Archives & Manuscripts
The Journal of the Australian Society of Archivists

REFLECTION ARTICLE

Documenting COVID-19 in Australia: An Interdisciplinary Perspective

Terhi Nurmikko-Fuller*

Centre for Social Research & Methods, Australian National University, Canberra, Australia

Abstract

Social media posts and unpublished student projects are just two examples of the digital content – a type of ephemeral popular culture – produced during the COVID-19 pandemic. Collecting this material would provide researchers and analysts with information that is complementary to other data used to report and capture the crisis, such as government policies and scientific documentation. But what are the long-term privacy implications of collecting this material? In this time of privacy paradoxes and the Data Economy, does the responsibility for the ethical use of this data fall onto the archivists and researchers?

Keywords: *Digital ephemera*; *Ethics*; *Privacy*; *Social media.*

In January of 2009 – some 15 years ago – three professors from the University of Southampton, UK (Halford, Pope, and Carr) uploaded a manifesto describing the World Wide Web (henceforth 'the Web') as an object worthy of study in its own right: a thing that was the result of a 'co-constitution of technology and society… with heterogeneous actors – human and [software]… with a focus on the significance of performativity, suggesting that the web is less a thing and more an unfolding, enacted practice, as people interact with HTTP to build 'the web' moment by moment'.[1] They called this research space *Web Science*. This 'deliberately multidisciplinary'[2] field has examined the ways in which the Web has, over the last 30 years, emerged as a global information repository. It is a hypermedia platform; a stage for the performance of the self;[3] it became the new town square[4]; and, *inter alia*, the spreader of deep fakes, alternative facts, and misinformation. To study the Web in this context is to recognise that technological developments are affected by social pressures and desires, but also that society too changes under the influence of technology. This article explores how digital content could be archived and (ab)used to document our collective experience of COVID-19.

A study by researchers at the Harvard Kennedy School of Government, published in December 2021, found that whilst many social media platforms claimed to explicitly prohibit

*Correspondence: Terhi Nurmikko-Fuller, Email: terhi.nurmikko-fuller@anu.edu.au

the sharing of COVID-19-related misinformation, only four (Facebook, Instagram, YouTube, and Twitter) had a COVID-19 misinformation policy. Platforms such as Twitch, Tumblr, Messenger, and WhatsApp were not found to have explicit policies that sought to limit or prevent the spreading of misinformation about COVID-19.[5] But how effective are these policies? As of 23 November 2022, Twitter was no longer enforcing the COVID-19 misleading information policy[6]; and a study by the Institute for Strategic Dialogue (who describe themselves as 'an independent, non-profit organisation dedicated to safeguarding human rights and reversing the rising tide of polarisation, extremism and disinformation worldwide') criticised Facebook for what they described as a 'failure to tackle COVID-19 disinformation'.[7] Suter, et al found that up to 25% of comments on YouTube videos contained misinformation, and that, perhaps unsurprisingly, 'fake comments receive more attention and attract more fake replies than factual comments'.[8] Instagram, owned by Facebook, adheres to the latter's strategies for 'reducing the spread of misinformation',[9] but do not appear to be trying to actively completely prevent it.

The onset of COVID-19 brought along with it a proliferation of academic research (at the time of writing, the term brings up 4,330,000 results on Google Scholar[10]; 2,327,342 hits on WorldCat[11]; ResearchGate returns 100 pages with 10 full papers per page[12]; Arxiv returns 6,886 results).[13] The Australian Data Archive ('a national service for the collection and preservation of digital research data' based at the Australian National University (ANU) in Canberra)[14] contains 25 datasets reporting on the views of Australians on a range of COVID-19-related issues, from mental health and economic stress[15] to volunteering and aged care,[16] to name just two examples. The Australian Federal Government provides access to 143 datasets via https://data.gov.au. Add to this the plethora of analysis by academics who have examined COVID-19 in Australia from a number of different perspectives, some more generic,[17] others focusing on specific aspects such as its effect on public perception of the unemployed,[18] the consumption of alcohol at the beginning of an era that became epitomised by social distancing and lockdowns,[19] or the impact of the pandemic on teachers and students at all levels.[20] With a cornucopia of scientific research data available, why focus on social media posts to document the COVID-19 experience, in Australia and globally?

