WEB HISTORIES IN THE MAKING: WEB ARCHIVES & THE LOGICS OF PRACTICE

Johannes Paßmann
Ruhr University Bochum

Lisa Gerzen
Ruhr University Bochum

Martina Schories
Ruhr University Bochum

Jessica Ogden
University of Bristol

Emily Maemura
University of Illinois Urbana-Champaign

Katie Mackinnon
University of Toronto

Historically-situated accounts of the Web have a long history within the field of internet studies. Drawing on diverse methodologies and forms of data, web histories of platforms, cultures and communities of practice have illuminated the rich, but often transient and shifting nature of life online. Many web histories rely upon researchers capturing, collecting, and generating their own data through time, though some have also engaged with web archives like the *Internet Archive Wayback Machine* as a means for studying the past online. Web archives, and their underlying data, align themselves with the promise of 'data revolutions' (Kitchin 2014) and so-called 'big data' for studying social and cultural worlds, allowing researchers to develop large-scale, longitudinal and 'multi-layered' approaches to platform historiographies (Helmond and van der Vlist

2019; Bruns and Weller 2016). While archives maintain a widespread preoccupation with scale, abundance and 'plentitude' of collections, and collection activities (De Kosnik 2016), these data continue to be underutilised in internet research.

Critical engagement with the nature and validity of web archives as historical sources has begun to be addressed by various researchers engaged in (what could be called) *web archive criticism*. Web archive criticism has identified the opportunities and limitations of web archives *as data* (Rogers 2017; Brügger 2018). Studies have shown how web archives are fundamentally shaped by the infrastructures of web archiving (Summers and Punzalan 2017; Ben-David and Amram 2018; Webster 2017; Dougherty and Meyer 2014; Hegarty 2022b; Ogden, Summers, and Walker 2023), the practices that produce web archives (Summers 2020; Ogden, Halford, and Carr 2017) and the 'archivability' of the Web (Kelly et al. 2013; Banos and Manolopoulos 2016). This form of source criticism is further complicated by the sociotechnical complexity of web archives (Ben-David and Huurdeman 2014; Brügger 2012; Rogers 2017), often making it difficult for researchers to address the provenance of data (Maemura et al. 2018), or trace back the decision-making processes to a single cause.

We contend that web archive criticism has to be taken seriously: Web archive data have never fulfilled the requirements of positivist ideals such as 'representativeness' (Hegarty 2022a) or objectivity, and the methodological consequences of this observation currently do not go far enough. This panel aims to shift and reframe current discussions of the 'promise' of web archives for web historiography, towards identifying what underlying *logics* or *ideals* drive and motivate various actors engaged in this work. We argue that not only do the logics underpinning the practices of collecting and archiving the Web deserve further attention, but also the practices of internet researchers who aim to use these materials for studying the Web. In doing so, we focus our discussion on what follows from web archive criticism: Is the result that web archive data is fundamentally different, or do we also need new ways of studying and conceptualising this difference in a qualitative way?

The first paper presents a "history of web histories" to reconstruct which ideals accompany research methods and practices of web historians, including the ways that scholars have addressed the perceived flaws in web archive data and the implications they have for tools and the positioning of methodological ideals. The second, examines the underlying logics that drive 'crisis collection' in web archiving, critically engaging how these logics often mask the complex power asymmetries and values at play when archiving is deployed at scale in the face of a 'dying platform'.

The third shows that practices of web archiving are shaped by path-dependent, material logics of the pre-platform web, resulting in procedures that do not account for more recent characterizations of the web according to platform-specific affordances,

vernaculars, and sensibilities. And the fourth proposes a method that is concerned with the question of how one can reconstruct practices *through* and *with* archived materials. Here 'archive promenades' provide a view into how people hold attachments to archived web materials in various ways that are often unpredictable. Each paper contribution in this panel builds on *web archive criticism* by situating archived web material as fundamentally tied to the *logics of practice*. These underpinnings affect not only the formation of web archives, but also the methodological approaches researchers take. We therefore suggest new ways for conceptualising the 'doing' of web histories, tying them to an assemblage of people, practice and data that shape how we can come to understand the Web.

