ARTICLE

Representing Biases, Inequalities and Silences in National Web Archives: Social, Material and Technical Dimensions

Kieran Hegarty*

Centre for Urban Research, RMIT University, Melbourne, Australia

Abstract

Contemporaneous collecting of the publicly available web has provided researchers with an invaluable source with which to interpret various aspects of the recent past. With millions of websites gathered, stored and made accessible in national web archives over the past 25 years, this paper argues for the need to reflect upon, and respond to, the biases, inequalities and silences that exist in these vast repositories. This article presents a research agenda for web archivists and web historians to together think broadly about the social, material and technical dimensions that shape what is included in web archives, and what is excluded. A key challenge impacting this effort is that various complexities and contingencies of archival formation are obscured. These include wider social inequalities, the entanglement of human and machine decision-making in the archiving process, changing dynamics of power over information online and the environmental impact of technical systems. Accounting for these social, material and technical factors that shape the formation of web archives provides opportunities to develop and use archives in ways that better acknowledge both the strengths and limitations of national web archives as a proxy for the web's past.

Keywords: *National web archives*; *Social inequality*; *Research ethics*; *Bias*.

In 2019, Ian Milligan challenged fellow historians to think about what it might mean to write a history of the 1990s or early 2000s. What would the archives look like? Whose voices would be heard, and whose would be silenced? What are the ethics of using the abundance of (sometimes very personal) information that is now only a few keywords away? These kinds of questions, whilst aimed at historians, are also critical for those building and providing ongoing access to contemporary archives. Consider Milligan's warning to historians:

> Imagine a history of 2019 that draws primarily on print newspapers, approaching this period as 'business as usual', ignoring the revolution in communications technology that fundamentally affected how people share, interact, and leave historical traces behind.[1]

*Correspondence: Kieran Hegarty Email: kieran.hegarty@student.rmit.edu.au

Milligan argues that historians need to put themselves in a better position to use those 'historical traces' that people leave behind online, some of which are captured in web archives. Milligan's work opens a dialogue between those doing contemporaneous collecting and the users of these collections, noting that it is key for scholars 'to become knowledgeable about the construction of the web archives they use'.[2] Untangling the social, material and technical dynamics that help shape the content and character of web archives illuminates both the strengths and limitations of these vast archives as a proxy for the web's past. This article uses the Australian context to examine various complexities and contingencies that are central in shaping web archives yet have either not been fully explored, or are treated separately, in the growing literature on web archives.[3] These include wider social inequalities, the entanglement of human and machine decision-making in the archiving process, changing dynamics of power over information online and the environmental impact of technical systems. Recognising the entanglement of these dynamics might allow the development and use of web archives in ways that better acknowledge both the strengths and limitations of national web archives as a proxy for the web's past.

Context

For the most part, the web of the past, at least in its publicly accessible form, has been gathered, preserved and made accessible by libraries and archives around the world.[4] From the mid-1990s, national and state libraries, particularly, have used their mandates to collect, preserve and provide ongoing access to the documentary record at national, state and regional levels to include online content. Whilst the first, most famous and largest web archiving institution, the US-based Internet Archive, is no doubt the key actor in this space, Australia has also played a key role in the global web archiving movement.[5] After several years of seeking to understand what it would take to absorb web resources into their collection, the National Library of Australia (NLA) released selection guidelines for 'online Australian publications intended for preservation by the National Library' in December 1996. By advancing the idea that 'anything that is publicly available on the Internet is published',[6] these guidelines provided a conceptual framework for the library to absorb a wide array of websites produced by individuals, government, businesses and community organisations into their collections (not just those 'published' in a formal sense).[7]

Whilst a comprehensive history of Australia's web archiving efforts is beyond the scope of this paper,[8] five key developments are important to detail when seeking to address the inequalities, silences and biases that exist in Australia's web archives. First, the NLA decided early on to take a *selective* approach to collecting, with an emphasis on the quality of capture and providing immediate access to content.[9] Because legal deposit did not at this time include online material, the library had to seek permission from the website owner to capture and provide access to their site. An estimate from the late-1990s suggests that archiving just one website would take 5 to 6 hours of staff time, whereas now 'we could shoot a title through in a few minutes', as one NLA staff member put it to me in a 2021 interview.[10] Second, selection is undertaken as a collaborative exercise amongst the NLA, Australia's various state libraries and other major collecting institutions. Third, the NLA has obtained annual 'contract crawls' of the entire Australian country code top-level domain (.au) from the Internet Archive since 2005.[11] Fourth, 2016 saw a key change in Australia's *Copyright Act* that allowed libraries to proactively capture and provide access to content without the written permission of website creators under revised legal deposit provisions.[12] Finally, in 2019, the various resources collectively captured since 1996, along with annual domain crawls since 2005, were made accessible and full-text searchable through the rebranded *Australian Web Archive*, accessible through *Trove*.[13] This history reflects a steady increase in the scale and pace of capturing and

making available content, driven by technical, organisational and legal factors. All these contingencies have shaped both the content and character of the web of Australia's past in critical ways.

Asking big questions of big collections

As suggested, a wide array of factors have shaped the web of Australia's past: the available technology, institutional resources, individual decisions of curators, the processes and ideals of international organisations like the Internet Archive, the prevailing legislative environment and more. These contingencies, I suggest, should not be seen as factors to be concealed on the way to inevitably larger and more accessible collections of content. Rather, these factors should provide critical reflection on the nature of the silences, biases and inequalities nestled within the masses of data that make up Australia's web archives. After a quarter of a century of collecting, I believe it is time to reflect upon questions rarely asked in reference to the archived web:

- What inequalities exist on the web that are carried over or amplified through the process of archiving?
- What gets lost in the process of assembling the web of Australia's past?
- Whose stories are silenced or invisible?
- What ethical protocols should surround the collection, ongoing dissemination and use of content from the web?
- What does all this mean for a critical understanding of the recent past?

These are, no doubt, big questions. But by asking them, one is in a better position to reflect and respond to the work that still needs to be done to ethically create and provide access to archives that respect and represent the complexity and diversity of networked life on this continent.

