



For Augustinian Archival Openness and Laggardly Sharing: Trustworthy Archiving and Sharing of Social Science Data From Identifiable Human Subjects

David Zeitlyn*

Institute of Social and Cultural Anthropology, School of Anthropology and Museum Ethnography University of Oxford, Oxford, United Kingdom

Keywords: human subjects, research ethics, embargo, GDPR, archives/ archiving, pseudonymisation, conflicting responsibilities

INTRODUCTION

Saint Augustine used to pray “Lord, make me chaste—but not yet.” This finds a parallel in qualitative social science where researchers may endorse data sharing and the calls for open science but with a hesitation that results from conflicting desires and obligations. How can one be open and share research material while respecting moral, ethical and legal requirements not to share? Many of the conflicts can be resolved by adding delay into the process: summarized in an Augustinian inspired archival prayer “to make me open—but not yet.” There is warrant for this but existing infrastructures may not easily accommodate it not so much because of technical problems but more because they are not resourced to plan over a long enough timescale.

Social science data involves other sorts of trust than those implicit in the call for contributions to the Frontiers Topic “Trust and Infrastructure in Scholarly Communications” (https://www.frontiersin.org/research-topics/18191/trust-and-infrastructure-in-scholarly-communications?utm_source=F-RTM&utm_medium=CFP_E2&utm_campaign=PRD_CFP_T1_RT-TITLE#overview) and in statements promoting the idea of open data, for example, by the Open Data Institute RRID:SCR_021681 (<https://theodi.org/about-the-odi/>). These are relevant to the design, implementation and funding of infrastructures supporting research beyond medical, chemical and physical disciplines. My areas of concern are qualitative social sciences as well as many humanities disciplines including oral history and life history research where, as a referee points out, “the tellers want their stories told” or the participants contribute because they want to contribute to the general development of science (Kuula, 2011). What is distinctive about these disciplines is that as an essential part of the research process, the researchers have human, interpersonal relationships with their interlocutors (which makes terms like informants, interviewees, research subjects or collaborators misleading if not inappropriate).¹ Also much of the importance of the material gathered lies precisely in the aspects that are destroyed by anonymization. Writing as an anthropologist I talk about anthropology but the discussion is relevant to all those undertaking qualitative research with living humans, especially those where more or less structured interviews are not the main form of data collection.²

¹This has led one pseudonymous author to suggest they should be coauthors. See Luther Blissett mss.

²It may even apply to those in humanities discussing the work of living authors: I suspect few in disciplines such as English etc have considered the relevance of GDPR and data sharing for their research.

OPEN ACCESS

Edited by:

Daniel. W Hook,
Digital Science, United Kingdom

Reviewed by:

Libby Bishop,
GESIS Leibniz Institute for the Social
Sciences, Germany

*Correspondence:

David Zeitlyn
david.zeitlyn@anthro.ox.ac.uk

Specialty section:

This article was submitted to
Scholarly Communication,
a section of the journal
Frontiers in Research Metrics and
Analytics

Received: 05 July 2021

Accepted: 13 September 2021

Published: 14 October 2021

Citation:

Zeitlyn D (2021) For Augustinian
Archival Openness and Laggardly
Sharing: Trustworthy Archiving and
Sharing of Social Science Data From
Identifiable Human Subjects.
Front. Res. Metr. Anal. 6:736568.
doi: 10.3389/frma.2021.736568

Trust type	Gloss	Infrastructural implications
Trust ₁ (T ₁)	Participants trust researcher	Long Term Embargos (census style)
Trust ₂ (T ₂)	Researchers trust participants	Readers can access data (data sharing)
Trust ₃ (T ₃)	Readers trust researcher	Readers can access data (data sharing)

I think it helps to distinguish different types of trust since they need different sorts of archiving infrastructure. Discussions of data sharing address mainly the third of these.

