
Interrogating the logics of web archiving in the era of platformization

Jessica Ogden*^{†1}, Katie Mackinnon*², and Emily Maemura*³

¹University of Bristol – United Kingdom

²UNIVERSITY OF TORONTO – Canada

³University of Illinois, Urbana-Champaign – United States

Abstract

This panel revisits the ‘promise’ of web archives for platform studies, by examining the various logics that inform how web archiving is done, and their implications for archival use. On the one hand, web archives offer possible alternative means for studying platforms through time, allowing researchers to supplement ethnographic or API-driven methods with longitudinal and ‘multi-layered’ approaches to platform historiography (Helmond and van der Vlist, 2019). On the other hand, web archives continue to be underutilized and have not typically taken into account the breadth and diversity of platform studies research, or how web archiving interacts with the situated (and often, carefully crafted) ethics and communities of use that online platforms afford.

This panel presents three papers on the factors that shape the ‘conditions of possibility’ through which researchers engage with the archived Web. Our research reflects different perspectives on the socio-technical processes, visions and motivations of archivists that have implications for platform archives and the logics that they inscribe. We also highlight key methodological, technical and ethical issues that arise when working with web archives which raise significant questions about how and when researchers *can* and *should* make use of this data.

We close with a discussion of ways to strengthen connections between web archives and platform researchers. We consider how the archiving process could include platform studies researchers to ensure collection practices take into account platform-specific norms, and the situated ethical considerations required for working with these materials. Through this session we hope to chart a new research agenda that leverages the expertise of both the web archiving and research communities and foregrounds critical engagement with the problems of mass data collection. This agenda also looks ahead as the Web continues to evolve beyond the ‘platform-centered’ era, questioning the role of web archiving for new and emerging studies of the decentralized Web.

Web archiving practices and the material logics of collection and discovery

The past 25 years of web archiving have been premised on automated collecting through use of web crawlers. In this context, proprietary platforms have come to present several challenges: in technical terms, platforms can resist web crawling since their site designs and

*Speaker

[†]Corresponding author: jessica.ogden@bristol.ac.uk

behaviours restrict access to automated bots; for web archivists, each platform may require new curatorial approaches to discover relevant materials; in legal terms, platform collecting practices must additionally adhere to each platform's terms of service. Taking a materialist perspective on platform challenges, this paper considers the misalignments between platform affordances and the logics underlying web archiving practices, i.e. the logics stemming from resource discovery via the archival Heritrix crawler and embedded in the design of the WARC file format.

Specifically, I identify how processes of crawling and web-resource discovery are rooted in a vision of the web which has not yet accounted for the configurations of actors and materials that comprise proprietary platforms. I present examples from fieldwork observations and interviews with archivists and researchers, revealing how the tools, workflows, and practices of institutional web archiving programs have inherited logics of material ordering from a 'pre-platform' web. A core finding of this work is that current institutional collecting practices center on datasets structured according to aggregating URIs. The focus on URIs ultimately limit the objects of computational analysis to URI components such as domains, status codes, file types, as well as HTML elements and text. Challenges arise as these more page-specific elements also come with significant demands of computational processing and reconfiguration to make them amenable to research. Additionally, these components do not necessarily align with researcher needs or driving questions, particularly for platform studies.

Looking towards recent work from platform studies, I consider how the discussion of platform affordances, vernaculars and sensibilities from Tiidenberg, Hendry and Abidin (2021) could be better integrated into web archiving tools and practices. Rather than applying existing discovery logics centered on seeds and keywords to platforms based on elements like hashtags, how might archival approaches embrace collecting practices accounting for platform-specific affordances? For example, Tiidenberg et al. note *tumblr*'s affordances of: low 'searchability,' high 'multimodality' and high degree of 'nonlinear temporality' i.e., "where some posts recirculate forever, others blink briefly before being forgotten, and many (networks of) blogs still function as archives" (p. 42-44). I reflect on possibilities for developing new tools and modes of collecting that attend to the situated features of platforms, how they are taken up by communities of users, and how these relationships are studied by platform scholars.

'Crisis collection': Web archiving logics in the face of dead and dying platforms

Despite their power, social media platforms are far from permanent. Shifts in platform policies, governance, revenue strategies (and more) have significant implications for the viability of platform futures, their networks of user communities, as well as our efforts to observe and study them. Public discourse surrounding recent high profile cases of platform decline (such as Twitter, tumblr, Vine and others) point to a growing interest in the role of web archiving (and web archives) in shaping access to platform histories over time. However, the logics that underpin these activities deserve further attention, particularly when examining how and to what extent platform studies researchers can or should make use of the archived Web. In short, how do platform web archives come to be and who, in fact, are they for?

Elsewhere, I have made the case for investigating the socio-cultural dimensions of web archiving, emphasising the ways that practices are fundamentally shaped by *who* is archiving. Here I expand this analysis to include further observations surrounding how web archiving is frequently framed through the discourse of an impending 'crisis' - or a "serious threat to the basic structures or the fundamental values and norms of a system" (Rosenthal, Charles and 't Hart, 1989). To illustrate, I draw on examples from my own ethnographic research, focusing on a case study encompassing the 2018-19 efforts to archive tumblr 'Not Safe for Work' (NSFW).