A HISTORY OF WEB HISTORIES: DEVELOPMENT OF WEB ARCHIVAL RESEARCH METHODS, TOOLS, AND THEIR IDEALS

Johannes Paßmann
Ruhr University Bochum

Lisa Gerzen
Ruhr University Bochum

Martina Schories
Ruhr University Bochum

This paper wants to foster a discussion on methods in web archiving and its research practices by reconstructing its methodological ideals: To a certain degree, every archive is shaped and structured by *ideals* of archiving. Its practices differ from the ideals (Ogden 2022), however, ideals affect practices insofar, as they are *hic et nunc* solutions oriented towards ideal solutions that are unattainable within a given situation. This is not only the case for practices of archiving, but also for their research practices: It makes a difference, whether the non-representativeness of web archives (Brügger 2012) is understood as an obstacle towards ideally representative web archive studies, or whether one rejects ideals such as representativity or objectivity of archived material and considers incompleteness or non-representativity as constitutive features of archived web data.

We argue that ideals of web archiving do not only shape and structure archiving, but also studies in web histories. Thus, rather than proposing new ideals for web archiving and web histories, we want to reconstruct their ideals retrospectively by analysing a broad range of documents and artefacts; most specifically 1) public accounts of web archives, 2) scholarly publications in web histories, and 3) research tools designed to work with web archive data. Thereby, we differentiate between ideals from *quantitative*

social sciences, namely representativity and objectivity on the one hand, and ideals from *qualitative* studies on the other hand, which can be summarised as 'contribution to a theoretical question' or short: theory.

Public accounts of web archives

Ever since Web Archives started capturing websites in the 1990s, researchers from a broad range of disciplines have emphasised the potential of these new data suddenly available. The promise of 'rogue archives' (De Kosnik 2016) challenging the logics of selection of other historical sources and large amounts of data being freely available to do research with, implied a great chance for various disciplines. The potential is said to have grown even more with the World Wide Web playing an increasingly central role in our everyday lives, making what is happening online seemingly even more important (Meyer et al. 2017).

The *Internet Archive*, which hosts the largest collection of archived web pages, reflects on its legitimacy with a comparison to other media whose need for archiving has been viewed as self-evident for a long time: 'Like newspapers, the content published was ephemeral – but unlike newspapers, no one was saving it' ('Internet Archive: About IA' n.d.). The goal was to archive everything on the World Wide Web, making a so far marginal part of history available and accessible forever.

Scholarly publications in web histories

However, the proposed paper shows not only that practices of web archiving are oriented towards ideals of completeness and identity, but these ideals are also inscribed into research practices. We want to reconstruct how these logics of web archives get in part reproduced in practices of researching with web archives, practices of researching web archives themselves and tools developed to research them and do research with them.

The restrictions and limitations of these new data were reflected early on in the field of Web Histories. Two characteristics in particular are written about a lot in Web Histories' methodological literature: incompleteness and non-identity. For technological and structural reasons, web archives of different kinds cannot archive everything on the web (Ankerson 2015; Burkhardt 2009) and the data is immensely biassed (Thelwall and Vaughan 2004; Ogden, Halford, and Carr 2017; Ben-David and Amram 2018; Paßmann 2021; Ogden, Halford, and Carr 2017). Web Historians believe archived web pages to be an 'actively created and subjective reconstruction of what was once online' (Brügger 2012, 320) and state that 'on the one hand, the archive is not exactly as the website *really* was in the past (we have lost something), but on the other, the archive may be

exactly as the Internet *never* was in the past (we get something different)' (Brügger 2005, 23).

These observations of fundamental incompleteness and non-identity have become the default methodological premise for most works using web archive data. However, the result of these observations is frequently that web archiving in each current form is *not ideal* to save the web. As a result, the implicit ideal active here is, that the data 'should be better', or at least that researchers should fix, clean or sort their data before using it and compare it to 'live data' (Hale, Blank, and Alexander 2017, 47) in order to get more representative results. Brügger recommends to get 'as close as possible to what was actually online' (Brügger 2013, 759) as the goal for comparisons between multiple archived snapshots, to then work with the 'least deficient version' (Brügger 2008, 167).

Methods of web histories are expected to work around the compromises of their data, adjusting their research questions to be answerable with the limited data available (Meyer et al. 2017, 23) or building their own, differently 'incomplete' archive to cater their research needs (Ackland and Evans 2017), detecting the right data sets to test their hypotheses (Cowls and Bright 2017), developing new computational methods to cope with the large amounts of data (Milligan 2017) and cross-referencing web archives with data from other sources (Schafer 2017).

Research tools for archived web data

The paper will also provide a brief look into how ideals are not only present in publications and other written documents, but also in tools. Looking at tools developed for collecting and analyzing data from the Internet Archive, we want to reconstruct which ideals shape the computational methods used in the field. The Internet Archive's *Wayback Machine* is for many researchers the entry point into the web of the past: Depending on how 'complete' their archiving is, it supports the impression of browsing websites as if they were in their original state. Other tools want to make it easier for scholars to use web archive data by providing examples or 'recipes' for a set of standard derivatives possible with most data sets (Ruest et al. 2020) or try to establish comparability with yearly archived versions closest to a certain date ('Internet Archive Wayback Machine Network Per Year: Instructions & Scenarios of Use' n.d.).

