This is a repository copy of *Participatory web archiving: multifaceted challenges*.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/203201/

Version: Accepted Version

Participatory Web Archiving: Multifaceted Challenges

Cui Cui[1, 3][0000-0002-0389-1262], Stephen Pinfield[1][0000-0003-4696-764X], Andrew Cox[1][0000-0002-2587-245X], Frank Hopfgartner[2][0000-0003-0380-6088]

[1] Information School, The University of Sheffield, 211 Portobello, Sheffield
[2] Institute for Web Science and Technologies (WeST) Universität Koblenz
[3] Bodleian Libraries, University of Oxford, Broad Street, Oxford

{ccui3, s.pinfield, a.m.cox}@sheffield.ac.uk,

hopfgartner@uni-koblenz.de

Abstract. There has been increasing interest in participatory web archiving in recent years. Indeed, it is widely regarded as a necessary step in the development of web archives. From a theoretical point of view, it has been seen as driven by the "archival turn" in which the origin of archives is critically analysed and it becomes clear that established power dominates archives while marginalized voices are absent. Web archiving benefits from this "archival turn", not only in addressing limitations inherited from conventional archives, but also in challenging embedded systemic and selection biases when choosing what to archive from the Web. Through a critical literature review, this paper addresses the need to analyse participatory web archiving practices, the mechanisms and power relations within them through political theories of power and participation.

Keywords: Web Archives, Web Archiving, Participatory, Archival Selection, Archival Appraisal, Archival Turn.

1 Context of the research

Archives are sites where power is exercised and legitimised. Ovenden [1] vividly described how colonial archives were created as parts of the archives of the colonialising power, and were later subjected to deliberate elimination in order to control how history is understood. These long-standing issues in conventional archives continue in the digital world. Web archives are a new form of archival material curated through a process of selecting and preserving websites, web pages or their contents using techniques, tools and platforms with long term preservation and access strategies in place. Major efforts in preserving the Web were first initiated by the Internet Archive. National libraries and archives, government organisations and research institutions have then taken on the

role, as evident from surveys conducted from 2010 to 2017 [2–7], and the list of web archiving initiatives globally from Wikipedia[1]. Some national archives, libraries and heritage institutions have a mandate to archive the national web domain arising from legislation or to produce thematic collections of websites/pages supported by heritage preservation traditions. However, web archives have developed as a new paradigm - one that has often stood alone in terms of theoretical development. Yet web archives need to respect conventional archival principles: they are still subject to many of the same practices [8] and face similar challenges in terms of collections.

The issue of *what* to archive has long been debated in traditional archival practice [9] and the Web is no different [10]. In conventional archives, the answer tends to reflect the nature of the archival institutions concerned and the power to choose what is important and so remembered vested in government administrative and bureaucratic hierarchies. Likewise, selection bias in web archives can be just as systemic and as much the consequence of human actions.

The "archival turn", in which the archive itself rather than just its contents becomes a subject of study, arises from a number of critical positions, including post-colonial theory [11, 12]. This "turn" encourages archivists to consider the ways in which vested interests and prevailing power structures shape the collection of materials in conventional archives. Web archives collate and curate a different type of material, but also do so within a particular political, social and economic context. Institutional control over technologies and the infrastructure facilitating these technologies is forming another force that is decisive in what ends up in web archives. Essentially, the issue is that of "power": who exercises it and the tools at hand to exercise it. This is to a large extent decisive in determining what materials are archived.

In resistance to this tendency, participatory archiving, which involves working in collaboration with different communities to build archives, is increasingly practiced in conventional archives [13]. It has now become one of the main themes of web archive development too [6, 10, 14–19]. Participatory web archiving is an approach deploying established strategies in a new paradigm to tackle inherent limitations rooted in the theory and practice of conventional archiving. Through community participation, it is attempted to redistribute the power to various, networked often interdependent actors/stakeholders that include community partners, content creators and users.

Through a critical literature review, this paper attempts to identify what are key issues in the development of participatory web archiving. They include: the mechanisms involved in participatory approaches; the ways in which the impacts of the mechanism can be measured and evaluated; the nature and composition of participation partners; the extent to which they have gained the power to shape the process and how that power is redistributed; and the extent to which their participation can address the inherent limitations and reduce or mitigate embedded biases.