To help me with these questions, I introduce two related concepts that I will be referring to throughout this paper. The first is the notion of 'representative' collections. Because exhaustive collecting is impossible in the context of the web, those developing national and state library collections aim for selections that are broadly 'representative' of a diverse range of groups, events and topics across society. In recent decades, these aims have been reflected in collection development policies and other strategic documents, which have stated aims to improve representation of 'groups who may not be well represented in library collections and programs'.[14] Traditional approaches, such as the use of census data to ensure linguistic diversity in collections, are well-established, even whilst they have been critiqued for assuming that people conform to strictly bounded identity categories.[15] More recent approaches aim for stronger community partnerships and building capacity amongst groups to help tell their own story (including through community archives).[16]

There are both conceptual and practical limits to the concept of a 'representative' collection. Over the past two decades, the field of critical archival studies has challenged the perception that the archivist is but 'an objective, neutral, passive… keeper of truth'[17] and instead seek to highlight how the *texture* of archives – their regularities, omissions and inconsistencies – reflect prevailing relations of power.[18] Reflecting this, South African archivist Verne Harris has argued 'that in any circumstances, in any country, the documentary record provides just a sliver of a window into the event. Even if archivists in a particular country were to preserve every record generated throughout the land, they would still have only a sliver of a window into that country's experience'.[19] I use Harris' notion of 'the archival sliver' to highlight the power-laden logics that underpin the formation of web of Australia's past. In short,

despite the unprecedented scale of contemporary web archives, they remain saturated with biases, inequalities and silences – it is the job of this article to explore several social, material and technical dynamics that shape contemporary archival formation. Because of the nature of the web – its materialities, its cultures of use, its power relations – these dynamics raise critical questions for those developing and using web archives.

Structure of paper

In thinking about the web of Australia's past as 'a sliver of a sliver of a window into process',[20] I first reflect upon what the web is, who has access to it and what has changed since its emergence in Australia in 1993.[21] From here, I dig into the process of archiving the web, and how changes in the character and governance of content on the web raise challenges to developing representative collections. I then reflect on recent ethnographic fieldwork at the NLA to explore how and when the ethics of capturing and making available the historical traces of people's lives on the web come into the picture. I illustrate in this section that silences do not necessarily reflect a 'gap' that needs to be 'filled'.[22] Rather, silences can reflect people expressing agency over their voice in the archives. Therefore, sometimes *not collecting* something may be the most respectful and ethical option. Finally, I expand the frame to explore what it might to 'represent' our collective, digitally entangled lives to highlight different dimensions of experience. To this, I present recent creative projects that highlight the material, ecological and affective dimensions of networked communication infrastructures. Taken together, these avenues offer promising directions for critical and reflexive engagement with the web of Australia's past.

Archiving inequalities

The first webserver was installed in Australia in 1993, and the system steadily expanded from being a sole concern of academics (initially coexisting with other information systems like Gopher) to include the websites of community organisations, individuals (1994), government, the media (1995) and businesses (1996).[23] As this progression suggests, the web has expanded to include more and more voices, and with it, more and more content. Whilst it is important to distinguish the web (the resource-layer) from the internet that enables its access, it is also crucial to reflect on who has access to the internet (and by extension the web) in Australia, and who has the skills, time and resources to contribute to it. Digital exclusion has significant implications for what voices are, and are not, included in web archives.

Internet access emerged as a social justice issue and concern for policymakers in the 1990s, usually conceived in terms of a 'digital divide'. Whilst access is clearly critical for many forms of social and economic participation, it is worth briefly noting the limitations of the 'digital divide' as a framework for addressing social inequality. As Daniel Greene notes, this narrative reduces 'the complex problem of… poverty to a much more basic binary: a digital divide that could be crossed with the right tools and skills'.[24] It also marginalises the many forms of digital engagement, innovation and resistance by those seen by policymakers to be on the 'wrong side' of the digital divide.[25] Nonetheless, demographic data on access to, and use of, the internet illustrate that social and digital inequalities are mutually constituted. Since it started being used as the measure of the digital divide in 2015, the Australian Digital Inclusion Index (ADII) has shown, again and again, that 'digital inclusion in Australia remains profoundly shaped by geographic and sociodemographic factors such as age, education, income, employment, and location'.[26] In short, digital exclusion is built on top of, and amplifies, broader social inequalities in Australia. This is important to reflect upon as collecting institutions go about building large-scale digital collections that seek to represent the complexity and diversity of life on this continent.

Digging a little deeper into the ADII for the purposes of this paper, I focus on how the skills, time and resources to *produce* content for the web are unevenly distributed across the population. Whilst in the 1990s, user-generated content may have looked like a personal website, and in the early 2000s a blog, over the past decade, this is more likely to be updates, posts and media distributed to an 'imagined audience' on social media platforms.[27] The kind of skills that fall under editing, producing and posting content are labelled as 'creative' in the ADII. As with other measures, these skills are not evenly distributed across the population, with the most 'digitally creative' more likely to be young, employed, abled bodied, on a higher income and with higher levels of education.[28] As such, the raw mass of content that people in Australia contribute to the web is but a sliver of representing the lives of all people on this continent.

Earlier, I mentioned that the contemporary 'historical traces' (to use Milligan's phrase) that people leave behind are now likely to be on social media. Yet, these platforms are hardly a perfect democracy. Whilst there is no authoritative source of data on social media use in Australia, studies have suggested that between 10 and 15% of people use Twitter.[29] The same sources suggest most Australians have a Facebook account, and just over half use the service reasonably regularly. Instagram and YouTube also remain popular amongst most people in Australia, whilst LinkedIn, Pinterest, and/or Snapchat were each used by around 10–20% of survey respondents.[30] However, these raw numbers tell us very little about how actively or passively these users engage with these services. What is clear, however, is that women, Indigenous people and LGBTQI+ people are much more likely to be trolled, harassed and vilified on social media.[31] Needless to say, despite some residual buzz around social media as being inherently more participatory or representative, it is worth taking a broader view to examine the inequalities, silences and biases embedded in who uses these services, and who benefits from their popularity and commercial-driven reliance on virality.