While *web archive criticism*, i.e. reflecting the limitations of data and defining the boundaries of one's analysis, is crucial, we ask for its consequences. What would it mean, and how would it change web histories if non-identity and incompleteness are not limitations or obstacles in a quest for representativity, but a taken-for-granted characteristic of web archive data? What if we understand the given tools and methods not as compromises towards ideals that cannot be achieved, but look for methodological ideals that fit better with the available data?

'CRISIS COLLECTION': ARCHIVAL LOGICS IN THE FACE OF DEAD AND DYING PLATFORMS

Jessica Ogden
University of Bristol

Despite their power and proliferation in everyday life, 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 (McCammon and Lingel 2022). However, the logics that underpin these activities deserve further attention, particularly when examining how and to what extent internet 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, the case for investigating the sociotechnical and cultural dimensions of web archiving has been made (Ogden 2022), emphasising how practices are fundamentally shaped by an assemblage of socio-cultural factors and technical actors, as well as *who* is doing the archiving. I build on this analysis to observe the ways that web archiving is frequently framed through the discourse of an impending 'crisis', or "serious threat to the basic structures or the fundamental values and norms of a system, which under time pressure and highly uncertain circumstances necessitates making vital decisions" (Rosenthal, Charles, and 't Hart 1989, 10). Further drawing on Boin et al. (2018), I outline three key elements of *crisis collection* in web archiving that fundamentally shapes practice: the presence and experience of a *threat*, and an impending sense of *uncertainty* and *urgency* that accompanies a crisis moment. Whereas this definition is applicable to a broad array of crises - from natural disasters to infrastructural decay and radical organisational change (ibid., 24) - I use the concept(s) to examine the subjective ways that web archiving is often urgently deployed in response to the impending 'death' of online platforms or removal of online user-generated content. *Crisis* therefore acts as an opportunity or analytical tool for qualitatively interrogating the logics that mobilise, underlie and enact this form of web archiving, whilst also considering the possible disconnects between the perception and experience of the specific threat of de-platforming by different stakeholders (i.e. platforms, archivists, and creator/user communities). I argue that all of the above have implications for using these archives in platform studies.

To illustrate each element of *crisis collection*, I draw on examples from ethnographic and documentary research with Archive Team, a loose collective of volunteer web archivists that have been archiving parts of the Web since 2009. In the first half of the paper I present an analysis of metadata associated with Archive Team projects alongside their running log of dead and dying websites (aptly named the *Deathwatch*). The analysis demonstrates two sets of observations about the perceived *threats* that mobilise archivists and practices associated with web archiving. First, I argue that *Deathwatch* says something about Archive Team itself (and perhaps the broader field); revealing a kind of 'fatalistic optimism' where the act of cataloguing and archiving becomes a hopeful, if perhaps futile, intervention in the face of platform power. Secondly, I reflect on the implications of *Deathwatch* for an understanding of the rise and fall of online platforms since 2009; illuminating a multiplicity in the root causes of platform retirement and consequentially, the inherent ephemerality of user-generated content online.

The second half of the paper focuses 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 as they attempted to archive NSFW tumblr after the platform announced their intentions to no longer allow 'adult content', nudity and sexually explicit posts in 2018. Here uncertainty 'pertains to both the nature and the potential consequences of the threat' (ibid., 25), which in this case is centred on concerns for how tumblr's deplatforming efforts would actually work in practice, as well as additional unknowns about how the cull would ultimately impact a community of tumblr users and the viability of the platform (Tiidenberg and van der Nagel 2020; Tiidenberg, Hendry, and Abidin 2021). The active sense of urgency induced by tumblr's announcement which gave users 14 days notice, in combination with the outcry and 'deep sense of betrayal' expressed by NSFW stakeholders (Tiidenberg and van der Nagel 2020, 74) became a rallying call for Archive Team.

Through examples, I link each element of crisis collection with two Archive Team 'tenets of practice' that are indicative of a particular set of moral commitments or *ideals* that fundamentally shape how the group frame and enact the value of web archiving in response to platform change. The cultural politics of web archiving are revealed through negotiations over the selection of which posts and platform components to save, as well as issues surrounding how archivists navigate platform access restrictions in the face of tumblr's (active) resistance to being archived. Wrapped in the urgency and uncertainty of an impending cull of NSFW accounts and posts, practice decisions and dilemmas concerning what to archive are shown to be contingent on the technical infrastructure available (bandwidth, IPs, worker-nodes), the value-judgements and priorities of participants, and the sociotechnical affordances of tumblr itself. I argue that the logics of 'abundance' and quantity (as measured in the number of URIs captured) ultimately trumped any passing commitments to 'representativity' voiced by some Archive Team participants.