---

[1] https://en.wikipedia.org/wiki/List_of_Web_archiving_initiatives accessed 26/11/2022

2   Web archives as a different form of archive

Web historians and theorists have tried to untangle the complexity of the components that constitute the Web [8, 20, 21]. These components are open to different interpretations, which has fundamental implications on for what and how the Web should be archived and, in fact, what web archives actually are. The Web could be thought of as having three components: content, context and technology. Brügger [22] has argued that the archived web distinguishes itself from other document types. It is the reborn digital version of a constructed unique representation of the online web, which is essentially different from the original. Moreover, it changes during the process of collection and preserving [20, 22, 23]. The archived fragments of the Web often hold significant features of incompleteness, temporal and spatial inconsistency.[8]. It is also difficult, indeed nearly impossible, to capture every single component of a website so that the archived version can be reconstructed as exactly as the original [8]. The contextual environment of the archived individual website or webpage may be cut out from certain points.

Material in web archives therefore does not consist of the original websites/webpages, but a type of new material originated from the live Web and reconstructed with human and technology interventions. When archiving the Web, the aim cannot be either to preserve the original, or totally replicate the original. Instead, it is collecting fragments of the website, parts of content, the design of the developer and the experience of users, technologies to support the system, and contextual links. These characteristics constitute a fundamental difference between archived web materials and conventional archival materials under the care of same archival organisations. Yet web archival practices have grown out of the established practices designed for types of materials that are different. This does not point to a fundamental break from conventional archives, especially when the responsibilities of archiving the Web are in the care of conventional archives [8].

During the early modern period, the legal dimension of archives became apparent. It did so in two respects: the preservation of original legal documents; and the reinforcement and preservation of the social and legal order [24]. Archives also served as the guarantors of political and administrative continuity due to the power of archives to help preserve memory of things [24]. As evidence of legal and business transactions, as well as memories of figures and events that were deemed worthy of celebrating, or memorializing, archives in effect legitimized established power and marginalized those without power [25]. To some extent, power structures are likely to be replicated in web archives in the same way.

It is the very survival of documents that determines what archives are composed of. The context of archive development concerns why some documents were made into archives in the first place, while others were not after "appraisal", and why in some cases archives were purposely destroyed or relocated. Without this knowledge, our understanding of archives is often limited. Understanding these limitations can offer the promise of richer intellectual and emotional approaches in our engagement with the past [26]. These considerations apply to web archives, too. The major theoretical de-

velopment and shift in archival thinking over the past 20 years offers a better understanding of a much more complex social context of archives. It also provides us with a new perspective when it comes to the future development of archives in respect of both traditional and born-digital materials.

3 Developing web archives in the direction of the "archival turn"

The "archival turn" – that is, the re-evaluation of archives, their purpose, and their development by various disciplines – has exerted a significant impact on archival theories [27]. One of these archival turns, grounded in post-colonial studies, points to "a move from archives as *sources* to archives as *epistemological sites* and the *outcome of cultural practices*" [27]. It treats archives themselves as primary documents of history, and re-examines their origins and establishments [11, 12]. In essence, this view arises from the struggle to navigate histories away from the imperial metropole in favour of anticolonial nationalist movements [11].

The development of institutional archives has always been closely associated with legal rights of established power. Thus, archives do not present the point of view and experiences of the whole of society but powerful record creators [28]. The recognition of the silencing of marginalised groups and communities in archives has promoted archival theorists to think about being more inclusive as well as diverse.

It follows that archives can be understood as "dynamic sites within a spectrum of pasts, presents, and futures" rather than being static [29]. Public and historical accountability demands that archives extend the definition of society's memory to offer citizens "a sense of identity, locality, history, culture and personal and collective memory" rather than one limited solely to the documentary residue left over (or chosen) by powerful record creators [28].

