



OPEN LETTER

The destruction of the 'Windrush' disembarkation cards: a lost opportunity and the (re)emergence of Data Protection regulation as a threat to longitudinal research [version 1; referees: 2 approved]

Andy Boyd ^{1,2}, Matthew Woollard ³, John Macleod⁴, Alison Park ²

¹ALSPAC, University of Bristol, Bristol, BS8 2BN, UK

²CLOSER, UCL Institute of Education, London, WC1H 0NU, UK

³UK Data Archive, University of Essex, Colchester, CO4 3SQ, UK

⁴Population Health Science, University of Bristol, Bristol, BS8 2BN, UK

v1 First published: 11 Sep 2018, 3:112 (doi: [10.12688/wellcomeopenres.14796.1](https://doi.org/10.12688/wellcomeopenres.14796.1))
Latest published: 11 Sep 2018, 3:112 (doi: [10.12688/wellcomeopenres.14796.1](https://doi.org/10.12688/wellcomeopenres.14796.1))

Abstract

Historical records and the research databases of completed studies have the potential either to establish new research studies or to inform follow-up studies assessing long-term health and social outcomes. Yet, such records are at risk of destruction resulting from misconceptions about data protection legislation and research ethics. The recent destruction of the Windrush disembarkation cards, which potentially could have formed the basis of a retrospective cohort study, illustrates this risk. As organisations across Europe transition to the EU General Data Protection Regulation (GDPR), this risk is being amplified due to uncertainty as to how to comply with complex new rules, and the requirement under GDPR that data owners catalogue their data and set data retention and destruction rules. The combination of these factors suggests there is a new meaningful risk that scientifically important historical records will be destroyed, despite the fact that GDPR provides a clear legal basis to hold historical records and to repurpose them for research for the public good. This letter describes this risk; details the legal basis enabling the retention and repurposing of these data; makes recommendations as to how to alleviate this risk; and finally encourages the research and research-active clinical community to contact their 'Data Protection Officers' to promote safe-keeping of historical records.

Keywords

GDPR, data protection, research archive, data retention, data repurpose, retrospective cohort study, follow-up, Windrush.



This article is included in the [Avon Longitudinal Study of Parents and Children \(ALSPAC\)](#) gateway.

Open Peer Review

Referee Status:

	Invited Referees	
	1	2
version 1 published 11 Sep 2018	 report	 report

1 **Natalie Banner** , Wellcome, UK

2 **Elizabeth Lomas** , University College London, UK

Discuss this article

[Comments](#) (0)

Corresponding author: Andy Boyd (a.w.boyd@bristol.ac.uk)

Author roles: **Boyd A:** Conceptualization, Investigation, Writing – Original Draft Preparation, Writing – Review & Editing; **Woollard M:** Writing – Review & Editing; **Macleod J:** Writing – Review & Editing; **Park A:** Conceptualization, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: AB and AP are funded by a research programme entitled ‘Cohorts and Longitudinal Studies Enhancement Resources’ (CLOSER), ESRC grant reference: ES/K000357/1, where AP is Principal Investigator. The ESRC funded a series of workshops and telephone discussions with members of the UK longitudinal research community, research funders and the UK Data Protection regulator (attended by AB, MW, AP); where the risk of unwarranted data destruction was initially identified (although not discussed in the context of retrospective research studies). The paper draws on insights from the governance work package of the Project to Enhance ALSPAC through Record Linkage (PEARL) – a Wellcome Trust award (WT grant reference: WT086118/Z/08/Z, where JM is Principal Investigator) that supports authors AB and JM and the information security work conducted in ALSPAC. Core funding for ALSPAC is provided by the UK Medical Research Council and Wellcome Trust (102215/2/13/2, where George Davey Smith and Nic Timpson are Principal Investigators) and the University of Bristol.