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 researchers and the mixed reception of these activities by creator/user communities themselves who question what purpose these archives ultimately serve. It is hoped that this intervention furthers debate on the critical ethical considerations and implications of working with web archives for internet studies.

RECONCILING WEB HISTORIES WITH THE MATERIAL LOGICS OF CRAWLERS

Emily Maemura
University of Illinois Urbana-Champaign

Over the past 25 years, methods for web archiving have been premised on automated collecting with web crawlers. Crawlers (aka spiders or bots) automatically discover new web content by taking a list of 'seed' URLs, analysing and visiting the URL links within them, and then continuing to add and discover new URLs with the potential to run 'infinitely' (Milligan 2016). The use of crawler technologies both enables and demands collecting at an unprecedented scale, and this form of web crawling is the basis of the Internet Archive Wayback Machine which has captured over 625 billion webpages and more than 15 petabytes of data (Goel 2016). However, several known problems arise in the use of web crawlers, in particular when capturing materials from proprietary platforms: in technical terms, platforms can resist web crawling since their site designs and behaviours restrict access to automated bots; for web archivists, changes to web sites and platforms require constant monitoring and updated curatorial approaches to capture relevant materials and configure the crawler with appropriate 'seeds'; in legal terms, collecting practices must additionally adhere to each platform's terms of service. In order to determine the underlying causes of these myriad challenges, this paper presents a material analysis of archival web crawler technologies and the forms of data they generate and rely upon, ultimately interrogating the logics that are embedded within these artefacts and identifying how their development is situated within organisational and historical contexts.

Specifically, a materialist perspective is adopted here to consider the misalignments between platform affordances and the logics underlying web archiving practices. The analysis is premised on viewing digital materials through the lens of 'materialities of information representation' which Dourish uses to consider how particular instantiations

of data serve to 'constrain, enable, limit and shape' subsequent uses and applications (Dourish 2017, 6). Using this approach, Dourish argues that "material arrangements of information—how it is represented and how that shapes how it can be put to work—matters significantly for our experience of information and information systems" (ibid., 4). Extending this perspective to study web archiving practices, I first analyse the material arrangements of the open-source Heritrix archival crawler as an information system, how Heritrix writes data into the standardised WebARChive (WARC) file format, and how both artefacts represent the needs and constraints of particular organisational contexts which subsequently shape and delimit the work that is possible with web archives. I additionally examine the effects of these design choices by studying the practices of researchers engaged in data-centred studies of web archives in two settings: the "Probing a Nation's Web Domain" project with the Danish Royal Library and researchers from Aarhus University's NetLab, and the "Archives Unleashed" project and a datathon event they hosted in Vancouver, B.C.

The material analysis reveals several ordering principles embedded in web archiving's central artefacts, demonstrating 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. For instance, a core finding (as explored in greater depth in Maemura 2023) is that the WARC's design pre-configures the URL as the central object of analysis for work with web archives. This design choice must also be read in historical context, since the WARC was developed alongside the Heritrix web crawler, whose development dates to 2004, and is based on older crawler technology from web analytics company Alexa Internet, which was founded by Internet Archive's Brewster Kahle in 1995. While Alexa Internet's crawlers were instrumental to support URL-based discovery and analytics for a Web that had evolved beyond the use of directory listings, that vision of web discovery and use is inherently tied to a specific period of the Web. While the Web has continued to evolve through periods of platformisation, and an emerging era of decentralisation, the older logics of discovery continue to persist in the foundational tools and technologies of web archiving today: examining the WARC as a data artefact reveals how it is ordered according to the logic of the crawler, and does not capture the broader range of curation choices and decisions made by archivists (Maemura 2023). In effect, the core tools, workflows, and practices of institutional web archiving programs have inherited logics of material ordering from a 'pre-platform' web.