The web archival sliver

From the 'sliver of a sliver of a window into process'[32] that is all the content on the web, what gets assembled in the archives? What is unable to be collected? What are the decisions and contingencies that underpin selection? Whilst web archives may be sizeable, 'web archives and the data they contain do not represent any form of objective or complete knowledge about the past, no more than any other inherently subjective historical method', as Milligan notes.[33] 'More' does not necessarily mean 'more representative'.

'Selection' may be the wrong word to think about what ends up in web archives. The decisions of individual curators and collecting guidelines, whilst important, can be tempered by the technical challenges of capturing a particular website, the prevailing legislative context, the in/ability to obtain permission, and the fact that online content can be removed or changed without a moment's notice. Sometimes one's decision to collect or not to collect is decided by whether it is technically possible to do so. As Valérie Schafer and colleagues note, 'the constitution of heritage is often contingent upon the accessibility of pages, rather than their content – the device determining the (im-)possibility of inclusion, the design becoming prescription'.[34] As such, 'contingencies', rather than 'selection' or 'curation', might be a more appropriate way to think about the factors that drive both the content and character of contemporary collections.

Content on the web, as Schneider and Foot note, is a 'unique mixture of the ephemeral and the permanent'.[35] Whilst librarians and archivists may consider the ephemerality and dynamism of content on the web to be leading to a 'digital dark ages',[36] for users, the fact that content about them is circulating online, or stored and used by third parties, means the right to permanently delete content might be a more pressing need, rather than selection and ongoing

preservation.[37] Yet, a sense of moral urgency over permanently losing information of potential cultural value sees web archiving institutions and actors generally attempt to collect desired content, even if providing access is presently legally or technically difficult.[38]

To illustrate the contingent nature of web archiving, I will briefly explain the process that sees content captured and included in the archives. Archiving websites is achieved through the deployment of automated software called 'crawlers'. After a site is specified in the software (called a 'seed'), the crawler contacts the server where the page is hosted and requests permission to collect the code and files that make up the page.[39] Depending on the specifications of the software, the crawler will then find and follow all hyperlinks on a page, capturing and storing content as it goes. The web archivist might limit the crawler by specifying that it does not stray beyond a particular domain, or a particular part of the website. This is a common practice for site-level curation that continues to be practiced by the NLA and many state libraries. For larger crawls (e.g., the entire .gov.au domain), the crawler is ceased 'when we hit a target, or when we run out of money', as one library staff member told me during my fieldwork at the NLA in 2021. The content is aggregated in a container file called a WARC file, and after a process of 'quality assurance', the content is reassembled using software (such as Wayback) to replay the content as it appeared during the time of the crawl.

From looking at the content, the web of technical, legal and organisational contingencies that lead to something being included in the archive is largely concealed. Science and technology studies scholars call this a 'black box', in that all that we see is an input (the seeds) and an output (the WARCs).[40] But, 'black boxing' web archives limits the questions that can be asked when thinking about representative collections. One may ask, why was *site Z* captured *X times* one year, and only *Y* the next? Why was *site A* captured but not *site B*? Why did the library stop archiving *site C* on a particular date? The mere fact of something existing in an archive – in the past suggesting significance because of its presence in the archives and the material resources taken to collect, catalogue and preserve it – may not signify something significant in the contemporary context. In the move to increasing pace and scale of collecting, the selection of content is increasingly driven by algorithms, rather than being determined solely according to a source's potential historical significance.[41]

With the contingent nature of collecting traces from the web in mind, it is important to recognise that there is a great deal of content on the web that collecting institutions simply cannot capture. Given crawlers travel through the web by following hyperlinks, there are many places where the crawler cannot go. Anything requiring user authentication (e.g., a CAPTCHA code, password, or IP authentication) is out of bounds. Really, any form of user interaction apart from clicking on a link impedes the crawler's journey through the web.[42] For these, and for ethical and legal reasons, web archives really only reflect the publicly accessible, or 'open', web.[43] Whilst those doing the web archiving have come up with an array of creative workarounds to potential problems,[44] there are limits. The migration of content from sites and blogs to platform environments is a key challenge.[45] Because of the nature of social media platforms, web archiving techniques, standards and tools do not translate to the so-called 'walled gardens' of social media platforms.[46] Facebook, for example, is largely closed to crawlers, and Facebook's *Terms of Use* explicitly prohibits 'data mining, robots, scraping or similar data gathering or extraction methods', regardless of the intent around its use.[47] On a web that is – in many cases –'unarchivable by design',[48] an awareness of the contingent nature of web archiving is critical to consider the array of forces that currently exert power over the character and content of the web of Australia's past.[49]

'Ethically important moments' in web archiving

In this section, I reflect on findings from my ethnographic fieldwork at the NLA to explore the process of negotiation that takes place between website creators and web archivists.

The dynamics of these negotiations can shape how and when the content is collected, and how it is made available in the archives. Through a short ethnographic episode that attends to this negotiation, I suggest that silences do not necessarily reflect a 'gap' that needs to be 'filled'.[50] Instead, silences can reflect people expressing agency over their voice in the archives. Therefore, sometimes *not collecting* something may be the most respectful and ethical option.[51] These 'ethically important moments'[52] highlight how the complex dynamics of online sociality challenge strict binaries between open and closed, and visible and invisible.

Whilst many materials that end up in libraries – particularly published materials – are by their nature firmly 'on the public record', content on the web occupies a more ambiguous position. Users share content with an 'imagined audience' in mind – does this include anyone who happens to locate this content in web archives?[53] Furthermore, the web blurs the boundaries between a 'public personality' and a 'private individual'. Keyword searching of web archives enables easy access to content relating to individuals, often dating back decades. Whilst this might be considered by some a mere embarrassment, for others, the implications may be more urgent. This has led a number of researchers using web archives to ask: what do creators and users of these collections owe to the people whose traces of lives they contain?[54] Unfortunately, public debate on privacy, visibility and surveillance often falls along simple binaries of open/closed, public/private and free/proprietary. As Kimberly Christen notes, 'these are not zero-sum games, and information sociality and creatively is more porous than these choices allow us to imagine'.[55]

The practical ethics of web archiving were illustrated to me when, in May and June 2021, I spent 6 weeks at the NLA conducting ethnographic fieldwork in the Web Archiving Section (WAS). During this time, I observed and participated in the everyday activities of staff, conducted formal and informal interviews with current and former NLA staff, and consulted 30 years of reports, memos, minutes and other documents relating to Australia's web archiving program. In relation to ethics and web archiving, I found that web archivists deftly navigate the complex challenges that privacy and visibility raises, whilst seeking to balance the various needs of the library, users of the archive and the creators of sources. To illustrate this, I present one story from my fieldwork.