It is crucial to remember that social media platforms enable the collection and recording of (at least some of) the socio-cultural ephemera of our times. The incidental data snippets that capture the mundane are an extremely valuable and important part of creating a comprehensive picture of life during the pandemic. It might seem trivial, but photos of meals published on Instagram tell us about how our diets have changed during lockdown; analysis of lengthy posts on Reddit can help highlight social phenomena ranging from common trends to topics as significant as detecting suicide ideation in individuals who are struggling with complicated grief.[21] We can question whether it is possible to digitally map the spread of previous pandemics so as to predict the spread of this one by modelling information with the option of reinterpreting the data by changing one of several known variables. Or, we could focus on the other types of information generated during lockdowns, such as behavioural data from online gaming: since data from video game strategies can be used to model real-world crime syndicate behaviour,[22] could we also use it to predict or understand socially detrimental behaviours (for example, the decision of individuals to break lockdown curfew)? All these are examples of research carried out by graduate students at the ANU. Unfortunately, to my knowledge many of their findings, completed in the context of their studies, remain unpublished.

There are other areas of investigation that we may not yet have had a chance to touch on. Could Twitter serve as the platform for recording the chorus of a million academics lamenting another lecture delivered to a wall of turned-off cameras? Perhaps not, given the acquisition of the platform by Elon Musk in 2022[23] and the subsequent exodus of many academics to

Mastodon.[24] And what of those videos on TikTok, the ones that did not go viral? Can they tell us about modern society and its values? And what can platforms that are popular beyond the anglophone world such as Sina Weibo[25] or QQ[26] tell us about the pandemic as a global phenomenon?

There is an abundance of user-generated data online. Researchers are keen to collect and analyse it, but where should we draw the line in our effort to document the pandemic? Considerations of privacy and ethics quickly come into play. Should we seek to collect and keep records of bored teenagers in lockdown sexting each other on SnapChat? Would any user want their private messages from encrypted platforms like WhatsApp recorded for posterity? Or exchanges over private email? But how does collecting this material (and the study of it by researchers of the future) differ from the work of historians reading private correspondence in archives and libraries? After all, these snippets from social media capture the world we live in, including the social phenomena of COVID-19: there is more (perhaps something more intangible, more inherently human) to the representation of the lived experience during the pandemic than data about the structures of the virus or the patterns of the spread and contaminations. Can we argue that social media posts are in some ways an insight (even if a curated one, and in some ways unrealistic one) into a global shared experience?

We know about some of the effects of 'long COVID', but there will be other longer-term effects of the pandemic that will only manifest to future generations. Is it our responsibility to provide future scholars with as comprehensive picture as possible? To document these exceptional times as comprehensively as we can? Perhaps so, but not without careful consideration of the challenges of preserving privacy and of the ethical collection, use, storage, and access to the collected data. This may be particularly poignant given phenomena such as the Privacy Paradox,[27] which stipulates that, although many may be aware of the risks of sharing data on social media (and this includes sites for the Quantified Self, such as Strava),[28] few users take concrete steps towards preserving their privacy when it comes to sharing content online. Does the onus of privacy protection rest with the researcher since the users seem unwilling to take on that responsibility? The last two decades have shown us that the platforms themselves (which have invested interests as commercial entities seeking to make a profit in the Data Economy)[29] are unlikely to step up to the plate.

Capturing the intangible culture of any place or context is undoubtedly riddled by the need to explicitly remember that any one individual's lived experience varies immensely from from person to person and not just from country to country or from one socio-economic or ethnic group to another. But moving aside for a moment from the social science context of focusing on individual people, how can we utilise the capacity of memory institutions and the GLAM (galleries, libraries, archives, and museums) sector to help document and enable research into the COVID-19? But what can our collections and our engagements with those collections tell us about COVID-19?