In parallel, the purpose of web archiving is to preserve web material deliberately and purposively and must in some degree reflect the specific reasons, such as for research purposes, record keeping, and as evidence. This may be to serve research project, or preserve memories and a nation's culture heritage [23, 30]. Both Ben-David & Amram and Ogden et al., have shown that the knowledge on the Internet Archive Wayback Machine is generated through a series of complex, intertwined socio-technical epistemic processes. They include proactive human curation and intervention, as well as editorial decisions by archival teams, and continuing efforts to repair and maintain, eventually facilitate access [31, 32]. During this process, unavoidably, both human and technology factors introduce embedded systemic and selection bias that in turn reflect wider power structures in societies.

The embedded bias in web archives starts from the digital divide in creating and accessing web content. Unavoidably, this division will be carried over to web archives that negatively affects the representation in web archives [8]. Furthermore, systemic bias in large scale international or national broad crawls are apparent, since popular websites are likely to be found within the scope of archiving strategies, and consequently are most likely to be archived [8]. Moreover, selection bias can be even more

prominent within smaller collections as they are often curated by a library, or an archive, or other institutions, which have particular collecting priorities. As technologies play an essential part when archiving the Web, the extent of technological expertise and the establishment of mass infrastructure may centralise decision-making when it comes to the selection of web archives [8].

As the choices made to decide what to archive are embedded in a complex, unsystematic, and less transparent environment [33], web archival practices are facing even greater challenges than conventional archives when coming to select what to archive and define the scope of the archive collections. Content created by least known or smaller organisations are likely to be less visible and influential on the Web and have a smaller chance to be archived. How to identify these marginalised materials can be challenging when the searching algorisms are in place to work in a different direction; various collecting strategies need to be in place purposely to address these problems.

4   Participatory web archiving

In the context of the "archival turn", greater focus of what has been understood to be valuable or worth archiving offers us an opportunity to reassess the contextual environment around archival development. Cook [13] summarised paradigm shifts of archival theory as those from "evidence to memory", to identity, and to participatory archiving as a community. Archivists shifted their focus from "truth, evidence, authenticity, and defending the integrity of the record" to consciously "co-creating the archive", to telling stories and narratives, and (Cook envisioned the possible fourth phase of development) to "share that appraisal function with citizens, broadly defined, where we engage our expertise with theirs in a blend of coaching, mentoring, and partnering" [34]. These paradigm shifts should arguably also be influential for web archives.

Participatory practices in the cultural heritage sector are not new. They are not limited to new technologies [35]. The potential benefits for institutions include deepening relationships, enhancing collections, potential cost savings in staff resources [36], promoting engaged access [37, 38], and extending existing user groups [39]. However, they do not necessarily lead to wider representation, due to the observed "long tail" participation pattern, in which the "crowd" is dominated by a small group of active participants [39, 40].

Participatory web archiving has gained increasing attention over the past few years both on practical and theoretical levels. An interview by Geeraert with Bingham [41] touched on some practical challenges for the development of web archives collaborating with wider communities, such as the curation decisions for a website with extreme views on Covid nominated by the "crowd", setting shared collecting standards with collaborating partners, overrepresented materials in English over other languages despite the fact that the collection contains materials in 51 languages. Schafer & Winters [19] have brought participatory web archiving within a "good governance" framework to address the political role of web archives. But less is known about how participation plays its role in this framework. Here we are dealing with an old problem in a new

situation: can participatory web archiving tackle inherent limitations rooted in conventional archival theories and practices? And can it address embedded bias that is intensified by the prevailing digital and political environment? If so, how? We need to understand participation within a new paradigm. Political science has the potential to help us answer these questions.

Since it first became a prominent topic in governance in the 1960s, the focus of interest in public participation has changed from debating whether it should be applied in public-policy making to its application and evaluating its impact, practices and methods [42]. As far as application is concerned, studies of participatory development have addressed the ineffectiveness of externally imposed and expert-oriented forms of research and planning that was increasingly evident in the 1980s [43].Participatory approaches emerged as an alternative to donor-driven and outsider-led development approaches. They have been justified in term of their sustainability, relevance and empowerment, making people central to development, allowing marginalised groups to be involved in interventions that affect them, and presenting their perspectives, knowledge, priorities and skills [43].