Copyright: © 2018 Boyd A *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution Licence](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Boyd A, Woollard M, Macleod J and Park A. **The destruction of the ‘Windrush’ disembarkation cards: a lost opportunity and the (re)emergence of Data Protection regulation as a threat to longitudinal research [version 1; referees: 2 approved]** Wellcome Open Research 2018, 3:112 (doi: [10.12688/wellcomeopenres.14796.1](https://doi.org/10.12688/wellcomeopenres.14796.1))

First published: 11 Sep 2018, 3:112 (doi: [10.12688/wellcomeopenres.14796.1](https://doi.org/10.12688/wellcomeopenres.14796.1))

Disclaimer

The views expressed in this article are those of the author(s). Publication in Wellcome Open Research does not imply endorsement by Wellcome.

Introduction

The UK state mistreatment of the ‘Windrush generation’ who migrated to the UK from the West Indies during the 1950s and 60s is a public scandal, but also illustrates a meaningful (re)emerging risk to longitudinal research. The scandal emerged from the ‘hostile environment’¹ resulting from government policy to reduce UK immigration rates. Individuals legitimately living in the UK were forced to demonstrate their residency; while in parallel, officials destroyed—reportedly due to ‘Data Protection’ requirements²—the disembarkation records that could help prove citizenship status. Aside from their (disputed) utility to demonstrate citizenship, these disembarkation cards, a record of ~0.5 million migrants from a defined geographical location arriving in Britain following the end of the Second World War³, could have been the starting point for a retrospective cohort study. With increasing realisation of record linkage strategies to enable retrospective and prospective follow-up, could a ‘Windrush Cohort’ have provided unique insights into the health and social outcomes of these migrants as they entered old age? Could such a cohort illustrate patterns of migration, community formation, economic and social integration and health outcomes within a population with distinct genomic and phenotypic characteristics and whom faced considerable socio-economic adversity? How would members of this community have felt about such a cohort? We may never know; while some equivalent records exist on a sub-set of this population, by destroying the disembarkation cards, this tantalizing possibility may be beyond reach. In this article we reflect on the value of repurposing historical records or historical research databases in long-term outcome studies; and then, assess the legal basis for retaining such records and repurposing them in this manner under the new EU General Data Protection Regulation (GDPR). We end by setting out provisional recommendations for clearer guidance and working practices that retain the potential for these records to inform future research, while retaining public acceptability.

The value of repurposing historical records

Repurposing historical records to define retrospective cohorts or to allow long-term follow-up of clinical trials is not a trivial exercise, but has immense epidemiological value. Such studies effectively shortcut the passing of time: allowing the assessment of early exposures, including randomly allocated interventions, on outcomes decades later without waiting for those years to pass. Cohort examples include the Lothian Birth Cohorts⁴, the Hertfordshire and Helsinki Cohorts⁵, and the Boyd Orr Cohort⁶, which have all repurposed historical records as a basis for contemporary studies. The Barry-Caerphilly⁷, Sorrento⁸, and Aberdeen⁹ trial follow-up studies illustrate the potential to measure health outcomes long after the original interventions. The Historical Sample of the Netherlands¹⁰ illustrates a variation on this theme: where registry records have been collated into a multi-generational longitudinal ‘spine’ linkable to other databases, an approach

recognised as having potential to facilitate longitudinal research¹¹. These studies have made innovative use of historical records, established platforms for interdisciplinary research, and in turn, produced a wealth of research outputs (see [Table 1](#)). None of these would have been possible without the (sometimes accidental) preservation of the underlying, identifiable, historical records.

Re-emerging risks of a ‘bonfire of the records’

The risk of a ‘bonfire of the records’¹² exists where ‘data protection’ concerns lead to the destruction of historical archives. The same risk also applies to existing cohort studies, particularly when: study participants are in transition (e.g. child participants reaching legal majority, or aging participants lacking the capacity for ongoing follow-up); studies face gaps in funding; intervention trials reach the end of their initial protocol; or when participant consent is no longer considered valid. A perceived ‘end’ of study could be coupled with pressure to destroy research databases or render the data anonymous. Even where anonymised records survive, these actions may preclude new data linkage opportunities, hamper integration into study consortia, or hinder assessments of long-term outcomes.