The pandemic might have defined the first few years of the 2020s, but it did not occur in isolation (all puns intended!). For Australia, and in the Australian Capital Territory (ACT) in particular, the year 2020 commenced with a big bang. The area suffered immense fires: in fact, the *Washington Post* declared the 2019–2020 Australian bushfire season as 'historic' and went on to state that on a global scale, the environmental impact of the fires was greater than that of COVID-19 lockdowns.[30] The National Museum of Australia (NMA) describes the fires as 'the longest and most intense experienced in our history'.[31] Then, in January 2020, the ACT experienced one of the single most severe hailstorms ever recorded in the area.[32] Following these events of fire and ice, a public health emergency was declared in March of the same year.[33]

The NMA made an effort to archive these tumultuous times: they called for members of the public to come forward with their stories to record them as part of Momentous[34]; a group called 'Fridge Door Fire Stories'[35] was established on Facebook[36]; and not only displayed a molten phonebooth from Cobargo,[37] but enabled students of the ANU to capture the phonebooth as 3D digital object[38] as part of their studies.[39] Other unpublished student projects include an animated children's book (aimed at helping young readers understand the impact and experience of the bushfires), as well as future-gazing mapping of the potential spread of the COVID-19 based on the historical records documenting the spread of pandemics of the past (particularly the so-called 'Spanish Flu'). As with the student-led examinations into social media, most of these projects tackling the challenges of capturing the lived experience of COVID-19 within the sphere of digital cultural heritage remain unpublished. It is hard to imagine that in this context the ANU would be singular example – rather, there are undoubtedly thousands of student-led and nascent investigations from hundreds of higher education institutions across Australia – and the world – experiencing the same.

During COVID-19, the slogan 'We're all in this together' became the government marketing rallying call. Whether attempting to boost public morale, compensate for social distancing, or 'mask public security',[40] it captures the notion of a shared experience, and a common benefit to be reaped from collaboration and the aggregation of resources. The same attitude could also be applied to data: aggregation and blending for benefit and synergies that provide insights that are much greater than the sum of their parts. What we need to do is take deliberate steps to capture and explicitly articulate tacit knowledge and minority (by which I mean something other than the mainstream media or the most popular social media) views and experiences. Even with the utilization of deep learning, machine learning and artificial intelligence (AI) (the tools *du jour*), we must invest in a comprehensive collection and capture of data, but also the representation of the information that we pull from that data that is derived in thoughtful ways.

As I have argued recently,[41] large-scale computational systems can and have gone horribly wrong in the past – such as the disgraceful (to say the least) case of Google Photos insouciantly racist tagging of images[42] or AI being used to generate recruitment advertisements for CEO positions that were then automatically sent exclusively to middle-aged white men (i.e., those who represented the historical model of previously successful applicants).[43] We have also seen AI reproduce bias in when used in criminal convictions of people in the USA.[44] We know that whilst technology is neither inherently good nor bad, per se, it is also clearly not neutral[45] and neither is the data, nor is the act of data collection. These are all riddled political decisions ranging from research aims to funding body decisions and beyond. The thoughtful and careful design of clear underlying data structures that represent different but equally valid knowledge perspectives are an absolute necessity if we seek to combine disparate datasets that hold complementary information.

Doing so at a vast scale at unprecedented speed is something that is technologically possible today but it is not just a matter of doing it, we need to do it well. As with any application of technological tools to a complex human problem we need to be careful not to succumb to temptation and the traps of technological determinism or lull ourselves into the belief that just because something is technological it is neutral, objective, and benign.[46] For it is none of them. We must also resist the ever-present temptation of editing and structuring the data that we collect to fit databases that we already have. Computational technologies should play a central role in our deliberate attempt to combine datasets from different places to accurately and faithfully document the COVID-19 experience. What we need is a coordinated, collaborative, shared effort and, crucially, the necessary investment into research to make this happen.

Notes on contributor

Dr Terhi Nurmikko-Fuller is a Senior Research Fellow at the Centre for Social Research & Methods at the Australian National University. Terhi's research focuses on interdisciplinary experimentation into ways digital technologies and computational methods can be used to support and diversify research in the Humanities, Arts, and Social Sciences in general, and in relation to public culture, including Web Science, and the cultural heritage sector in particular. Terhi's publications centre on topics related to Linked Data, knowledge representation, and digital libraries, but cover a range of other topics from the role of gamification and informal online environments to 3D digital models in museums. Terhi is a CI on an Australian Research Council-funded project (*Nyingarn: a Platform for Primary Sources in Australian Indigenous Languages*, led by University of Melbourne); and a member of the Territory Records Advisory Council, Australian Capital Territory Government.