Until 2016, WAS staff would have to contact the creator of a website to seek written permission to capture their site and make it available. Staff would note how time intensive this was. Nonetheless, it provided an opportunity for staff to interact with website creators. Staff could often find themselves giving advice on how the creator could make their site more amenable to web archiving. Or the creator would have questions about the process or express pride in the fact that their website was included in the national collection. Following a 2016 change in the *Copyright Act*, the library could capture material without having to gain explicit written permission, increasing the efficiency of web archiving considerably. However, it is worth reflecting on how interaction was an opportunity for navigating the ethics of capturing online traces.

During my fieldwork, I was processing titles at the library and noticed that a specific title had conditions attached to the publisher's granting of permission to capture it and make it available.

The correspondence read:

I have decided to grant permission for the Library to [collect my website] … HOWEVER: I wish to be credited simply as [my pseudonym] and do not give permission for my full name to be used in the catalogue record.

This negotiation allowed access to proceed, whilst respecting the rights of the creator. Now, these interactions between humans have been supplanted by machine-to-machine interaction.

Behind every website, there is a user, who will have their own reasons for the creation of the website. Extending its life, or including it in a national collection, may converge or come into conflict with these aims, involving negotiation and compromise.

Reflecting on these 'ethically important moments'[56] illustrates that silences do not necessarily reflect a 'gap' that needs to be 'filled': silences can reflect people expressing agency over their voice in the archives. Collecting is not a binary proposition (collect/do not collect). Navigating the web of the past involves negotiating tensions between the responsibility of collecting institutions to preserve the documentary record, the rights of individuals and groups to decide the fate of their digital traces and the ongoing popularity of social media platforms that seek to control and profit from these traces. All this raises a raft of ethical challenges that require ongoing negotiation and offer methodological possibilities to advance a more 'care-full' research practice.[57]

Embodied, affective and material dimensions of web archives

In this final section, I highlight recent creative projects that illustrate the material, ecological and affective dimensions of networked communication infrastructures. These elements, I suggest, offer promising directions for developing a critical and reflexive mode of archiving and using the web of Australia's past. Together, they push the boundaries of what it might mean to represent contemporary digital life, opening different avenues of experience to critical reflection.

First, whilst it is easy to consider internet-enabled communications as transparent, they are, in fact, deeply material.[58] As Fiona Cameron notes, understanding digital heritage and curation in a 'more-than-human' world means attending to data centres, sensors, robots, cables, earth minerals, land and so on.[59] How can collecting institutions and their users bring these material dimensions of web archiving to the fore?

A range of art projects have sought to raise awareness of the materials required to sustain contemporary technological production and internet-enabled communication infrastructures. For example, Kate Crawford and Vladan Joler's *Anatomy of an AI System* (anatomyof.ai) traces the materials, places and systems required to produce and power one specific AI-powered gadget, the *Echo*, a 'smart' speaker by Amazon. As Crawford and Joler note in a 2018 interview:

> The Echo sits in your house, looks very simple and small, but has these big roots that connect to huge systems of production: logistics, mining, data capture, and the training of AI networks. It's an entire infrastructural stack you never see. You just give a simple voice command… and it feels like magic.[60]

This critical reflection on the planetary costs of commercial data infrastructures offers the viewer an opportunity to understand, and challenge, the increasing scale and pace of contemporary communication systems. Similarly, artist Joana Moll's 2014 online installation *CO2GLE* (janavirgin.com/CO2/) displays in real-time the amount of carbon dioxide emitted from visits to the most popular site in the world – google.com. It starkly displays 'GOOGLE. COM EMITTED [#] KG OF CO2 SINCE YOU OPENED THIS PAGE' in black text on a white background, whilst the number grows each second 'GOOGLE.COM EMITTED 510.49… 1020.98… 1531.47… 2041.94… KG OF CO2…' (see Figure 1).[61] The artwork offers a very different reading of Google, shifting the user's focus from a commercial product with technical affordances to the materiality and environmental costs of data-driven convenience. In this spirit, what would it take to 'read' the web of Australia's past as a web of social and material relations, rather than simply a collection of archived websites?
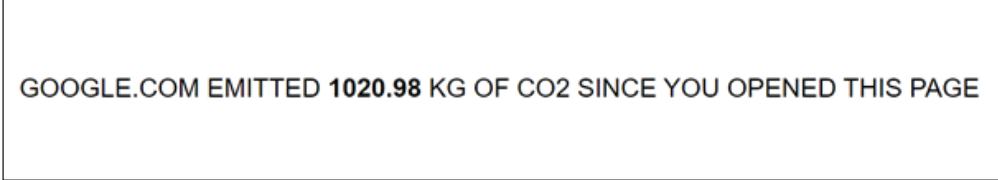
GOOGLE.COM EMITTED **1020.98** KG OF CO2 SINCE YOU OPENED THIS PAGE

Figure 1. Screenshot on Joana Moll's 2014 online art installation, CO2GLE, that displays the amount of real-time carbon dioxide emissions from global visits to google.com (screenshot taken on 13 December 2021). © Joana Moll. Reuse not permitted.