These material logics embedded in crawler design do not only affect collecting practices, but extend to the ordering of WARC data which is increasingly being made available for research analysis. Observation of research practices with the "Probing a Nation's Web Domain" and "Archives Unleashed" projects reveals how a focus on computationally aggregating and filtering this URL-centric data ultimately limits the researchers' objects of analysis. The order and construction of the WARC's data

artefacts guide and lead researchers to extract and examine URL components such as domains and subdomains, HTTP status codes, file types, as well as HTML elements and text. Yet these elements are not always translatable or transformable into objects of scholarly interest such as online communities, events, or discourse. Additional challenges arise for researchers since extracting and reconfiguring granular elements from within web pages also come with significant demands of computational processing, time, and financial resources to make them amenable to specific analysis methods.

Therefore, even if it were possible to address the laundry list of 'known problems' that archivists encounter with web crawling, a higher-level issue remains: current web archiving methods pre-configure tools, data artefacts, and objects of analysis which are not easily aligned with researcher needs and their driving questions. I conclude with a discussion of ways forward for updating and revisiting approaches that web archiving could employ. For instance, 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 centred on seeds and keywords to platforms based on elements like hashtags, how might archival approaches embrace collecting practices accounting for platform-specific affordances? How might web archiving incorporate what Tiidenberg et al. identify as the affordances of tumblr as a platform: 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" (ibid., 42-44). I conclude by reflecting 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.

TOWARDS ALTERNATIVE METHODS FOR QUALITATIVE WEB HISTORIES

Katie Mackinnon
University of Toronto

Since 1996, the Internet Archive has crawled and archived snapshots of the open web, making it available from anywhere in the world through the Wayback Machine. Over 673 billion archived snapshots are available on the IAWM, which have served as legal evidence in judicial processes (Eltgroth 2009; Howell 2006) and in digital journalism (Ryfe, Mensing, and Kelley 2015). However, The Internet Archive can be interpreted as not so much an archive, but rather as a "large, unsorted store" that lacks assessment of content, communities, and cultures of use, making it more of a container and dumpster, where "things are neither lost nor preserved" (Sofia 2000; Hogan 2015). This is seen in

the large-scale dumping of early web histories, which are often only undertaken as an emergency measure; when online communities are under threat of deletion or removal, web archivists often step in to prevent the loss of materials (Dillon et al. 2019; Ogden 2022).

This scrape and dump approach to preserving the web is, on the one hand, what makes much of early web history research possible: without these records, much would be lost. On the other hand, it is volatile and incredibly costly in terms of ethical concerns and environmental impact. What happens to memory when we opt towards hoarding over curation, appraisal, sorting and organising (Hogan 2015)? Rather, we should retain the right to curate materials that create stories, and intentionally include deletions, gaps, silences and absences.

Since the logics of web archiving prioritise 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 they are often seen as secondary to the primary goal of preservation and its technological challenges. Additionally, the logics of platforms, including an emphasis on data quantity, efficiency, visibility over ethical responsibility, and 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.

Early Internet Memories

The Early Internet Memories (EIM) project explores 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 pairs oral interview research with web archives of youth data, called *archive promenade*. In this approach, researchers take the position of a "vulnerable companion" (Atuk 2022) 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.

The archive promenades activate Walter Benjamin's adaptation of Charles Baudelaire's *flâneur*: a mode of critical wandering through urban modernity. This character has been used previously by danah boyd (2007), as the "digital flâneur:" who straddles the line between public and private in digital spaces, looking to see and be seen, and contributing to the making of a mass public. The "information flâneur" (Dörk, Carpendale, and Williamson 2011) as a mode of data analysis is one way to grapple with the magnitude of big data by "browsing through data without specific questions or goals in mind" while pursuing "curious, creative and critical information seeking". Lauren Berlant (2011) similarly invoked the flâneur to describe how the historical

Euro-modernist shock of mass society was mollified by the practice of the flâneur, who scans and collects the present, relieving them from crisis as they can maintain mental distance between themselves and the rest of the world (ibid., 8).

In my use of the term, the researcher and the participant take on flâneur modes of inquiry as they move through a psychogeographic landscape; that is, digital space that evokes memories and emotions. In mourning the death of old internet communities, Evgeny Morozov declared that "the internet is no longer a place for strolling," suggesting that this motion through the web was a thing of the past. In the archive promenades, I attempt to revive this mode of perusal, of critical strolling, as we move through the ruins of once-vibrant digital landscapes searching for digital traces and adding thick context to web materials. For example, some participants wanted to explore personal websites and blogs, GeoCities homepages, BlackPlanet, MySpace and NeoPets accounts, while others were interested in their involvement in smaller, marginal web social spaces popular in their unique geographic locations, like BlueKaffee, Nexopia, Outaouaisweb and hobby-based forums.