However, participation is often situated within complex and fluid power relations. As such, its outcomes depend on many factors and are often limited by three sets of participatory development "tyrannies" identified by Cooke and Kothari [43]; "the tyranny of decision-making and control, the tyranny of the group, the tyranny of method". That is the tyranny where facilitators control over the process, when the existing dominating interests are reinforced and participatory methods become the only legitimate approaches. They raise key points on evaluating the power dimensions in participatory web archiving: actors who have the power of decision making, interests of groups they represent, and the dominance of participatory methods undermining other valid approaches and reinforcing the existing power.

Later more attention was paid to the transformative impact of participation. Cornwall [44] focused on the dynamics of power and difference within invited spaces and analysed participation as a spatial practice that emphasises the transformative possibilities of participation. Similarly, Williams [45] addressed the spatial and temporal aspects of empowerment that participation seeks to achieve. Furthermore, from the evaluation point of view, he [45] offered an alternative way of "examining the effects of participation on political capabilities: how, if at all, do specific instances of participation contribute to processes of political learning, reshape networks of power, and change patterns of political learning, reshape networks of power and change patterns of political representation?" These studies indicate we could analyse the transformation impact of participatory web archiving on representation through knowledge sharing and learning.

Furthermore, from an analytical point, Foucault's concept of "governmentality"[2], offers us a framework to address the dimensions of power, and how various forms of knowledge and theories clustered under the heading of governance may inform its exercise in political and administrative reforms [46]. This term contains a wider analytical scheme (genealogy) [46]. Within it, "the analytics of government not only concentrates

---

[2] The term "governmentality" was introduced in the lecture, given at the Collège de France in February 1978 by Foucault [48].

on the mechanisms of the legitimisation of domination or the masking of violence, be-yond that it focuses on the knowledge that is part of the practices, the systematisation and 'rationalisation' of a pragmatics of guidance"[47]. "Governmentality" thus offers a potential analytic framework for articulating these hidden power dominance created and legitimised during the process of knowledge creation and sharing in participatory web archiving. .

Although the purpose of public participation may differ in various situations, we argue for a need to focus on some key issues, which are drawn from above theoretical frameworks. Firstly, there is a need to understand how participation in web archiving influences the power structure which not only arises from its origins in conventional archiving but is also reshaped when the power is redistributed to multiple networked yet independent stakeholders in a new digital environment. Secondly, it is important to articulate the extent to which participation in web archiving can reduce or mitigate em-bedded biases during the knowledge creation process. Fundamentally, there are three main themes emerging from this analysis: power relations, knowledge creation, and representation. To collect evidence and make these relationships and power dynamics in web archiving explicit, it is important to revisit and ground the research in the foun-dation of political science and methodologies.

5 Conclusion

This paper has reviewed the relevant literature on web archives and participation. It links web archives with conventional archives, as well as considering their theoretical and practical connections with each other. It also examines ways in which the "archival turn" has shaped the direction of travel of web archiving, and how participation has gained attention in the field of web archiving. Applying existing approaches in a new paradigm, it stresses the need to ground the analysis of power relations, knowledge creation, and representation in political theories of power and participation. It highlights emerging research needs to seek ways in which power relations in web archiving, and contradictions and conflicts can be articulated and decoupled, so that opportunities of participation can be exploited to advance the practice of web archiving.

References

1. Ovenden, R.: Burning the books: A history of knowledge under attack. John Murray, London (2020).
2. Gomes, D., Miranda, J., Costa, M.: A survey on web archiving initiatives, (2011). https://doi.org/10.1007/978-3-642-24469-8_41.
3. Costa, M., Gomes, D., Silva, M.J.: The evolution of web archiving. Int. J. Digit. Libr. 18, 191–205 (2017). https://doi.org/10.1007/s00799-016-0171-9.
4. NDSA: National Digital Stewardship Alliance web archiving survey: Report produced by the NDSA content working group. (2012).
5. Bailey, J., Hanna, K., Archive, I., Hartman, C., Mccain, E., Taylor, N.: Web archiving in the United States: A 2013 survey. (2014).

6. Bailey, J., Grotke, A., McCain, E., Moffatt, C., Taylor, N.: Web archiving in the United States : An NDSA report results of a survey of organizations preserving web content. (2017). https://doi.org/10.17605/OSF.IO/R5PQK.