The risk of data destruction is re-emerging through the new EU GDPR and national implementations of the regulation^{13,14}. Specifically, as we discuss below, through the potential for misinterpretation and as a result of the requirements within GDPR to audit information, compile catalogues of information with assigned data retention and destruction requirements. GDPR, like predecessor legislation, relates to ‘personal data’, i.e. data that can be related to specific living individuals. GDPR includes rephrased data retention requirements (GDPR Article 5(1)(e)), and new requirements that organisations systematically catalogue data assets (GDPR Article 30) and define retention and destruction rules for each type of asset (GDPR Recital 39). In practice, this means that universities, government departments and health bodies across Europe are conducting data audits, cataloguing data—including historical records and medical research databases—assessing their ‘lifespan’ and determining when they should be destroyed. Whether from sincere attempts to comply with new legislation, ignorance of historical data’s future research potential, or from pressure to clear shelf/server space, it is likely that historical records—equivalent to those used to establish the Boyd Orr, Lothian, Hertfordshire and Helsinki cohorts—are now under considerable threat of destruction. To counter this threat, it is important that the ethico-legal basis for retaining these records is clarified and communicated: here, we illustrate a case using UK Data Protection law as an example.

The legal basis for retaining and repurposing historical records for research

The former UK Data Protection Act 1998 (DPA98) stated that “Personal data shall be obtained only for one or more specified and lawful purposes, and shall not be further processed in any manner incompatible with that purpose” (DPA98 Principle 2) and that data “shall not be kept for longer than is necessary

Table 1. Example longitudinal resources sampled from historical records.

Study	Location (City/Region, Country)	Historical records used for sampling	Sampling frame era
Lothian Birth Cohorts	Scotland, UK	School administered intelligence tests	1921 and 1936
		<i>Aim:</i>	To assess cognitive change over the life-course
		<i>Notable findings/outputs:</i>	Identified that childhood cognitive ability accounts for half the variance in ability in older age
Hertfordshire Cohort Study	Hertfordshire, UK	Midwifery registers	1931–1939
Helsinki Cohort Study	Helsinki, Finland	Child welfare clinic records	1934–1944
		<i>Aim:</i>	To generate evidence to support the 'fetal origins hypothesis'
		<i>Notable findings/outputs:</i>	Established association between early developmental conditions and adult health outcomes
Boyd Orr Cohort	UK	Family Diet and Health' survey	1937–1939
		<i>Aim:</i>	To investigate early-life dietary exposures on adult health outcomes
		<i>Notable findings/outputs:</i>	1) Identified links between child diet and cancer outcomes 2) Demonstrated the impact of nutrition on subsequent inter-generational health, and of breastfeeding on later cardiovascular mortality
Barry Caerphilly Growth Study	Wales, UK	Intervention trial records	1972–1974
Sorrento Maternity Hospital Study	Birmingham, UK	Intervention trial records	1979–1980
Aberdeen Folic Acid Supplementation Trial (AFAST)		Intervention trial records	1966–1967
		<i>Aim:</i>	To conduct long-term follow-up of trials of nutritional manipulation during pregnancy to enable the examination of effects on outcomes in adult offspring
		<i>Notable findings/outputs:</i>	1) In Barry Caerphilly those given the intervention (free milk in pregnancy and early childhood) had lower Insulin-like growth factors (IGF-1) than the control group. 2) The Sorrento study found no evidence that nutritional supplements given to pregnant women are an important influence on adult disease risk; 3) In AFAST, findings suggest that suggest that maternal folic-acid supplement use is associated with changes in the DNA methylation of the offspring that persist for many years after exposure in utero.
Historical Sample of the Netherlands	Netherlands	Birth, death, marriage certificates; population registers	1812–1920
		<i>Aim:</i>	To produce a representative resource for demographic, social science and epidemiological research
		<i>Notable findings/outputs:</i>	Established a national 'life history' database which can be linked to other resources for diverse research applications

for that purpose” (DPA98 Principle 5): suggesting that retaining and repurposing historical records is prohibited. However, the value of research is acknowledged in research exemptions (DPA98 s33); in regulatory guidance stating that “records selected for permanent preservation as archives, with a view to their use in historical or other research” is a legitimate ‘purpose’

within DPA98¹⁵, and a regulator endorsed code of practice stating:¹⁶

“4.3.2 There is a danger that over-cautious interpretation of the Act may lead to the weeding, anonymising or destruction of files containing

personal data that would otherwise be passed to the archives repository. An archivist's ability within the Act permanently to retain personal and sensitive personal data for the purposes of research (see 4.2.1) should therefore be made clear to potential depositors. The legislation contains the necessary safeguards for depositors."