This work engages with a feminist ethics of care for archived web data to demonstrate 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. Reflecting upon the ethics of web archival research, Stine Lomborg (2018) asks, "how can we ensure that the voluntary sharing of personal data at one point in time does not come to negatively impact the research subject at a later point in time?" (ibid., 204), noting that "researching children, social or politically marginalized groups and physically or mentally vulnerable individuals entails a great ethical responsibility to protect participants from being bullied or put on undesirable public display, regardless of whether the online activities we study revolve around their vulnerability" (ibid., 205).

In this project I explore how digital traces evoke different affective responses and feelings of attachment, intimacy, and connection that expand the potential of web archives. 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, digital privacy, and the effects of datafication on life and livability on the web. This paper demonstrates how the construction of qualitative web histories that engage with web archives would benefit from methods that weave in personal narrative and reflection and co-constructed knowledge of digital space.

References

Ackland, Robert, and Ann Evans. 2017. 'Using the Web to Examine the Evolution of the Abortion Debate in Australia, 2005– 2015', 159–89. https://doi.org/10.25969/mediarep/12522.

Ankerson, Megan Sapnar. 2015. 'Read/Write the Digital Archive: Strategies for Historical Web Research'. In *Digital Research Confidential: The Secrets of Studying Behavior Online*, edited by Eszter Hargittai and Christian Sandvig, 29–54. Cambridge and London: MIT Press.

Atuk, Tankut. 2022. 'Hungry Ethics: From Doing Research to Becoming-Sangtin'. *Gender, Place & Culture* 29 (7): 946–60. https://doi.org/10.1080/0966369X.2021.1916446.

Banos, Vangelis, and Yannis Manolopoulos. 2016. 'A Quantitative Approach to Evaluate Website Archivability Using the CLEAR+ Method'. *International Journal on Digital Libraries* 17 (2): 119–41. https://doi.org/10.1007/s00799-015-0144-4.

Ben-David, Anat, and Adam Amram. 2018. 'The Internet Archive and the Socio-Technical Construction of Historical Facts'. *Internet Histories*, March. https://www.tandfonline.com/doi/abs/10.1080/24701475.2018.1455412.

Ben-David, Anat, and Hugo Huurdeman. 2014. 'Web Archive Search as Research: Methodological and Theoretical Implications'. *Alexandria* 25 (1): 93–111.

Berlant, Lauren. 2011. *Cruel Optimism.* Duke University Press.

Boin, Arjen, Paul 't Hart, and Sanneke Kuipers. 2018. 'The Crisis Approach'. In *Handbook of Disaster Research*, edited by Havidán Rodríguez, William Donner, and Joseph E. Trainor, Second, 23–38. New York, NY: Springer. https://doi.org/10.1007/978-3-319-63254-4_2.

Boyd, Danah. 2007. 'Why Youth (Heart) Social Network Sites: The Role of Networked Publics in Teenage Social Life'. SSRN Scholarly Paper. Rochester, NY. https://papers.ssrn.com/abstract=1518924.

Brügger, Niels. 2005. *Archiving Websites: General Considerations and Strategies*. Aarhus: Center for Internetforskning.

———. 2008. 'The Archived Website and Website Philology'. *Nordicom Review* 29 (2): 155–75. https://doi.org/10.1515/nor-2017-0183.

———. 2012. 'Web History and the Web as a Historical Source'. Application/pdf, 77 KB. https://doi.org/10.14765/ZZF.DOK-1588.

———. 2013. 'Web Historiography and Internet Studies: Challenges and Perspectives'. *New Media & Society* 15 (5): 752–64. https://doi.org/10.1177/1461444812462852.

———. 2018. 'Toward a Source Criticism of the Archived Web'. In *The Archived Web: Doing History in the Digital Age*, 137–48. https://direct.mit.edu/books/book/4215/chapter/177576/Toward-a-Source-Criticism-of-the-Archived-Web.

Bruns, Axel, and Katrin Weller. 2016. 'Twitter as a First Draft of the Present: And the Challenges of Preserving It for the Future'. In *Proceedings of the 8th ACM Conference on Web Science*, 183–89. Hannover, Germany: ACM Press. https://doi.org/10.1145/2908131.2908174.

Burkhardt, Marcus. 2009. 'Is There a Way Back or Can the Internet Remember Its Own History?' *Digital Memories: Exploring Critical Issues*, 129–38.

Cowls, Josh, and Jonathan Bright. 2017. 'International Hyperlinks in Online News Media', 101–16. https://doi.org/10.25969/mediarep/12519.

De Kosnik, Abigail. 2016. *Rogue Archives: Digital Cultural Memory and Media Fandom*. https://doi.org/10.7551/mitpress/10248.001.0001.