7. Farrell, M., McCain, E., Praetzellis, M., Thomas, G., Walker, P.: Web archiving in the United States: A 2017 Survey. (2018). https://doi.org/10.17605/OSF.IO/3QH6N 4.

8. Milligan, I.: History in the age of abundance? : How the web is transforming historical research. McGill-Queen's University Press, Montreal ; London ; (2019).

9. Cook, T.: Building an archives: Appraisal theory for architectural records. Am. Arch. 59, 136–143 (1996).

10. Leetaru, K.: Why it's so important to understand what's in our web archives-annotated, https://www.forbes.com/sites/kalevleetaru/2015/11/25/why-its-so-important-to-understand-whats-in-our-web-archives/#6c71e0f73f8c, last accessed 2021/05/15.

11. Dirks, N.: Autobiography of an archive: A scholar's passage to India. Columbia University Press, New York, NY (2015).

12. Stoler, A.: Colonial archives and the arts of governance. Arch. Sci. 2, 87–109 (2002). https://doi.org/10.1007/BF02435632.

13. Cook, T.: Evidence, memory, identity, and community: Four shifting archival paradigms. Arch. Sci. 13, 95–120 (2013). https://doi.org/10.1007/s10502-012-9180-7.

14. Brügger, N., Schroeder, R.: The Web as history. UCL Press, London (2017).

15. Summers, E., Punzalan, R.: Bots, seeds and people: Web archives as infrastructure. Proc. ACM Conf. Comput. Support. Coop. Work. CSCW. 821–834 (2017). https://doi.org/10.1145/2998181.2998345.

16. Gail, T.: Web archiving environmental scan. Harvard Library Report (2016).

17. Dougherty, M., Meyer, E.T., Madsen, C., van den Heuvel, C., Thomas, A., Wyatt, S.: Researcher engagement with web archives: State of the art. Joint Information Systems Committee Report, London (2010).

18. Bingham, N.: Harnessing the crowd: Coronavirus topical collection at the UK, https://blogs.bl.uk/webarchive/2020/04/harnessing-the-crowd-coronavirus-topical-collection-at-the-uk-web-archive.html, last accessed 2021/05/15.

19. Schafer, V., Winters, J.: The values of web archives. Int. J. Digit. Humanit. 2, 129–144 (2021). https://doi.org/10.1007/s42803-021-00037-0.

20. Brügger, N.: The archived web: Doing history in the digital age. The MIT Press, Cambridge, Massachusetts ; London, England (2018).

21. Masanès, J.: Web archiving: Issues and methods. In: Web archiving. pp. 1–54. Springer, Berlin (2006).

22. Brügger, N.: The archived website and website philology: A new type of historical document? Nord. Rev. 29, 155–175 (2008). https://doi.org/10.1515/nor-2017-0183.

23. Brügger, N., Finnemann, N.O.: The Web and digital humanities: Theoretical and methodological concerns. J. Broadcast. Electron. Media. 57, 66–80 (2013). https://doi.org/10.1080/08838151.2012.761699.

24. Friedrich, M., Dillon, J.N.: The birth of the archive: A history of knowledge. University of Michigan Press, Ann Arbor, Michigan (2018).

25. Cook, T.: What is past is prologue: A history of archival adeas since 1898, and the future paradigm shift. Archivaria. 43, 17–63 (1997).

26. Prescott, A.: The Textuality of the Archive. In: Craven, L. (ed.) What are archives? Cultural and theoretical perspectives: A reader. pp. 31–52. Ashgate, Aldershot (2012).

27. Ketelaar, E.: Archival turns and returns: Studies of the archive. In: Gililand, A.J., McKemmish, S., and Lau, A.J. (eds.) Research in the archival multiverse. pp. 228–268. , Clayton, Victoria (2017).

28. Cook, T.: Archival science and postmodernism: New formulations for old concepts. Arch. Sci. 1, 3–24 (2001). https://doi.org/10.1007/BF02435636.

29. Lee, J.A.: Beyond pillars of evidence: Exploring the shaky ground of queer/ed archives and their methodologies. In: Gililand, A.J., McKemmish, S., and Lau, A.J. (eds.) Research in the archival multiverse. , Clayton, Victoria (2017).

