

---

# Making social media archives: current trends and archiving practices in the development of (representative) social media collections.

Beatrice Cannelli\*<sup>1</sup>

<sup>1</sup>School of Advanced Study, University of London [London] (SAS) – University of London, Senate House, Malet Street, London, WC1E 7HU, United Kingdom

## Abstract

In the last couple of decades, social media has rapidly become one of the main ways for people to interact online and participate to global, ongoing public discussions about contemporary events. Social media has defined and continues to define the 21st century, becoming an essential source of information for the interpretation and contextualisation of present times. Recent political and health crisis have made even more clear the central role social media has been playing in facilitating the diffusion of official critical information, reporting from war zones or exposing human rights abuses (Stewart, 2019; Wong et al., 2020). Furthermore, content created and shared on social media constitutes not only an essential primary source for historical research related to contemporary events, but also a means to reflect, remember and build upon the past, contributing to the making of our collective memory (Fondren & Menard McCune, 2018; Henninger & Scifleet, 2016; Winters, 2017).

The cultural and historical value of social media content has been recognised by a wide range of cultural heritage institutions worldwide which are more and more seeking to preserve the material published on these platforms in order to ensure its safeguard and accessibility in the long term (Bingham et al., 2020; Bingham & Byrne, 2021; Fondren & Menard McCune, 2018; Mengzhen, 2022; Storrar, 2014). Although they are still quite few in number, the amount of memory institutions that are or are planning to start archiving social media is steadily increasing. As web archiving practices and *ad-hoc* workflows keep being developed following the evolution of social media platforms, the content comprised in such collections continues to be shaped by a wide array of legal, technical, and ethical challenges, also generating discussions among curators about establishing strategies to ensure representativeness of social media collections.

This presentation will first offer a brief overview of the issues surrounding representation in the social media archiving panorama based on the geographical location of social media archiving initiatives, highlighting under-represented areas and new trends in the collection of content from social platforms. Moreover, drawing from interviews conducted with web archivists and curators worldwide as part of a wider, on-going PhD research investigating the challenges and opportunities related to the development of social media archives, this paper will offer a glimpse on new, open questions and practices related to the challenging task of selecting material adopted by some of the interviewed memory institutions that are (or plan to be) involved in the important and complex endeavour of preserving content from

---

\*Speaker

social media platforms. Finally, the paper will shed light on solutions adopted by national memory institutions to mitigate issues related to the representativeness of collections and ethical concerns, providing a set of information that could help better understand how social media collections are shaped, their intrinsic biases, and how to critically approach such collections for research purposes.

## References

Bingham, N., Byrne, H., Lelkes-Rarugal, C., & Rossi, G. C. (2020). *Using Webrecorder to archive UK political party leaders' social media after the UK General Election 2019-UK Web Archive blog*. <https://blogs.bl.uk/webarchive/2020/05/using-webrecorder-to-archive-uk-political-party-leaders-social-media-after-the-uk-general-election-2.html>

Bingham, N. J., & Byrne, H. (2021). Archival strategies for contemporary collecting in a world of big data: Challenges and opportunities with curating the UK web archive. *Big Data & Society*, 8(1), 2053951721990409. <https://doi.org/10.1177/2053951721990409>

Fondren, E., & Menard McCune, M. (2018). Archiving and Preserving Social Media at the Library of Congress: Institutional and Cultural Challenges to Build a Twitter Archive. *Preservation, Digital Technology & Culture*, 47(2), 33–44. <https://doi.org/10.1515/pdtc-2018-0011>

Henninger, M., & Scifleet, P. (2016). How are the new documents of social networks shaping our cultural memory. *Journal of Documentation*, 72(2). <https://www.emeraldinsight.com/doi/abs/10.1108/JD-06-2015-0069>

Mengzhen, W. (2022). *China's national library to archive 200 billion Weibo posts-CGTN*. CGTN News. <https://web.archive.org/web/20220725143541/https://news.cgtn.com/news/3d3d674d79677a4e3445>

Stewart, E. (2019, August 20). *How China used Facebook, Twitter, and YouTube to spread disinformation about the Hong Kong protests*. Vox. <https://www.vox.com/recode/2019/8/20/20813660/china-facebook-twitter-hong-kong-protests-social-media>

Storarr, T. (2014). Archiving social media (Text). *The National Archives Blog*. <https://blog.nationalarchives.gov.uk/social-media/>

Winters, J. (2017). *Will history survive the digital age?* HistoryExtra. Available at: <https://www.historyextra.com/period/20th-century/will-history-survive-the-digital-age/>

Wong, A., Ho, S., Olusanya, O., Antonini, M. V., & Lyness, D. (2020). The use of social media and online communications in times of pandemic COVID-19. *Journal of the Intensive Care Society*, 1751143720966280. <https://doi.org/10.1177/1751143720966280>

**Keywords:** Social media archiving practices, Representativeness of collections