Dillon, Lindsey, Rebecca Lave, Becky Mansfield, Sara Wylie, Nicholas Shapiro, Anita Say Chan, and Michelle Murphy. 2019. 'Situating Data in a Trumpian Era: The Environmental Data and Governance Initiative'. *Annals of the American Association of Geographers* 109 (2): 545–55. https://doi.org/10.1080/24694452.2018.1511410.

Dörk, Marian, Sheelagh Carpendale, and Carey Williamson. 2011. 'The Information Flaneur: A Fresh Look at Information Seeking'. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1215–24. Vancouver BC Canada: ACM. https://doi.org/10.1145/1978942.1979124.

Dougherty, Meghan, and Eric T. Meyer. 2014. 'Community, Tools, and Practices in Web Archiving: The State-of-the-Art in Relation to Social Science and Humanities Research Needs'. *Journal of the Association for Information Science and Technology* 65 (11): 2195–2209.

Dourish, Paul. 2017. *The Stuff of Bits: An Essay on the Materialities of Information*. https://doi.org/10.7551/mitpress/10999.001.0001.

Eltgroth, Deborah. 2009. 'Best Evidence and the Wayback Machine: Toward a Workable Authentication Standard for Archived Internet Evidence'. *Fordham Law Review* 78 (1): 181.

Goel, Vinay. 2016. 'Defining Web Pages, Web Sites and Web Captures'. *Internet Archive Blogs* (blog). 23 October 2016. https://blog.archive.org/2016/10/23/defining-web-pages-web-sites-and-web-captures/.

Hale, Scott A., Grant Blank, and Victoria D. Alexander. 2017. 'Live versus Archive: Comparing a Web Archive to a Population of Web Pages', 45–61. https://doi.org/10.25969/mediarep/12516.

Hegarty, Kieran. 2022a. 'Representing Biases, Inequalities and Silences in National Web Archives: Social, Material and Technical Dimensions'. *Archives & Manuscripts* 50 (1): 31–45. https://doi.org/10.37683/asa.v50.10209.

———. 2022b. 'The Invention of the Archived Web: Tracing the Influence of Library Frameworks on Web Archiving Infrastructure'. *Internet Histories* 6 (4): 432–51. https://doi.org/10.1080/24701475.2022.2103988.

Helmond, Anne, and Fernando van der Vlist. 2019. 'Social Media and Platform Historiography: Challenges and Opportunities'. *TMG Journal for Media History* 22 (November): 6–34. https://doi.org/10.18146/tmg.434.

Hogan, Mél. 2015. 'The Archive as Dumpster'. *Pivot: A Journal of Interdisciplinary Studies and Thought* 4 (1). https://doi.org/10.25071/2369-7326.39565.

Howell, B. A. 2006. 'PROVING WEB HISTORY: HOW TO USE THE INTERNET ARCHIVE.' *Journal of Internet Law* 9 (8).

'Internet Archive: About IA'. n.d. Accessed 28 February 2023. https://archive.org/about/.

'Internet Archive Wayback Machine Network Per Year: Instructions & Scenarios of Use'. n.d. Digital Methods Initiative Wiki. Accessed 28 February 2023. https://wiki.digitalmethods.net/Dmi/ToolInternetArchiveWaybackMachineToNetwork.

Kelly, Mat, Justin F. Brunelle, Michele C. Weigle, and Michael L. Nelson. 2013. 'On the

Change in Archivability of Websites Over Time'. In *Research and Advanced Technology for Digital Libraries*, edited by Trond Aalberg, Christos Papatheodorou, Milena Dobreva, Giannis Tsakonas, and Charles J. Farrugia, 8092:35–47. Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-40501-3_5.

Kitchin, Rob. 2014. *The Data Revolution: Big Data, Open Data, Data Infrastructures & Their Consequences*. Los Angeles, California: SAGE Publications Ltd.

Lomborg, Stine. 2018. 'Ethical Considerations for Web Archives and Web History Research'. In *Handbook of Web History*, 199–219. SAGE.

Maemura, Emily. 2023. 'All WARC and No Playback: The Materialities of Data-Centered Web Archives Research'. *Big Data & Society* 10 (1): 20539517231163172. https://doi.org/10.1177/20539517231163172.

Maemura, Emily, Nicholas Worby, Ian Milligan, and Christoph Becker. 2018. 'If These Crawls Could Talk: Studying and Documenting Web Archives Provenance'. *Journal of the Association for Information Science and Technology* 69 (10): 1223–33. https://doi.org/10.1002/asi.24048.