30. Brügger, N.: Web archiving: Between past, present, and future. In: Consalvo, M; Ess, C. (ed.) The handbook of internet studies. pp. 24–42. Wiley-Blackwell, Malden (2011).

31. Ben-David, A., Amram, A.: The Internet Archive and the socio-technical construction of historical facts. Internet Hist. 2, 179–201 (2018). https://doi.org/10.1080/24701475.2018.1455412.

32. Ogden, J., Halford, S., Carr, L.: Observing web archives: The case for an ethnographic study of web archiving. In: Proceedings of the 2017 ACM on web science conference. pp. 299–308. ACM (2017). https://doi.org/10.1145/3091478.3091506.

33. Milligan, I.: Historiography and the Web. In: Brügger, N. (ed.) The SAGE handbook of web history. SAGE Publications Ltd, London (2018).

34. Cook, T.: "We are what we keep; we keep what we are": Archival appraisal past, present and future. J. Soc. Arch. 32, 173–189 (2011). https://doi.org/10.1080/00379816.2011.619688.

35. Benoit, E., Eveleigh, A.: Defining and framing participatory archives in archival science. In: Benoit, E. and Eveleigh, A. (eds.) Participatory archives : theory and practice. pp. 1–12. Cambridge Core, London (2019).

36. Ridge, M.: Crowdsourcing our cultural heritage: Introduction. In: Ridge, M. (ed.) Crowdsourcing our cultural heritage. pp. 1–13. Farnham, Ashgate (2014).

37. Carletti, L.: Participatory heritage: Scaffolding citizen scholarship. Int. Inf. Libr. Rev. 48, 196–203 (2016). https://doi.org/10.1080/10572317.2016.1205367.

38. Owens, T.: Digital cultural heritage and the crowd. Curator Museum J. 56, 121–130 (2013). https://doi.org/10.1111/cura.12012.

39. Bonacchi, C., Bevan, A., Keinan-Schoonbaert, A., Pett, D., Wexler, J.: Participation in heritage crowdsourcing. Museum Manag. Curatorsh. 34, 166–182 (2019). https://doi.org/10.1080/09647775.2018.1559080.

40. Reinsone, S.: Searching for deeper meanings in cultural heritage crowdsourcing. In: Hetland, P., Pierroux, P., and Esborg, L. (eds.) A history of

participation in museums and archives : Traversing citizen science and citizen humanities. Routledge, London (2020).

41. Geeraert, F., Bingham, N.: Exploring special web archives collections related to COVID-19: The case of INA, an interview with Nicola Bingham (British Library) conducted by Friedel Geeraert (KBR). WARCnet. (2020).

42. Quick, K.S., Bryson, J.M.: Public participation. In: Ansell, C. and Torfing, J. (eds.) Handbook on theories of governance. pp. 158–168. Edward Elgar Publishing, Northampton (2022).

43. Cooke, B., Kothari, U.: The case for participation as tyranny. In: Cooke, B. and Kothari, U. (eds.) Participation: The new tyranny? pp. 1–15. Zed Books, London (2001).

44. Conrwall, A.: Space for transformation? Reflections on issues of power and different in participation in development. In: Hickey, S. and Mohan, G. (eds.) Participation: from tyranny to transformation?: Exploring new approaches to participation in development. Zed, London (2004).

45. Williams, G.: Towards a re-politicisation of participatory development: political capabilities and spaces of empowerment. In: Hickey, S. and Mohan, G. (eds.) Participation: from tyranny to transformation?: Exploring new approaches to participation in development. pp. 92–107. Zed, London (2004).

46. Triantafillou, P.: Governmentality. In: Ansell, C. and Torfing, J. (eds.) Handbook on theories of governance. pp. 378–388. Edward Elgar Publishing, Northampton (2022).

47. Lemke, T.: Foucault, governmentality, and critique. Rethink. Marx. 14, 49–64 (2002). https://doi.org/10.1080/089356902101242288.

48. Foucault, M.: Governmentality. In: Burchell, G., Gordon, C., and Miller, P. (eds.) The Foucault effect: Studies in governmentality, with two lectures by and an interview with Michel Foucault. pp. 87–104. University of Chicago Press, Chicago (1991).