If the Windrush archive was destroyed due to concerns around data protection (specifically the requirements of the DPA98, which was in force at that time), either these safeguards were not sufficiently communicated; or insufficient safeguards existed to protect against destruction from individuals who could not perceive the wider value of the records. The impetus for the new GDPR, along with the UK Data Protection Bill (DPA18), is to protect citizen interests in a digitized world where personal data are a 'monetized' commodity with both legitimate and transparent use and illegitimate and opaque use. Such protection is established in core principles (GDPR Article 5). Following extensive lobbying¹⁷, GDPR recognises the benefits of research conducted for the public good and provides archive and research exemptions to the purpose and storage limitations (subject to Article 89(1) safeguards). This results in a permissive legislative framework for research and archiving, while raising data management standards and providing freedom of academic expression (Article 85(1)). GDPR recognises, in Article 5(1)(e), that any data can be used for research and stored as such for many years:

"personal data may be stored for longer periods insofar as the personal data will be processed solely for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes in accordance with Article 89(1)"

The UK government have clarified that archiving is permitted within the proposed DPA18¹⁸. However, research data or archive records are not exempt from GDPR requirements to catalogue 'information assets' and set retention timescales. Therefore, the risk of records of research importance being destroyed comes from operational decisions made by administrative rather than research staff; and likely from a place of not fully understanding the research aspects of the new legislation.

A call for clearer guidance

There is therefore an urgent need for operational guidance describing the options for retaining and repurposing this class of information for research purposes. The default should be that information with potential research/historical interest—a deliberately broad category—should be retained rather than destroyed. And that retention should be indefinite, albeit subject to periodic review where the benefits and risks of holding data can be assessed in a contemporaneous context. We argue this from the position that GDPR is not intended to curtail research,

that research aiming to improve public goods typically enjoys strong public support, and the low risk to data subjects resulting from archiving or research follow-up. Following lobbying from the research community, the Information Commissioner's Office (the UK Data Protection regulator) has issued guidance supporting this position¹⁹. The research ethics community should produce new guidance for researchers designing new study (e.g. RCT) protocols, and new guidance for ethics committee members assessing them. Best practice standards (such as those evident in the findings of the 'Understanding Patient Data' taskforce in the UK) should be followed in order to develop clear and consistent public-facing language that explains the principles and reasons for retaining and repurposing records. There remains an open question as to how best communicate data repurposing, and how to engage the public in this activity. While the mechanism for this is likely to be context specific, we suggest that prior to repurposing comprehensive efforts should be made to engage data subjects, or where this is not possible, to test the public acceptability of the specific data reuse. We support the need that the retention of records is subject to meeting safeguarding requirements (GDPR Article 89) insofar as these do not limit future (and as yet unspecified) needs; and that repurposing records requires appropriate governance safeguards (e.g. research ethics approval with meaningful public input); and both should occur within socially acceptable frameworks²⁰.

We set out provisional recommendations for addressing this risk within [Box 1](#), but consider that these need developing with professional (e.g. the Archives and Records Association and The National Archives in the UK) and public input. Resulting clear guidance should be communicated to institution 'Data Protection Officers'—a staff role that all organisations must now have, as mandated under GDPR—to clarify the legal basis for retaining and repurposing records. Institutional and national archives could curate information from dormant studies (e.g. the UK Data Service's [ReShare](#) online repository, or the University of Bristol's [data.bris](#) archive). However, while such repositories could have value in reducing risks and increasing public acceptability, these will do little to address our primary concern that the research value of historical records is not clearly perceived.