To read the materiality of this particular web of Australia's past, one might start when it is physically located – on unceded Ngunnawal land. Web servers exist in physical space, and when one uses the archive, one uses land. To illustrate the connection between the web and occupation of Indigenous lands, Brooklyn-based designer Caleb Stone developed *Web Acknowledgement*, an extension for the Google Chrome browser that performs an acknowledgement of country based on where the website one is visiting is physically stored (see Figure 2).[62] *Web Acknowledgement* offers an alternative reading of content on the web, mobilising the possibility of unceded land itself as 'a recording medium, an embodiment of the context of creation'.[63]

Finally, I consider what it would mean to capture not only the traces left behind on the web but also the affective dimensions of its use. The 'surfing' (via hyperlinks) of the 1990s is a very different experience of the web than the 'searching' of the 2000s or the 'scrolling' of today.[64] How could these experiential dimensions of the web be captured? The browser emulator, *OldWeb.today* by developer Ilya Kreymer, allows one to navigate web archives using a range of emulated browsers, including the now defunct Mosaic, Netscape and Internet Explorer.[65] This presents the user not only the *content* of an archived webpage but also some of the *experience* of the web in, say, 2001. Understanding the experience of using the web in 2001 would not only involve consideration of the visual culture of the web at this time but also its affective dimensions – the waiting as a website slowly loads, the purring of a bulky desktop computer, the limits on use imposed by cost and access. With this comes the recognition that the web of the past is at once material and affective, produced at a time and place, and involving an array of people, things, machines and environments.

Conclusion

Web archives should not be treated as a 'black box', but rather as a site from which creators and users of these sources can reflect upon the material, cultural and affective dimensions of contemporary digital life. Attending to the contingent nature of archival production illustrates the web of actors and factors that sustain the inequalities, silences and biases existent in these vast repositories. A critical and reflexive approach to developing and using web archives would involve understanding, respecting and, in some cases, challenging, the plurality of ideas of what the web is, could and should be. The seemingly relentless pace and the scale of content creation and distribution in a mediatised world mean that it is time to rethink what 'representative' means, remembering that 'more' does not mean 'more representative'. The next step is to collectively reckon with the ongoing task of collecting, providing access to and using archives in ways that respect and reflect the complexity and diversity of contemporary networked life on this continent.
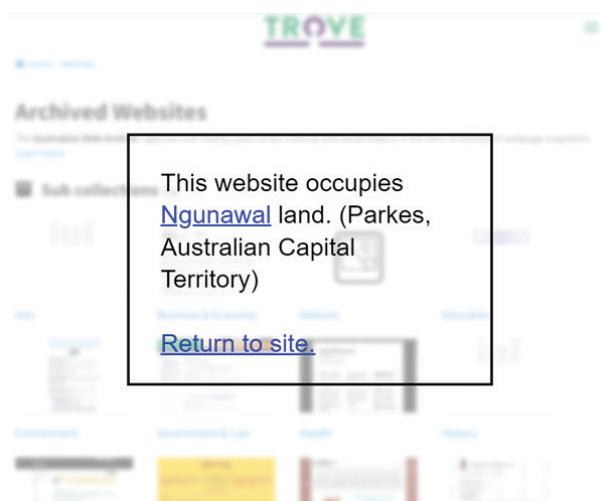
Figure 2. An acknowledgement that the Australian Web Archive website is physically stored on unceded Ngunnawal land, using the Web Acknowledgement extension for Google Chrome by Caleb Stone. The website used is webarchive.nla.gov.au (screenshot taken on 11 December 2021). © Caleb Stone. Reuse not permitted.

Reflecting upon, and responding to, the inequalities, silences and biases that exist in web archives opens a space for several critical interventions for both archivists and researchers. First, the web is certainly not universally experienced as a democratic means to express oneself. As such, archiving institutions could use measures of social and digital inequality to identify those marginalised on dominant channels of online communication, and consider why this may be the case. As I have outlined in this article, silences should not be treated *a priori* as a 'gap' that needs to be 'filled'; rather absence can reflect people expressing agency over the contexts in which they interact. The dominant understanding of absence as fundamentally negative, and the inability to adequately represent absence in institutional metrics, might require new ways of both doing and representing archival work that centre relations sustained through care, ethical responsibility and radical empathy, rather than machine-driven efficiency.[66]

Second, archivists could surface the labour involved in developing, maintaining and providing access to collections. Emily Maemura's recent push for an 'infrastructural description of archived web data' is a step in this direction.[67] 'Web archival labour' could be reflected in catalogue records; however, there are many other ways to illustrate the complexities and contingencies of archival work.[68] For example, from 2012 to 2015, staff responsible for web archiving at Australian's national and state libraries developed a series of regular blog posts that highlighted some of the challenges and peculiarities of collecting online content.[69] For users, these posts provide an engaging and insightful look at the logics of these collections, and the various sociotechnical contingencies that shape web archives. For those wanting to use web archives for research, sustained engagement with those doing the collecting is critical.[70]

Third, there are ways of representing collections that go beyond the dominant mode of access (i.e., playback via Wayback software). For example, web archives could build in optional browser emulators, such as *OldWeb.today*, so that the user can better understand what the page may have looked and felt like a particular moment in history. Providing another example, the State Library of New South Wales has partnered with CSIRO's Data61 Business

Unit to visualise the affective dimensions of social media activity in the state using an 'Emotion Clock' as part of their Social Media Archive.[71] Archiving institutions could also provide users with an insight into the material dimensions of their collections by encouraging artists to experiment with collections or offering users a 'backstage' look at the operation of the institution. This was done very effectively during the NLA's 50th anniversary of the current library building, where the library ran a '50 People of the NLA' promotion on Instagram (see @NLA50ppl) that included photos of library users, staff and machines along with their response to two simple questions: what they do in the library and what they love most about it. Initiatives like this highlight that libraries and archives are more than collections: they involve people and their labour, and buildings, materials and technologies that require regular maintenance and care.

Finally, the web archives community should continue to engage with those using their collections, including the lively field of Internet Studies, which incorporates perspectives from media studies, sociology, cultural studies and more. This need not be onerous and could include signing up for the Association of Internet Researchers mailing list, attending some sessions of their affordable online events, or attending events run by various institutions leading the way with internet research. This will help with developing collections that are used and useful and provide space for greater dialogue between archivists and archive users (and may help productively blur these distinctions). For example, an exciting research agenda is presently being pursued by Canadian researcher Katie Mackinnon, who is forging new modes of using web archives that pay better attention to the contextual and relational nature of ethics involved in researching young people online.[72] Mackinnon wisely encourages us to 'begin with the person rather than their data', and her work has seen her engage website creators in a walkthrough of their archived website, allowing the research participant 'to reconstruct a history of what it meant to them to exist in this space'.[73] Methodological innovations such as these offer a way of understanding the myriad ways researchers are using web archives as part of understanding social and cultural life. Following the lead of these various inventions, innovations and interventions allows us to both acknowledge the strengths and limitations of using national web archives as a proxy for the web's past and push the development and use of web archives in exciting new directions.