As institutionalized by research ethics boards when a researcher starts talking to an “informant” a first step is to work through a consent process. As has been long pointed out (Neale and Bishop, 2012; Bell, 2014), for qualitative research subjects such as anthropology this is often inadequate and misleading but it continues, driven by forms of institutional inertia and what Wynn and Israel (2018) describe as a *fetishization* of signed consent forms. Commonly, researchers promise to respect the confidentiality of what they have been told by anonymizing the individuals concerned. Until relatively recently this was taken (usually tacitly) to mean anonymizing *in publications* (strictly this is pseudonymisation since versions with identifiers were retained). The GDPR in the EU (and in UK) and its implementation by administrators in research ethics boards has raised the possibility of its application to the actual data collected on which publications are based (see Yuill, 2018) despite its explicit provisions for data archiving. For many researchers the “research material” (a more generous term than data) may be scrappy, poorly structured and may seem inadequate to others without the advantage of having taken part in the research process. For the original researcher they may be more *aide memoire* than hard data but not necessarily the worse for this. Sharing such material may not help assessment of publications based on it. With apologies for a double negative, this is *not* to say they are untrustworthy. Between researcher and reader there is an “ethnographic pact” (paralleling Philippe Lejeune’s “autobiographical pact” Lejeune, 1989).³ For all that sharing such material may have other benefits. It makes the material available for use by other researchers from other disciplines, and to those from the places where it may have originated. Usefulness/comprehensibility is not all or nothing property. Unquestionably the original researcher has hugely privileged access but that does not mean no one else can make any sense at all of fieldnotes etc (another intentional double negative). And in terms of *trust* a refusal to share raises the question of what they might be hiding while an openness to sharing implies of itself an degree of trustworthiness, even if the sharing may be far in the future. In other words, many of the reservations researchers may have about data sharing can be addressed by making very long embargoes the norm.

In publications from the harder sciences, issues of trust revolve around whether the published results and the text discussing them are trustworthy (T₃). To help assess this access to the underlying data is helpful. However, other issues of trust accrue in qualitative

social sciences before similar questions about the trustworthiness of publications can be asked. Usually tacit or unspoken, there are different questions of trust underlying consent giving and subsequent interactions: the respondents or interlocutors have to assess the trustworthiness of the researchers (T₁): can these individuals and can the promises they make be trusted? These are different points to those about the trustworthiness of publications. This is fundamental since, of course, if participants really don’t trust the researchers then they will not continue and no material will be collected so no publications will result. If they have doubts they may still participate but may provide more or less unhelpful or misleading answers (Kuula, 2011:15). In some cases researchers acting in good faith have made promises they have not been able to keep. The clearest and most notorious instance of this is case of an oral history project about the “troubles” in Northern Ireland which was archived in Boston (United States). The researchers sought to resolve the issue of informant confidentiality by promising former participants in the conflict that their interview data would not be released until after the interviewee’s death. But the Northern Irish police used legal subpoenas to break these promises. The researchers had made promises in good faith that in the end legal process meant they were not able (were not allowed) to keep.⁴

In cases where consent is withheld, the potential participant has made a judgment that, for whatever reasons, the researcher cannot be trusted (T₁), and no data is collected. Conversely, once someone has agreed to participate, the researcher has to make judgments about them: is their account to be trusted (T₂)? This may change over time. A claim to have special access and insight may be initially accepted but then later changed as it becomes clear that the person in question does not have access and indeed may be widely regarded by others in the community as a notorious liar. Anthropologists have to manage being simultaneously credulous fools and ultimate cynics. I call this professional bad faith or adopting a position of ironic detachment; Johannes Fabian talks about “the duplicity without which ethnographic research would be impossible—a duplicity which makes us cross borders but not without establishing a record that lets us return to our professional roles and habits” (2008: 6).⁵ That said there is a lot that can be learned from liars such as what is deemed to be a plausible alternative version of events and motivations. Disputes in politics (and elsewhere) are revealing about process no matter what

³Pursuing this is in danger of taking me off topic so I will not further discuss wider purposes and limits of data sharing here.

⁴See Lowman and Palys (2013) as well as news coverage such as <https://www.bbc.co.uk/news/uk-northern-ireland-27238797> and also <https://www.irishtimes.com/news/crime-and-law/attempt-to-access-former-ira-man-s-boston-college-tapes-replete-with-errors-court-told-1.3357750> Both accessed 26 Sept 2018.