Conclusion

It is clear from the examples considered above that repurposed historical research and administrative records can make unparalleled contributions to driving improved scientific understanding and improving health and social policy. To neglect the safe keeping of these records will be to neglect this aspect of longitudinal research. We encourage researchers and research-active clinicians across Europe to contact their Data Protection Officer to alleviate this risk. The wider research community should work with regulators, archivists, data managers and other stakeholders to ensure long-term retention of these data in a manner that is fair and transparent to data subjects.

Box 1. Recommendations for long-term retention and subsequent repurposing of records in a publicly acceptable manner

1) That EU Data Protection regulators (referred to as national authorities in the GDPR) consider the risk of unnecessary record destruction and either:

- i. provide clear guidance to records managers regarding the permissive nature of GDPR for the long-term retention of personal data in archives, and potential for repurposing of these archives for research purposes;

Or

- ii. ratify a code of conduct providing this guidance that has been developed by appropriate national organizations.

2) That this precise scenario is communicated to key individuals; i.e. those likely to be conducting data audits and making decisions about retention schedules (e.g. all Data Protection Officers (a post mandated within GDPR), research managers, records managers, research ethics chairs and facilitators, research funders archivists).

3) That data retention decisions are recorded by those with responsibility for managing data (with oversight from Data Protection Officers), that decisions are internally audited and retained to ensure institutional memory of the value of specific records.

4) That where historical personal data are scheduled for destruction, and there is doubt regarding their retention value, that guidance is sought and due process is followed. It should be recognized that individuals may be confident there is 'no doubt' about the (lack of) value in a record: and that this can only be countered by raising the general appreciation of the value of these data. Consideration should be made at a national level as to the best means to achieve this in a transparent manner with input from diverse stakeholders. Options could include a panel of interdisciplinary experts to provide such guidance, or a system of public notifications where destruction is embargoed until a 'consultation' time period expires (i.e. a system akin to the UK land use planning permission system).

5) Both new and existing studies (both observational and interventional) shall (where practicable) inform potential participants that the personal data they provide will be retained for long periods (perhaps indefinitely with periodic retention reviews) and potentially repurposed within a given governance framework.

6) That research funders produce clear guidance on personal data retention and repurposing; and promote their funding mechanisms to support long-term archiving of important records.

7) That, prior to repurposing taking place, a code of practice is developed that establishes key principles. These principles should include guidance on:

- i) the requisite safeguards;
- ii) how representatives of the participant community are engaged and how their views are integrated into the research design in a meaningful manner;
- ii) how participants are informed about repurposing (given that direct contact with all individuals may be impracticable) and if they are given the right to object;
- iii) how the principle of fairness is retained during the repurposing of records for new research purposes

8) As a research and archive community we should learn from recent data scandals and consult the public on their understanding and expectations relating to retention and repurposing, and ensure appropriate safeguards are implemented to ensure continuing acceptability.

9) Information about archived records should be discoverable to the public (to ensure retention and repurposing is transparent) and researchers (in order to maximise appropriate use).

Grant information

AB and AP are funded by a research programme entitled 'Cohorts and Longitudinal Studies Enhancement Resources' (CLOSER), ESRC grant reference: ES/K000357/1, where AP is Principal Investigator. The ESRC funded a series of workshops and telephone discussions with members of the UK longitudinal research community, research funders and the UK Data Protection regulator (attended by AB, MW, AP); where the risk of unwarranted data destruction was initially identified (although not discussed in the context of retrospective research studies). The paper draws on insights from the governance work package of the Project to Enhance ALSPAC through Record Linkage (PEARL) – a Wellcome Trust award (WT grant reference: 086118, where JM is Principal Investigator) that supports authors AB and JM and the information security work conducted in ALSPAC. Core funding for ALSPAC is provided

by the UK Medical Research Council and Wellcome Trust (102215), where George Davey Smith and Nic Timpson are Principal Investigators) and the University of Bristol.