Acknowledgements

Notes on contributor

Kieran Hegarty is a librarian and sociologist. After a decade of working in the Australian library sector, Kieran is currently a PhD candidate at RMIT University and a 2022 Digital Humanism Junior Fellow at the Institute for Human Sciences in Vienna. His research focuses on digital technology, archives and collective memory and examines how digitalisation is changing how the past is understood and how the future is imagined. You can find out more about Kieran's research at his website (assemblingtheweb.com) or on Twitter (@assemblingweb).

ORCID

Kieran Hegarty

Notes

1. Ian Milligan, *History in the Age of Abundance?: How the Web Is Transforming Historical Research*, McGill-Queen's Press, Montreal, 2019, p. 28.
2. Milligan, *History in the Age of Abundance?*, p. 87.
3. Emily Maemura, Nicholas Worby, Ian Milligan and Christoph Becker, 'If These crawls Could Talk: Studying and Documenting Web Archives Provenance', *Journal of the Association for Information Science and Technology*, vol. 69, no. 10, 2018, pp. 1223–1233; Ed Summers, 'Appraisal Talk in Web Archives', *Archivaria*, vol. 89, no. 1, 2020, pp. 70–102.
4. Daniel Gomes, João Miranda and Miguel Costa, 'A Survey on Web Archiving Initiatives', in Stefan Gradmann, Francesca Borri, Carlo Meghini and Heiko Schuldt (eds), *International Conference on Theory and Practice of Digital Libraries*, vol. 6966, Lecture Notes in Computer Science, Springer, Berlin, Heidelberg, 2011, pp. 408–420, available at http://link.springer.com/10.1007/978-3-642-24469-8_41, accessed 2 August 2020.
5. Peter Webster, 'Users, Technologies, Organisations: Towards a Cultural History of World Web Archiving', in Niels Brügger (ed.), *Web 25: Histories from 25 Years of the World Wide Web*, Peter Lang, Bern, 2017, pp. 179–190.
6. National Library of Australia, *Guidelines for the Selection of Online Australian Publications Intended for Preservation by the National Library*, National Library of Australia, 1996, available at: https://webarchive.nla.gov.au/awa/19970212062836/http://www.nla.gov.au/1/scoap/scoapgui.html, accessed 13 December 2021, sec. 3.2.
7. Kieran Hegarty, 'The Invention of the Archived Web: Tracing the Influence of Library Frameworks on Web Archiving Infrastructure', *Internet Histories*, forthcoming, 2022. Available at https://doi.org/10.1080/24701475.2022.2103988.
8. Edgar Crook, 'The Work of Pandora', *National Library of Australia Gateways*, August 2006, available at https://webarchive.nla.gov.au/awa/20120322043418/http://www.nla.gov.au/pub/gateways/issues/82/story01.html, accessed 24 August 2020; Hegarty, 'The Invention of the Archived Web: Tracing the Influence of Library Frameworks on Web Archiving Infrastructure'.
9. Paul Koerbin, 'Revisiting the World Wide Web as Artefact: Case Studies in Archiving Small Data for the National Library of Australia's PANDORA Archive', in Niels Brügger (ed.), *Web 25: Histories from 25 Years of the World Wide Web*, Peter Lang, Bern, 2017, pp. 191–206.
10. Ian Morrison, 'www.nla.gov.au/pandora: Australia's Internet Archive', *The Australian Library Journal*, vol. 48, no. 3, 1999, pp. 271–284.
11. Lachlan Glanville, 'Web Archiving: Ethical and Legal Issues Affecting Programmes in Australia and the Netherlands', *The Australian Library Journal*, vol. 59, no. 3, 2010, pp. 128–134.
12. Kelly Buchanan, *Australia: National Library to Implement Digital Legal Deposit from January 2016*, Global Legal Monitor, 2015, available at https://www.loc.gov/law/foreign-news/article/australia-national-library-to-implement-digital-legal-deposit-from-january-2016/, accessed 12 August 2020; Paul Koerbin, 'National Web Archiving in Australia: Representing the Comprehensive', in Daniel Gomes, Elena Demidova, Jane Winters and Thomas Risse (eds), *The Past Web: Exploring Web Archives*, Springer, 2021, available at https://www.springer.com/gp/book/9783030632908, accessed 26 March 2021.
13. Axel Bruns, *The Australian Web Archive Is a Momentous Achievement – But Things Will Get Harder from Here*, The Conversation, 2019, available at http://theconversation.com/the-australian-web-archive-is-a-momentous-achievement-but-things-will-get-harder-from-here-113542, accessed 19 October 2020; George Nott, *National Library Launches 'Enormous' archive of Australia's Internet*, Computerworld, 2019, available at https://www.computerworld.com/article/3488134/national-library-launches-enormous-archive-of-australia-s-internet.html, accessed 19 July 2020.
14. National and State Libraries Australia, *Strategic Plan 2020–2023*, National and State Libraries Australia, Melbourne, 2020, p. 5.
15. Kieran Hegarty, 'Defining and Inscribing "Multicultural Library Services" in Australia: A Case Study of the Working Group on Multicultural Library Services (Victoria)', *Journal of the Australian Library and Information Association*, vol. 70, no. 1, 2021, pp. 44–59.
16. National and State Libraries Australia, *Strategic Plan 2020–2023*.
17. Joan M. Schwartz and Terry Cook, 'Archives, Records, and Power: The Making of Modern Memory', *Archival Science*, vol. 2, no. 1–2, 2002, pp. 1–19, p. 5.
18. Verne Harris, 'The archival sliver: Power, memory, and archives in South Africa', *Archival Science*, vol. 2, no. 1–2, 2002, pp. 63–86; Ann Laura Stoler, 'Colonial Archives and the Arts of Governance', *Archival Science*, vol. 2, no. 1–2, 2002, pp. 87–109.

19. Harris, 'The archival sliver', pp. 64–65.

20. Harris, 'The archival sliver', p. 84.

21. Roger Clarke, 'Morning Dew on the Web in Australia: 1992–95', *Journal of Information Technology*, vol. 28, no. 2, 2013, pp. 93–110.