The case study follows the activities of Archive Team, a "loose collective" of volunteer web archivists, as they attempted to archive NSFW tumblr after the platform announced their intentions to no longer allow 'adult content', nudity and sexually explicit posts. Drawing on Boin and 't Hart (2007), I outline three key elements of 'crisis collection' in web archiving:

threat, uncertainty and urgency. Through examples, I then link each element with Archive Team ‘tenets of practice’ that illustrate how cultural priorities become a set of situated moral commitments in web archiving that shape how we will come to understand platforms in future. The cultural politics of web archiving are revealed through practice dilemmas and negotiations over the selection of which posts and platform components to save, as well as issues surrounding how archivists engage platform access restrictions in the face of tumblr’s resistance to being archived.

I conclude by returning to the value of interrogating web archiving logics, with the aim of critically engaging how the crisis discourse often masks the complex power asymmetries and values at play when archiving is deployed at scale in the face of a ‘dying platform’. By acknowledging the subjective processes behind the identification of particular ‘crises’ and the regular positioning of web archiving as ‘solution’, this framing foreshadows tensions between the goals of archivists, the desires of platform researchers and the mixed reception of these activities by creator/user communities themselves who question what purpose these (often haphazard) archives ultimately serve. It is hoped that this intervention furthers debate on the critical ethical considerations and implications of working with web archives for platform studies.

The Archive Promenade: working against platform logics in web archival research

Historical research often involves primary sources where the authors of personal texts, like diaries or letters, are deceased and the rights to academic use have been approved. Research about the history of the web - both its infrastructures and content - must adapt to engaging with the liveliness of human research subjects whose data are collected and stored in web archives in abundant quantities (Milligan, 2019). The logics of web archiving prioritize achieving access to proprietary data and responding to perceived crises in technological failure and decline. Consideration for the sensitivity of materials captured can be overlooked in these instances, as it might be seen as secondary to the primary goal of preservation and its technological challenges. The logics of platforms, including an emphasis on data quantity, efficiency and visibility over ethical responsibility - as well the terms of service/use that determine limits of individuals’ data privacy - often become embedded into the archive and form the foundation upon which historical internet research can be conducted with web archives.

A feminist ethics of care for archived web data requires new ways of engaging with materials held in web archives. By bringing participants in to find and examine their own digital traces, researchers achieve more than informed consent, they are able to design methodologies that value the relationships people hold with data they have produced online throughout their lives. The Early Internet Memories (EIM) project explored ‘millennials’ (b. 1981-1996) memories of growing up online in the mid 1990s-early 2000s in Canada. In this work, I developed an ethico-methodological intervention that paired oral interview research with web archives, called *The Archive Promenade*. Here researchers take the position of a “vulnerable companion” (Atuk, 2020) with the participant as they move through archived web material – digging, as scavengers, to retrieve what materials of theirs have been dumped. They are moving through the ruins of a digital space that participants once called home, working as active participants in the co-analysis of what these materials mean for the historical record.

I explore how digital traces evoke different affective responses and feelings of attachment, intimacy, and connection. While some affective responses are bolstered by nostalgia for the early web, many of these responses demonstrate the myriad of ways in which people are *databound*: attached to the data they have produced throughout their lives in ways that they both can and cannot control through their ability to socially modulate and determine their information privacy. This framing assists in theorizing the long-term implications of online engagement, the effects of datafication on life and livability on the web, and the role of web archives. This paper demonstrates how the construction of cultural web histories that engage with web materials would benefit from methods that weave in personal narrative,

reflection and co-constructed knowledge of digital space, not only to engage ethics of care, but also to work against the extractive logics of the platform the permeate web archiving practices, collections, and use.

References

Atuk, T. (2021). "Hungry Ethics: from doing research to becoming-sangtin," *Gender, Place & Culture*, DOI: 10.1080/0966369X.2021.1916446

Boin, A., & 't Hart, P. (2007). The Crisis Approach. In H. Rodríguez, E. L. Quarantelli, & R. R. Dynes (Eds.), *Handbook of Disaster Research* (pp. 42–54). Springer.

Helmond, A., & van der Vlist, F. (2019). Social Media and Platform Historiography: Challenges and Opportunities. *TMG Journal for Media History*, 22(1), 6–34. <https://doi.org/10.18146/tmg.434>

Milligan, I. (2019). *History in the Age of Abundance*. McGill-Queen's University Press.

Rosenthal, U., Charles, M. T., & 't Hart, P. (1989). *Coping with Crises: The Management of Disasters, Riots and Terrorism*. Charles C. Thomas.

Tiidenberg, K., Hendry, N. A., & Abidin, C. (2021). *Tumblr*. Polity.

Keywords: platform studies, web archiving, platformization, ethics