⁵The ambiguities involved have also been discussed as the part of the dilemmas of ‘insider/outside’ positionality (Zavella 1996).

actually did occur or who did what to whom. Such material even if considered to be unreliable (T_2) may yet be revealing hence useful and so is worth keeping.

When it comes to data sharing a researcher might well want to anonymize versions (I stress *versions*: not everything) of research material so it can be made accessible relatively soon after the data has been collected—paralleling the ways that the UK Census Office removes personal ID information from some material from recent censuses. Importantly, even though the publicly accessible version may be anonymized, if the key to the anonymization exists, then legally it remains personal data and subject to GDPR rules (and strictly should be called a pseudonymized version).⁶ I very much hope researchers retain the keys because the *longue-durée* view is, in my opinion, that the full data should be retained and made available but only in the long term. Perhaps, as with census data, this should default to being 100 years after collection. Even this, in some circumstances, may have to be restricted but for most material it is plausible to suggest that access could be enabled with only a very light administrative touch.⁷ The data may still be described as FAIR (meeting principles of findability, accessibility, interoperability, and reusability) only a delay has been introduced into the aspects of accessibility and reusability.

An example of why it is important to retain complete data records including the names may be found in the title of a book by Michel Foucault based on research in judicial archives.

I, Pierre Rivière, having killed my mother, sister and brother (1835) Published 1973 (filmed in 1976).

I think similar books and films should be possible in 100 years' time with the full names given rather than:

I, Homme03061835, having killed my mother, sister and brother.

In other words I am arguing that any anonymization should be reversible. As already mentioned technically this is pseudonymization. Either a complete version of the material must be kept without pseudonyms, or a key retained to the pseudonyms used so that at a much later date the pseudonymised file can be re-edited to reinsert the actual names and other identifying information. (I suspect that in practical terms it is easier to keep two versions of the file but data managers may take a different view).⁸

ETHICAL CONTRADICTIONS

There are contradictions and conundrums in the ethical position of archiving. In related work Kirsten Bell (Bell, 2018 and Bell and

Wynn, 2020) considers conflicting relationships and stances towards consent and participation. She discusses how researchers and those they work with can develop a “procedural ethics framework.” Neale and Bishop (2012) describe an example of such a framework for research archives. This addresses the conundrum that prior “informed” consent cannot be given for future research by unknown others asking unknown research questions.⁹ Bishop has also discussed other ethical concerns about qualitative data archiving and shown how these can be addressed, hence enabling data archiving and sharing (2009). In the short and medium terms archivists must act as trusted proxies,¹⁰ implementing the processes that the participants consented to, and serve as trusted gate-keepers to the data long into the future (with all the administrative and resource implications that go with this). Russell and Barley (2019) raise questions about “who owns research data” which has a more direct bearing on data archiving.

Consider the possibly conflicting, certainly different, responsibilities that researchers have towards:

- Informants
- Colleagues
- Informants at a later date and their descendants
- Future colleagues (including older versions of themselves)
- Research funders/institutions

These responsibilities and obligations point in different, mutually conflicting, directions. Respect for the privacy of individuals suggests anonymization, closure or not archiving (and data destruction), whereas respect for the descendants of those individuals in the distant future suggests openness and archiving for the long term.

To make concrete the importance of preserving data such as photographs without blurring or other forms of anonymization let me give a clear example of why it would be unethical to destroy or fully anonymise data. Consider the following field photograph that I took in May 1986 (Figure 1).

Yanele Blandine, the girl in the white headscarf, is long dead. Her son, Serge Donat, had no photograph of his mother until recently, when I was able to send him a copy. Imagine his response if I had said either:

I took a photograph of your mother long ago but I cannot send you a copy because I destroyed it on completing my doctorate.

Or

I took a photograph of your mother long ago but I cannot send you a copy because I do not have her permission to share it.

Or

⁶Note that if truly anonymized and there is no key to reverse the process then they are no longer personal data and therefore not subject to GDPR regulations.

⁷Users may still have to register and the archives may wish their role to be acknowledged in any resulting publications.