McCammon, Muira, and Jessa Lingel. 2022. 'Situating Dead-and-Dying Platforms: Technological Failure, Infrastructural Precarity, and Digital Decline'. *Internet Histories* 6 (1–2): 1–13. https://doi.org/10.1080/24701475.2022.2071395.

Meyer, Eric T., Taha Yasseri, Scott A. Hale, Josh Cowls, Ralph Schroeder, and Helen Margetts. 2017. 'Analysing the UK Web Domain and Exploring 15 Years of UK Universities on the Web', 23–44. https://doi.org/10.25969/mediarep/12515.

Milligan, Ian. 2016. 'Lost in the Infinite Archive: The Promise and Pitfalls of Web Archives'. *International Journal of Humanities and Arts Computing* 10 (1): 78–94. https://doi.org/10.3366/ijhac.2016.0161.

———. 2017. 'Welcome to the Web: The Online Community of GeoCities during the Early Years of the World Wide Web', 137–58. https://doi.org/10.25969/mediarep/12521.

Ogden, Jessica. 2022. '"Everything on the Internet Can Be Saved": Archive Team, Tumblr and the Cultural Significance of Web Archiving'. *Internet Histories* 6 (1–2): 113–32. https://doi.org/10.1080/24701475.2021.1985835.

Ogden, Jessica, Susan Halford, and Leslie Carr. 2017. 'Observing Web Archives: The Case for an Ethnographic Study of Web Archiving'. In *Proceedings of WebSci'17, Troy, NY, USA., June 25–28, 2017*. ACM SIGGRAPH. https://doi.org/10.1145/3091478.3091506.

Ogden, Jessica, Edward Summers, and Shawn Walker. 2023. 'Know(Ing) Infrastructure: The Wayback Machine as Object and Instrument of Digital Research'. *Convergence*, March, 13548565231164759. https://doi.org/10.1177/13548565231164759.

Paßmann, Johannes. 2021. 'Medien-theoretisches Sampling. Digital Methods als Teil qualitativer Methoden'. *Zeitschrift für Medienwissenschaft* 13 (2): 128–40. https://doi.org/10.25969/MEDIAREP/16779.

Rogers, Richard. 2017. 'Doing Web History with the Internet Archive: Screencast Documentaries'. *Internet Histories* 1 (1–2): 160–72. https://doi.org/10.1080/24701475.2017.1307542.

Rosenthal, Uriel, Michael T. Charles, and Paul 't Hart. 1989. *Coping with Crises: The Management of Disasters, Riots and Terrorism*. Springfield, Illinois: Charles C. Thomas.

Ruest, Nick, Jimmy Lin, Ian Milligan, and Samantha Fritz. 2020. 'The Archives Unleashed Project: Technology, Process, and Community to Improve Scholarly Access to Web Archives'. In *Proceedings of the ACM/IEEE Joint Conference on Digital Libraries in 2020*, 157–66. JCDL '20. New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/3383583.3398513.

Ryfe, David, Donica Mensing, and Richard Kelley. 2015. 'What Is the Meaning of a News Link?' *Digital Journalism* 4 (October): 1–14. https://doi.org/10.1080/21670811.2015.1093269.

Schafer, Valérie. 2017. 'From Far Away to a Click Away: The French State and Public Services in the 1990s', 117–33. https://doi.org/10.25969/mediarep/12520.

Sofia, Zoë. 2000. 'Container Technologies'. *Hypatia* 15 (2): 181–201. https://doi.org/10.1353/hyp.2000.0029.

Summers, Ed. 2020. 'Appraisal Talk in Web Archives'. *Archivaria* 89: 70–102.

Summers, Ed, and Ricardo Punzalan. 2017. 'Bots, Seeds and People: Web Archives as Infrastructure'. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, 821–34. CSCW '17. New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/2998181.2998345.

Thelwall, Mike, and Liwen Vaughan. 2004. 'A Fair History of the Web? Examining Country Balance in the Internet Archive'. *Library & Information Science Research* 26 (2): 162–76. https://doi.org/10.1016/j.lisr.2003.12.009.

Tiidenberg, Katrin, Natalie Ann Hendry, and Crystal Abidin. 2021. *Tumblr*. Polity Press.

Tiidenberg, Katrin, and Emily van der Nagel. 2020. *Sex and Social Media*. Society Now. Bingley, UK: Emerald Publishing.

Webster, Peter. 2017. 'Users, Technologies, Organisations: Towards a Cultural History of World Web Archiving'. In *Web 25: Histories from the First 25 Years of the World Wide Web*, edited by Niels Brügger, 170–90. Peter Lang.