Notes

1. Available at https://www.researchgate.net/publication/43652198_A_Manifesto_for_Web_Science, accessed 03 May 2023.
2. A phrase quoted from Kieran O'Hara and Wendy Hall, Four internets: Data, Geopolitics, and the Governance of Cyberspace, Oxford University Press, 2021, p. 245.
3. Erving Goffman, The Presentation of Self in Everyday Life, University of Edinburgh, Social Sciences Research Centre, Edinburgh, 1956.
4. Katrina Grant, 'Performances of Power-The Site of Public Debate', in Shirley Leitch and Paul Pickering (eds.), Rethinking Social Media and Extremism, ANU Press, Canberra, 2022, pp. 143–158.
5. Available at https://misinforeview.hks.harvard.edu/article/research-note-examining-how-various-social-media-platforms-have-responded-to-covid-19-misinformation/, accessed 07 June 2023.
6. Available at https://transparency.twitter.com/en/reports/covid19.html#2021-jul-dec, accessed 07 June 2023.
7. Available at https://www.isdglobal.org/isd-publications/ill-advice-a-case-study-in-facebooks-failure-to-tackle-covid-19-disinformation/, accessed 07 June 2023.
8. Viktor Suter, Morteza Shahrezaye, and Miriam Meckel, 'COVID-19 Induced Misinformation on YouTube: An Analysis of User Commentary', Frontiers in Political Science, vol. 4, 2022, p. 849763.
9. Available at https://help.instagram.com/1735798276553028, accessed 07 June 2023.
10. Available at https://scholar.google.com.au/, accessed 07 June 2023.
11. Available at https://www.worldcat.org/, accessed 07 June 2023.
12. Available at https://www.researchgate.net/, accessed 07 June 2023.
13. Available at https://arxiv.org, accessed 07 June 2023.
14. Available at https://ada.edu.au/, accessed 07 June 2023.
15. Nicholas Biddle, 'ANU Poll 48 (October 2021): Mental Health and Economic Stress During COVID 19', doi: 10.26193/THF1VZ, ADA Dataverse, V1, 2021, available at https://dataverse.ada.edu.au/dataset.xhtml?persistentId=doi:10.26193/THF1VZ, accessed 07 June 2023.
16. Nicholas Biddle, 'ANU Poll 50 (April 2022): Volunteering, Aged Care, Policy Priorities and Experiences with COVID-19', doi: 10.26193/AXQPSE, ADA Dataverse, V1. 2022, available at https://dataverse.ada.edu.au/dataset.xhtml?persistentId=doi:10.26193/AXQPSE, accessed 07 June 2023.
17. Dominic O'Sullivan, Mubarak Rahamathulla, and Manohar Pawar, 'The Impact and Implications of COVID-19: An Australian Perspective', The International Journal of Community and Social Development, vol. 2, no. 2, 2020, pp. 134–151. See also Jane Shakespeare-Finch, Holly Bowen-Salter, Miranda Cashin, Amalia Badawi, Ruth Wells, Simon Rosenbaum and Zachary Steel, 'COVID-19: An Australian perspective', Journal of Loss and Trauma, vol. 25, no. 8, 2020, pp. 662–672.
18. Aino Suomi, Timothy P. Schofield, and Peter Butterworth, 'Unemployment, Employability and COVID19: How the Global Socioeconomic Shock Challenged Negative Perceptions toward the Less Fortunate in the Australian Context', Frontiers in Psychology, vol. 11, 2020, p. 594837.
19. Erica Neill, Denny Meyer, Wei Lin Toh, Tamsyn Elizabeth van Rheenen, Andrea Phillipou, Eric Josiah Tan, and Susan Lee Rossell, 'Alcohol Use in Australia during the Early Days of the COVID-19 Pandemic: Initial Results from the COLLATE Project', Psychiatry and Clinical Neurosciences, vol. 74, no. 10, 2020, pp. 542–549.