⁸A Frontiers editor has suggested that cryptographic protocols or infrastructures such as the Blockchain could provide mechanisms for deterministically reversible pseudonymisation (i.e. reversible at a pre-destined point in time based on either computational difficulty to crack anonymity or on a smart contract). I find this an intriguing suggestion but fear it may involve too complex a bet on future infrastructures to achieve wide take up. Keeping two copies, only one of which is pseudonymised seems simpler.

⁹An example might be a linguist studying the phonetics of vowel quality in interview sound recordings. The linguist might be interested because the interviewer had recorded where participant and their grand-parents lived but be entirely uninterested in the content of the discussion that was recorded.

¹⁰Hence the importance of certification schemes for repositories such as Core Trust Seal (RRID:SCR_021679 <https://www.coretrustseal.org/>) by which they can be assessed as ‘Trusted Digital Repositories’.



FIGURE 1 | Caption: L-R: Barmi (alive 2018), Nde Donat (d 1987?), Ndignoua Salomon, Suzana Thia (alive 2018), Ngon Luise (died), Blandine (died), Dissi (Mougna's child, died) Jacqueline (alive 2018), the two children in the front: the boy is Kounaka Fidèle (Salomon's junior brother (alive 2018), and the girl is Mbitti Josephine (alive 2018). The names were provided in 2018 by Serge Donat, son of Blandine (photo David Zeitlyn Reference: 24_34. jpg 01/05/1986).

I took a photograph of your mother long ago but I can only send you a copy with her face blurred because I do not have her permission to share it.

I would suggest that all of these possible responses as suggested by the ethos of many ethics panels would be inhumanly rude and indeed would be deeply unethical.

There is a further almost banal point to be made—unnecessary for most of this readership but important nonetheless. Openness and Archiving does *not* mean:

- 1) giving free access to all comers
- 2) putting all material online without regulation (even for digital archives)

The UK Data Archive (RRID:SCR_014708 <https://www.data-archive.ac.uk>) is a digital repository for social science data. It includes quite a lot of anthropological material, and other qualitative data. It has developed ways of working with the sorts of unstructured material that anthropologists tend to produce. Access is not open in the sense of being uncontrolled: users have to register and in some cases *they* have to give a form of consent,¹¹ so they enter into the same sort of relationships of trust with the informants that the original researcher did. Moreover, material placed into the archive

¹¹Those reusing the data have to sign a User Agreement from the archive with binding terms and conditions that define what they can, and cannot, do with the data they access. This has parallels with the original consent process.

can in some circumstances be embargoed (usually only for a relatively few years). However, the UK Data Archive has also worked with the UK Census to develop protocols (and technical solutions) for “Secure Access Points” which enable remote access to highly sensitive datasets in constrained and controlled fashion. To summarize: Digital Archives can have a wide range of different types of access—ways that have more resemblance to access to physical archives than most digital evangelists might imagine. My suggestion is that a default for qualitative data might be 100 years, then in some unusual cases longer embargoes could still be called for, in many others much shorter ones. Even within the embargo period tightly controlled access could be arranged (again as is currently possible for census data). This is to suggest that what could be called “archival hesitancy,” a suspicion about data archiving (sharing) among qualitative social scientists, can be, partly, addressed by shifting to a census-style default from which one could discuss variations. Such an environment would only be possible if research funders allow data to be archived under census-type protocols *and* if archiving services such as the UK Data Archive are given the resources to enable this.

In short anthropology and its fellow subjects can manage to resolve the injunctions of morality and ethical responsibility to our research collaborators with the conflicting pressures to be open and share our data by resolving to be open in ways that parallel how the census is open. A recent example of how openness can be achieved over a century or more is the publication of Alfred Haddon's diaries (Herle and Philp 2021). This was undertaken in collaboration with the descendants of the people Haddon worked with in the Torres Straits in 1888 and 1898. This is the timescale we need to start

thinking about, this is the timescale that openness in anthropology and cognate disciplines can accept as ethically consistent with the promises we make. The challenge now is to resource infrastructures that can accommodate ethical complexity: to enable us to be open but not quite yet (Parry and Mauthner, 2004; Bishop, 2009; Blisset, unpublished¹²).

ETHICS STATEMENT

Written informed consent was obtained from the individuals and/or minors' legal guardian/next of kin for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and has approved it for publication.

