphere Analysis: An Approach to Studying Online Action. In C. Hine (Ed.). Virtual Methods: Issues in Social Research on the Internet. United Kingdom: Berg Publishers.

Society of American Archivists. (2011). SAA: Council handbook (App. K-A Code of Ethics with Commentary). Retrieved from http://www2.archivists.org/statements/saa-core-values-statement-and-code-of-ethics.

Zimmer, M. (2010). "But the Data is Already Public": On the Ethics of Research in Facebook. *Ethics and Information Technology, 12*(4), 313-325.

License