20. Rachael H. Dodd, Kevin Dadaczynski, Orkan Okan, Kirsten J. McCaffery, and Kristen Pickles, 'Psychological Wellbeing and Academic Experience of University Students in Australia during COVID-19', International Journal of Environmental Research and Public Health, vol. 18, no. 3, 2021, p. 866.
21. This refers to on-going work by Xinyuan Xu, building on her prior work on social media mourning: see for example Xinyuan Xu, Terhi Nurmikko-Fuller, and Bernardo Pereira Nunes, 'Tweets, Death and Rock'n'roll: Social Media Mourning on Twitter and Sina Weibo', in Proceedings of the 10th ACM Conference on Web Science, 2018, May, pp. 297–306.
22. Robert Fleet and Terhi Nurmikko-Fuller, 'The Potential for Serious Spaceships to Make a Serious Difference', in Proceedings of the 11th ACM Conference on Web Science, 2019, June, pp. 97–104.
23. Available at https://www.nytimes.com/2022/10/27/technology/elon-musk-twitter-deal-complete.html, accessed 08 June 2023.
24. Available at https://www.insidehighered.com/news/2022/12/01/academics-twitter-disperse-wake-musk-take-over#:~:text=Fiesler%20remains%20on%20Twitter%20but,much%20higher%20than%20on%20Twitter, accessed 08 June 2023.
25. Available at https://english.sina.com/weibo/, accessed 07 June 2023.
26. Available at https://im.qq.com/index/, accessed 07 June 2023.
27. Available at https://stories.uq.edu.au/business/2020/privacy-paradox-data/index.html#:~:text=However%2C%20only%20a%20small%20percentage,studied%20about%20two%20decades%20ago, accessed 07 June 2023.
28. Available at https://www.strava.com/, accessed 07 June 2023.
29. Available at https://www.technologyreview.com/2021/11/16/1040036/capitalizing-on-the-data-economy/, accessed 08 June 2023.
30. Available at https://www.washingtonpost.com/weather/2021/07/27/australian-bushfires-smoke-climate-covid/, accessed 07 June 2023.
31. Available at https://www.nma.gov.au/explore/features/momentous, accessed 07 June 2023.
32. Available at https://knowledge.aidr.org.au/resources/hailstorm-act-january-2020/#:~:text=One%20of%20the%20most%20severe,to%20the%20inner%20southern%20suburbs, accessed 07 June 2023.
33. Available at https://www.cmtedd.act.gov.au/open_government/inform/act_government_media_releases/rachel-stephen-smith-mla-media-releases/2020/public-health-emergency-declared-for-act, accessed 07 June 2023.
34. Available at https://www.nma.gov.au/explore/features/momentous, accessed 07 June 2023.
35. The significance of the fridge is explained in this article from Canberra Weekly: Available at https://canberraweekly.com.au/famous-firies-fridge-on-display/, accessed 07 June 2023.
36. Available at https://www.facebook.com/groups/fridgedoorfirestories/, accessed 07 June 2023.
37. Available at https://www.abc.net.au/news/2022-07-03/natural-disaster-cobargo-melted-phone-booth-canberra-museum/101196906, accessed 07 June 2023.
38. Available at https://sketchfab.com/3d-models/cobargo-phonebooth-2-e0b984f0253148e58328f0007b5f10ef, accessed 07 June 2023.
39. The courses as delivered in the previous years are described in Katrina Grant and Terhi Nurmikko-Fuller, 'Students as Collaborators: A Digital Humanities and GLAM Sector Collaboration to Produce New Web-Based Content Through Student Led Projects', in MuseWeb 2021, available at https://mw21.museweb.net/paper/students-as-innovators-a-digital-humanities-and-glam-sector-collaboration-to-produce-new-web-based-content-through-student-led-projects/index.html, accessed 07 June 2023.
40. Available at https://www.afr.com/politics/pandemic-slogans-mask-public-scrutiny-20210616-p581n6, accessed 07 June 2023.
41. Terhi Nurmikko-Fuller, Linked Open Data for Digital Humanities, Routledge, 2023.
42. Available at https://www.theverge.com/2018/1/12/16882408/google-racist-gorillas-photo-recognition-algorithm-ai, accessed 27 May 2022.
43. Available at https://www.independent.co.uk/life-style/gadgets-and-tech/news/google-s-algorithm-shows-prestigious-job-ads-men-not-women-10372166.html, accessed 25 February 2021.
44. Available at https://www.technologyreview.com/2019/01/21/137783/algorithms-criminal-justice-ai/, accessed 25 February 2021.
45. Melvin Kranzberg, 'Technology and History: "Kranzberg's Laws"', Technology and Culture, vol. 27, no. 3, 1986, pp. 544–560.
46. Terhi Nurmikko-Fuller and Paul Pickering, 'Crisis, what crisis?', in Shirley Leitch and Paul Pickering (eds.), Rethinking Social Media and Extremism, ANU Press, Canberra, 2022, pp. 159–178.