REFERENCES

- Bell, K. (2014). Resisting Commensurability: Against Informed Consent as an Anthropological Virtue. *Am. Anthropologist* 116 (3), a–n. doi:10.1111/aman.12122
- Bell, K. (2018). The 'problem' of Undesigned Relationality: Ethnographic Fieldwork, Dual Roles and Research Ethics. *Ethnography* 20 (1), 8–26. doi:10.1177/1466138118807236
- Bell, K., and Wynn, L. (2020). Research Ethics Committees, Ethnographers and Imaginations of Risk. *Ethnography* 1466138120983862, 146613812098386. doi:10.1177/1466138120983862
- Bishop, L. (2009). Ethical Sharing and Reuse of Qualitative Data. *Aust. J. Soc. Issues* 44 (3), 255–272. doi:10.1002/j.1839-4655.2009.tb00145.x
- Herle, A., and Philp, J. (2021). *Recording Kastom: Alfred Haddon's Journals from His Expeditions to Torres Strait and New Guinea, 1888 and 1898*. Sydney: Sydney University Press.
- Kuula, A. (2011). Methodological and Ethical Dilemmas of Archiving Qualitative Data. *Iq* 34 (3-4), 12–17. doi:10.29173/iq455
- Lejeune, P. (1989). *On Autobiography*. Minneapolis: University of Minnesota Press.
- Lowman, J., and Palys, T. (2013). The Betrayal of Research Confidentiality in British Sociology. *Res. Ethics* 10 (2), 97–118. doi:10.1177/1747016113481145
- Neale, B., and Bishop, L. (2012). The Ethics of Archiving and Re-using Qualitative Longitudinal Data: a Stakeholder Approach. *Timescapes Methods Guides Ser.* [Online] 18. 12(1), Available at: <http://www.timescapes.leeds.ac.uk/assets/files/methods-guides/timescapes-neale-ethics-archiving.pdf>.
- Parry, O., and Mauthner, N. S. (2004). Whose Data Are They Anyway? *Sociology* 38 (1), 139–152. doi:10.1177/0038038504039366
- Russell, L., and Barley, R. (2019). Ethnography, Ethics and Ownership of Data. *Ethnography* 21 (1), 5–25. doi:10.1177/1466138119859386
- Wynn, L. L., and Israel, M. (2018). The Fetishes of Consent: Signatures, Paper, and Writing in Research Ethics Review. *Am. Anthropologist* 120 (4), 795–806. doi:10.1111/aman.13148
- Yuill, C. (2018). 'Is Anthropology Legal?' *Anthropol. Action*. 25 (2), 36–41. doi:10.3167/aia.2018.250205
- Zavella, P. (1996). "Feminist Insider Dilemmas: Constructing Ethnic Identity with Chicana Informants," in *Feminist Dilemmas in Fieldwork*. Editors L.W. Diane and D. Carmen Diana (London: Routledge), 138–159.
- Zeitlyn, D. (Forthcoming 2022). Archiving Ethnography? the Impossibility and the Necessity. *Damned if We Do, Damned if We Don't. Ateliers d'anthropologie* 51.

ACKNOWLEDGMENTS

The example cited in this piece and surrounding text has been taken from (Zeitlyn, Forthcoming 2022) with permission of the author. The Author is happy to acknowledge helpful and clarifying discussions at the Colloquium "Valoriser les archives des ethnologues Usages contemporains des collections" BNF, Paris, 4-5-6 October 2018. Some of the issues featured in a talk also in October 2018 at All Souls College, Oxford for the *Open Science Seminar* convened by Lisa Lodwick and Jasmine Nirody and at L'école d'été UMC—Marseilles "Understanding Mediterranean Collections," 17 July 2019. Another, shorter, version was presented remotely in November 2018 at a AAA Roundtable "Approaches to Expanding the Use of Anthropological Archives" convened by Diana Marsh. The Author is very grateful for all the invitations and for the feedback from these presentations and on-going conversations with Kirsten Bell, who also made extremely helpful comments on a draft.

Conflict of Interest: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Zeitlyn. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

¹²Blissett, L. 2006. The Hau of the Paper and Dividual Authors: Reimagining Authorship in Anthropology. Unpublished manuscript.