

Qualitative Data Archives

By Kirsty Patrick¹, Jessica Scantlebury², Robert Meckin^{3,4}
and Mark Elliot^{3,4}

This Methods Futures Briefing focuses on changes to qualitative data archives. It first outlines what qualitative data archives are using the UK's Mass Observation Archive and its COVID-19 Collection as an exemplar. The advent of various technologies and social changes may impact the meaning and status of qualitative data archives, and this briefing highlights some social science and humanities (SSH) methodological concerns that may arise due to these changes.

What are Qualitative Data Archives?

Qualitative data archives are secure storage facilities that contain various forms of qualitative data. This data can exist in diverse forms, including physical formats such as handwritten and printed documents or photographs, audiovisual material, and digital formats such as email, digital documents and digital photograph or film.

The UK's **Mass Observation Archive** is one such archive. First launched in 1937 as an '*anthropology at home . . . a science of ourselves*' (New Statesman, 1937), it sought to document everyday life in Britain, using a range of methods including observations, surveys and diaries. Capturing people's behaviour, thoughts and feelings, its first incarnation ran until the 1950s. Today the Archive is a national charity in the care of the University of Sussex, based at [The Keep](#). It continues to generate qualitative data through its national panel of volunteer contributors and annual open calls. In 2020, the Archive sought to document the unfolding events of COVID-19 resulting in a collection of nearly 10,000 documents. This was wide-ranging in scale and format, and beyond the staff teams expectations (Patrick & Scantlebury, 2025).

Since 1981, the Archive has cultivated a dedicated panel of 600 self-selecting volunteers, known as 'Mass Observers.' These individuals regularly respond to open-ended questionnaires—called 'Directives'—that

invite them to reflect on personal, political, and social topics. The prompts range from everyday curiosities like '*What's on your bedside table?*' to broader societal questions such as '*How do you feel about the upcoming election?*' or '*How do you use technology at home?*'

During the COVID-19 pandemic, directives explored themes like '*Time*,' '*Vaccines*,' and '*Testing*,' alongside topics such as '*Recycling*' and '*Black Lives Matter*.'

As a result, the Archive has amassed an extensive and diverse collection of historical documentation, spanning the periods before, during, and after World War II, as well as contemporary contributions. While its core material is narrative and longitudinal, thanks to long-standing Mass Observers, it also includes photographs, illustrations, and artwork, in formats ranging from handwritten and printed to digital.

Contributors anonymously share what matters most to them in real time, offering invaluable insights into social change. The Archive also strives to amplify voices that are often overlooked by working closely with grassroots organisations.

This means that qualitative data archives can contain a wide variety of heterogeneous data, often collected over varying times, demographics, and sometimes in response to notable events or issues.

Contributing to an Archive

There are two ways in which people contribute to the Mass Observation Archive today: participating in the national **panel** and responding to annual public **open**

¹ Mass Observation Archive. Contact author: K.Patrick@sussex.ac.uk

² Mass Observation Archive. Contact author: j.c.scantlebury@sussex.ac.uk

³ University of Manchester

⁴ National Centre for Research Methods

calls. People participate for a variety of reasons. For some they are leaving a legacy “*for future Historians*” and enjoy writing. Others see value in being part of research and the opportunity to have their opinions heard. Being a Mass Observer on the national panel involves responding to the regular ‘Directive’ questionnaires in spring, summer and autumn, each containing 2-3 themes. A lesser commitment is Mass Observation’s annual call for day diaries written on 12th May in which anyone can participate both in narrative and creative formats.

Changes in Qualitative Data Archiving and Access

A range of socio-technical changes has the potential to significantly influence the practices and meanings of qualitative data archiving.

The ways that people contribute to qualitative archives are diversifying. Submitting material to an archive can be done through conventional postal services, generating handwritten submissions on different types of paper such as notepaper or letter paper and creative formats including scrapbooks, illustrations, paintings and textiles. Alternatively, responses can be submitted by email or through our online submission form in an electronic format chosen by the contributor. This means that archives need to be able to **handle and securely store** an ever-increasing array of physical and digital submissions.

Mass Observation has remained attentive to the evolving needs of researchers. For those seeking more comprehensive demographic data, an optional biographical form was introduced to capture ethnicity, disability, sexual orientation, and religion. However, it is crucial to **preserve the unique appeal that drives contributions**—namely, the anonymity, the freedom to respond voluntarily and without deadlines, and the distinction from conventional market research. Striking a balance between contributors' preferences and researchers' requirements is essential to maintaining the integrity of Mass Observation’s work and the continuation of the Mass Observation Project.

Historically, researchers would need to visit the Archive in person to view material. The rise in digital submissions has transformed our archival processes, enabling remote access to responses. To ensure both preservation and accessibility, the Archive will digitally preserve the original format while also providing an

access copy for research use in PDF via secure cloud storage.

Qualitative data archives must remain vigilant about the various ways **individuals may unintentionally disclose personal information** in electronic submissions. For instance, contributions may contain sensitive details that identify the contributor or others, sometimes due to hidden metadata within digital files. To safeguard privacy, Mass Observation redacts this data when creating access copies, intended for research uses, of the responses

The shift towards digital access alters the **relationships** of users, the archive and its staff, as reduced levels of personal interaction raise questions of care, particularly when users are accessing troubling or sensitive material. This was particularly relevant when the Archive collected material during the COVID-19 pandemic. In its duty of care, Mass Observation amended its website to include contact details of support services. Additionally, users receive guidance on vicarious trauma when accessing materials remotely to offer appropriate support.