22. Jodie Boyd, 'Filling the 'Gaps' in the Representation of Australia's Cultural Diversity: The Multicultural Imaginary of the NLA's Oral History Collection', *Journal of the Australian Library and Information Association*, vol. 71, no. 1, 2002, pp. 4–26.

23. Clarke, 'Morning Dew on the Web in Australia: 1992–95'.

24. Daniel Greene, *The Promise of Access: Technology, Inequality, and the Political Economy of Hope*, MIT Press, Cambridge, Massachusetts, 2021, p. 40.

25. Rose Barrowcliffe, 'Closing the Narrative Gap: Social Media as a Tool to Reconcile Institutional Archival Narratives with Indigenous Counter-Narratives', *Archives and Manuscripts*, vol. 49, no. 3, 2021, pp. 151–166; Bronwyn Carlson and Ryan Frazer, *Indigenous Digital Life: The Practice and Politics of Being Indigenous on Social Media*, Springer, Cham, Switzerland, 2021.

26. Julian Thomas, Jo Barraket, Sharon Parkinson, Chris K. Wilson, Indigo Holcombe-James, Jenny Kennedy, Kate Mannell and Abigail Brydon, *Measuring Australia's Digital Divide: The Australian Digital Inclusion Index 2021*, Report, RMIT University, Swinburne University of Technology, Telstra, October 15, 2021, available at https://apo.org.au/node/314284, accessed 9 November 2021, p. 11.

27. Alice E. Marwick and Danah Boyd, 'I Tweet Honestly, I Tweet Passionately: Twitter Users, Context Collapse, and the Imagined Audience', *New Media & Society*, vol. 13, no. 1, 2011, pp. 114–133.

28. Thomas et al., *Measuring Australia's Digital Divide*.

29. Axel Bruns, Brenda Moon, Felix Münch and Troy Sadkowsky, 'The Australian Twittersphere in 2016: Mapping the Follower/Followee Network', *Social Media + Society*, vol. 3, no. 4, 2017, pp. 1–15; Annabel Crabb, 'Facebook or YouTube? What Does Your Favourite Social Media Site Say About You?', *ABC News*, June 27, 2021, available at https://www.abc.net.au/news/2021-06-28/annabel-crabb-austra-lia-talks-social-media/100241888, accessed 23 November 2021; Yellow, *Yellow Social Media Report 2020: Consumer Statistics*, Thryv Australia, 2020, available at https://www.yellow.com.au/wp-content/uploads/sites/2/2020/07/Yellow-Social-Media-Report-2020-Consumer-Statistics.pdf, accessed 9 January 2021.

30. Crabb, 'Facebook or YouTube? What Does Your Favourite Social Media Site Say about You?'; Yellow, *Yellow Social Media Report 2020: Consumer Statistics*.

31. Carlson and Frazer, *Indigenous Digital Life*; Ariadna Matamoros-Fernández, 'Platformed Racism: The Mediation and Circulation of an Australian Race-Based Controversy on Twitter, Facebook and YouTube', *Information, Communication & Society*, vol. 20, no. 6, 2017, pp. 930–946.

32. Harris, 'The archival sliver', p. 84.

33. Milligan, *History in the Age of Abundance?*, p. 58.

34. Valérie Schafer, Francesca Musiani and Marguerite Borelli, 'Negotiating the Web of the Past: Web Archiving, Governance and STS', *French Journal For Media Research*, vol. 6, 2016, pp. 1–23, p. 12.

35. Steven M. Schneider and Kirsten A. Foot, 'The Web as an Object of Study', *New Media & Society*, vol. 6, no. 1, 2004, pp. 114–122, p. 115.

36. Terry Kuny, 'A Digital Dark Ages? Challenges in the Preservation of Electronic Information', in *63rd IFLA Council and General Conference*, International Federation of Library Associations and Institutions, Copenhagen, Denmark, August 27, 1997.

37. Viktor Mayer-Schonberger, *Delete: The Virtue of Forgetting in the Digital Age*, Princeton University Press, Princeton, N.J., 2009.

38. Jessica Ogden, '"Everything on the Internet Can Be Saved": Archive Team, Tumblr and the Cultural Significance of Web Archiving', *Internet Histories*, vol. 6, no. 1–2, 2021, pp. 113–132.

39. Janne Nielsen, *Using Web Archives in Research: An Introduction*, NetLab, Aarhus, 2016, pp. 12–19.

40. Michel Callon and Bruno Latour, 'Unscrewing the Big Leviathan: How Actors Macro-Structure Reality and How Sociologists Help them to Do So', in K Knorr and A Cicourel (eds), *Advances in Social Theory and Methodology: Toward an Integration of Micro- and Macro-Sociologies*, Routledge, London, 1981, pp. 277–303.

41. Fiona R. Cameron, *The Future of Digital Data, Heritage and Curation: In a More-Than-Human World*, Routledge, London, 2021, available at https://doi.org/10.4324/9781003149606, accessed 6 May 2021; Milligan, *History in the Age of Abundance?*; Schafer et al., 'Negotiating the Web of the Past: Web Archiving, Governance and STS'.

42. Nielsen, *Using Web Archives in Research*.

43. Niels Brügger, 'Webraries and Web Archives – The Web Between Public and Private', in Wendy Evans and David Baker (eds.), *The End of Wisdom?: The Future of Libraries in a Digital Age*, Chandos, Hull, 2016,

pp. 185–190, available at https://linkinghub.elsevier.com/retrieve/pii/B9780081001424000233, accessed 5 September 2020.

44. Jessica Ogden, Susan Halford and Leslie Carr, 'Observing Web Archives: The Case for an Ethnographic Study of Web Archiving', in *Proceedings of the 2017 ACM on Web Science Conference – WebSci '17*, ACM Press, Troy, NY, 2017, pp. 299–308, available at http://dl.acm.org/citation.cfm?doid=3091478.3091506, accessed 20 July 2020.

45. Quentin Lobbé, 'Where the Dead Blogs Are', in Milena Dobreva, Annika Hinze and Maja Žumer (eds.), *Maturity and Innovation in Digital Libraries*, Lecture Notes in Computer Science, Springer, Cham, 2018, pp. 112–123.

46. Kieran Hegarty, 'Unlocking Social Media Archives: Creative Responses to the Challenge of Access', in *Bring IT On!*, Melbourne, June 14, 2022, available at https://www.vala.org.au/vala2022-proceedings/vala2022-online-session-6-hegarty/, accessed 23 July 2022.

47. Frank McCown and Michael L. Nelson, 'What Happens When Facebook Is Gone?', in *Proceedings of the 2009 Joint International Conference on Digital Libraries – JCDL '09*, ACM Press, Austin, TX, 2009, p. 251, available at http://portal.acm.org/citation.cfm?doid=1555400.1555440, accessed 10 August 2021, p. 251.

48. Anat Ben-David, 'Counter-Archiving Facebook', *European Journal of Communication*, vol. 35, no. 3, 2020, pp. 249–264, p. 251.

49. Kieran Hegarty, 'Shhh…: What a Library's Social Character Reveals about the Logics and Politics of Source Creation', *The Sociological Review*, The Sociological Review, 2021, available at https://thesociologicalreview.org/magazine/december-2021/postgraduate-research/shhh/, accessed 17 January 2022.

50. Boyd, 'Filling the "Gaps" in the Representation of Australia's Cultural Diversity'.

51. Eira Tansey, 'No One Owes Their Trauma to Archivists, or, the Commodification of Contemporaneous Collecting', *Eira Tansey*, June 5, 2020, available at https://eiratansey.com/2020/06/05/no-one-owes-their-trauma-to-archivists-or-the-commodification-of-contemporaneous-collecting/, accessed 27 August 2021.

52. Marilys Guillemin and Lynn Gillam, 'Ethics, Reflexivity, and "Ethically Important Moments" in Research', *Qualitative Inquiry*, vol. 10, no. 2, 2004, pp. 261–280.

53. Marwick and Boyd, 'I Tweet Honestly, I Tweet Passionately'.

54. Jimmy Lin, Ian Milligan, Douglas W. Oard, Nick Ruest and Katie Shilton, 'We Could, but Should We?: Ethical Considerations for Providing Access to GeoCities and Other Historical Digital Collections', in *Proceedings of the 2020 Conference on Human Information Interaction and Retrieval*, ACM, Vancouver, BC, March 14, 2020, pp. 135–144, available at https://dl.acm.org/doi/10.1145/3343413.3377980, accessed 6 April 2022; Stine Lomborg, 'Ethical Considerations for Web Archives and Web History Research', in Niels Brügger and Ian Milligan (eds.), *The SAGE Handbook of Web History*, SAGE, London, 2019, pp. 99–111, available at http://sk.sagepub.com/reference/the-sage-handbook-of-web-history/i1130.xml, accessed 24 February 2021; Katie Mackinnon, 'Ethical Approaches to Youth Data in Historical Web Archives', *Studies in Social Justice*, vol. 15, no. 3, 2021, pp. 442–449; 'Critical Care for the Early Web: Ethical Digital Methods for Archived Youth Data', *Journal of Information, Communication and Ethics in Society*, vol. 20, no. 3, 2022, pp. 349–361.

55. Kimberly Christen, 'Does Information Really Want to be Free? Indigenous Knowledge Systems and the Question of Openness', *International Journal of Communication*, vol. 6, 2012, p. 24, p. 2874.

56. Guillemin and Gillam, 'Ethics, Reflexivity, and "Ethically Important Moments" in Research'.

57. Mackinnon, 'Critical care for the early web'.

58. Kate Crawford, *The Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*, Yale University Press, New Haven, Connecticut, 2021; Nicole Starosielski, *The Undersea Network*, Duke University Press, Durham, North Carolina, 2015.

59. Cameron, *The Future of Digital Data, Heritage and Curation*.

60. James Vincent, *This Beautiful Map Shows Everything that Powers an Amazon Echo, from Data Mines to Lakes of Lithium*, The Verge, 2018, available at https://www.theverge.com/2018/9/9/17832124/ai-artificial-intelligence-supply-chain-anatomy-of-ai-kate-crawford-interview, accessed 13 December 2021.

61. Joana Moll, *CO2GLE*, 2014, available at https://www.janavirgin.com/CO2/CO2GLE_about.html, accessed 13 December 2021.

62. Caleb Stone, *Web Acknowledgement [Chrome Extension]*, 2021, available at https://chrome.google.com/webstore/detail/web-acknowledgement/dimpmephpcddcichkcicdljjgdhjkgop, Accessed 13 December 2021.

63. Jeannette A. Bastian, 'Records, Memory and Space: Locating Archives in the Landscape', *Public History Review*, vol. 21, 2014, pp. 45–69, p. 52.

64. Richard Rogers, 'Doing Web History with the Internet Archive: Screencast Documentaries', *Internet Histories*, vol. 1, no. 1–2, 2017, pp. 160–172.

65. Ilya Kreymer, *OldWeb.today*, 2015, available at https://oldweb.today/, accessed 13 December 2021.

66. Michelle Caswell and Marika Cifor, 'From Human Rights to Feminist Ethics: Radical Empathy in the Archives', *Archivaria*, vol. 81, 2016, pp. 23–43.

67. Emily Maemura, *Towards an Infrastructural Description of Archived Web Data*, WARCnet Papers, WARCnet, Department of Media and Journlism Studies School of Communication and Culture, Aarhus University, Aarhus, 2022.

68. Ogden et al., 'Observing Web Archives'.

69. National Library of Australia, *Web Archiving Blog Posts*, Pandora Archive, 2020, available at https://pandora.nla.gov.au/pandoranews.html, accessed 31 May 2022.

70. Jessica Ogden and Emily Maemura, '"Go fish": Conceptualising the Challenges of Engaging National Web Archives for Digital Research', *International Journal of Digital Humanities*, vol. 2, no. 1, 2021, pp. 43–63.

71. Hegarty, 'Unlocking Social Media Archives: Creative Responses to the Challenge of Access'.

72. Mackinnon, 'Critical Care for the Early Web'.

73. Mackinnon, 'Critical Care for the Early Web', pp. 2, 9.