The **scale** of the Archive, both in terms of how the data grows over time, means the size of the overall corpus of qualitative data is **attracting new and different disciplines**. Researchers analyse Mass Observation data thematically (Walsh 2023), and longitudinally (Lindsey, 2022; Lindsey and Bulloch 2014). More recently, Robinson et al (2023) used a distant reading method to analyse around 7500 diaries. For the authors ‘[computational linguistic techniques] are a *beneficial way of identifying trends in broader patterns of thoughts and behaviours which reflect lived-experiences that are particularly prominent of a given dataset*’. The COVID-19 [database](#) enables users to search the narratives by keyword and free text but also export the data to be used in such ways.

The **digitisation** of Mass Observation material by [AM Digital](#) transformed access to the collection, making it more widely available, especially to international audiences. Through Mass Observation Online and Mass Observation Project Online, users at HE institutions can explore digitised archive material enhanced with interactive maps and essays, creating an accessible research experience. These resources use Quartex technology, which further enhances the discoverability of the documents through keyword searching. The publication of this resource generates vital royalties that sustain the Archive, supporting staff and operational costs.

Future

The **representativeness** of Mass Observation has been widely debated, but how does technology and social change affect this? In a landscape of **social media and blogging**, Mass Observation continues to provide a space for people to contribute anonymously and be part of a collection of voices stored for both current research and perpetuity. Furthermore, creative engagement activities through open calls and grassroots partnerships enable the generation of valuable data, with the voices of marginalised groups.

Despite the popularity of social media, Mass Observation continues to thrive, drawing new contributors and maintaining a panel of around 600 writers. While many life loggers are attracted to different and emerging technology devices, Mass Observation does not actively seek audiovisual content, due to the challenges of maintaining anonymity in these formats and digital preservation resources. Instead, the archive primarily relies on specific Directive responses and day diary calls rather than daily documentation, and thus records **deliberate and reflective responses** from a range of contributors.

The use of Instagram and LinkedIn has enabled Mass Observation to reach new and wider audiences. **Active engagement** with 'posts' and 'reels' supports Mass Observations' national reach, with 'behind-the-scenes' of the archive activities. This changes the interactions of the archive with audiences and may have implications for how data are generated in the future.

Like many institutions, the Mass Observation Archive is exploring the potential of AI to enhance its processes and support research. Staff are currently investigating how generative AI-powered tools can assist in cataloguing archival materials and improving collection discoverability. Early explorations suggest that these technologies could significantly **accelerate cataloguing** efforts, enabling the creation of detailed records in formats aligned with open-standard metadata vocabularies such as the Dublin Core. However, Mass Observation staff will continue to review both the catalogues and qualitative narratives to safeguard anonymity, implement necessary redactions, and uphold the trust-based relationship with contributors.

Digitised resources offer users seamless **remote access** to valuable collections through intuitive, user-friendly platforms. Advanced search tools provide efficient navigation, helping researchers explore vast amounts of qualitative data with ease. For instance,

users can generate samples of diarists or Mass Observers based on biographical metadata, then conduct keyword or free-text searches. Additionally, the ability to export 'big qualitative data' into discipline-specific software, such as computational linguistics tools, enhances analytical possibilities. These raise some of the care issues mentioned earlier.

However, many curated resources accessed by researchers are purchased by HE institutions, many of which face financial challenges. Meanwhile, open-access resources are steadily increasing, driven in part by funding requirements. As AI continues to shape the research landscape, the ways in which users discover and interact with digitised resources are also evolving, requiring ongoing adaptation and innovation

Mass Observation was founded with the aim of being '*an anthropology of ourselves*' (*First Year's Work 1937-1938*). In a time of technical and social change it is important to not lose sight of why people contribute to such archives and ensure we can continue to provide a range of access digitally, electronically and in-person to meet the needs of different disciplines and those outside of academia their function as a repository of social thought, reaction and expression.

References

Lindsey, R. (2022) 'Using Mass Observation as a source of qualitative secondary data for interdisciplinary longitudinal research on voluntary action', in Dean, J. and Hogg, E. (eds.) *Researching Voluntary Action: Innovations and Challenges*. Bristol: Policy Press, pp. 84–95.

Lindsey, R. and Bulloch, S.L. (2014) 'A sociologist's field notes to the Mass Observation Archive: a consideration of the challenges of "re-using" Mass Observation data in longitudinal mixed-methods study', *Sociological Research Online*, 19(3), pp. 1–12. doi:10.5153/sro.3362.

New Statesman (1937) 'Anthropology at home: a science of ourselves', *New Statesman*, 30 January.

Mass-Observation (1938) *First Year's Work, 1937–1938*

Walsh, K. (2023) 'Night-time bedroom soundscapes: embodied geographies of housing and home', *Social & Cultural Geography*, 25(7), pp. 1064–1083. <https://doi.org/10.1080/14649365.2023.2245800>

National Centre for Research Methods
Social Sciences
University of Southampton
Southampton, SO17 1BJ
United Kingdom.

Web <http://www.ncrm.ac.uk>
Email info@ncrm.ac.uk
Tel +44 23 8059 4539
X @NCRMUK