Acknowledgements

This paper originates in GDPR discussion workshops organised by the UK Economic Social Research Council. We thank Sarah Dickson, Alex Bailey, Geraldine Clement-Stoneham (UK Medical Research Council); Allan Sudlow, Ben White (British Library); and George Davey Smith (University of Bristol) for their helpful discussions. We note the views in this manuscript do not necessarily reflect their views or those of their institutions. Authors AB and AP conceived this paper, AB drafted the manuscript and AB, MW, JM and AP developed the final version. AB is guarantor of the content.

References

1. Consterdine E: **Hostile environment: the UK government's draconian immigration policy explained.** *The Conversation*.
[Reference Source](#)
2. The Guardian: **Home Office destroyed Windrush landing cards, says ex-staffer.** 2018.
[Reference Source](#)
3. Peach C: **The Caribbean in Europe: contrasting patterns of migration and settlement in Britain, France and the Netherlands.** Coventry: Centre for Research in Ethnic Relations, University of Warwick; 1991.
[Reference Source](#)
4. Deary IJ, Gow AJ, Pattie A, *et al.*: **Cohort profile: the Lothian Birth Cohorts of 1921 and 1936.** *Int J Epidemiol.* 2012; **41**(6): 1576–84.
[PubMed Abstract](#) | [Publisher Full Text](#)
5. Barker DJ, Osmond C, Kajantie E, *et al.*: **Growth and chronic disease: findings in the Helsinki Birth Cohort.** *Ann Hum Biol.* 2009; **36**(5): 445–58.
[PubMed Abstract](#) | [Publisher Full Text](#)
6. Martin RM, Gunnell D, Pemberton J, *et al.*: **Cohort profile: The Boyd Orr cohort—a historical cohort study based on the 65 year follow-up of the Carnegie Survey of Diet and Health (1937-39).** *Int J Epidemiol.* 2005; **34**(4): 742–9.
[PubMed Abstract](#) | [Publisher Full Text](#)
7. Ben-Shlomo Y, Holly J, McCarthy A, *et al.*: **Prenatal and postnatal milk supplementation and adult insulin-like growth factor I: long-term follow-up of a randomized controlled trial.** *Cancer Epidemiol Biomarkers Prev.* 2005; **14**(5): 1336–9.
[PubMed Abstract](#) | [Publisher Full Text](#)
8. Macleod J, Tang L, Hobbs FD, *et al.*: **Effects of nutritional supplementation during pregnancy on early adult disease risk: follow up of offspring of participants in a randomised controlled trial investigating effects of supplementation on infant birth weight.** *PLoS One.* 2013; **8**(12): e83371.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
9. Richmond RC, Sharp GC, Herbert G, *et al.*: **The long-term impact of folic acid in pregnancy on offspring DNA methylation: follow-up of the Aberdeen Folic Acid Supplementation Trial (AFASST).** *Int J Epidemiol.* 2018; **47**(3): 928–937.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
10. Mandemakers K: **Historical sample of the Netherlands.** Handbook of international historical microdata for population research. 2000; 149–77.
[Reference Source](#)
11. Davis-Kean P, Chambers RL, Davidson LL, *et al.*: **Longitudinal Studies Strategic Review: 2017 Report to the Economic and Social Research Council.**
[Reference Source](#)
12. Statement from the Archives and Records Association: **ARA Urges Modernisation of UK Government Record keeping.** 2018.
[Reference Source](#)
13. McCall B: **What does the GDPR mean for the medical community?** *Lancet.* 2018; **391**(10127): 1249–50.
[PubMed Abstract](#) | [Publisher Full Text](#)
14. John B: **Are you ready for General Data Protection Regulation?** *BMJ.* 2018; **360**: k941.
[PubMed Abstract](#) | [Publisher Full Text](#)
15. National Records of Scotland: **Data Protection Code of Practice: Archive Collections.** 2013.
[Reference Source](#)
16. The National Archives, Society of Archivists, Records Management Society, *et al.*: **Code of Practice for Archivists and Records Managers under Section 51(4) of the Data Protection Act 1998.** 2007.
[Reference Source](#)
17. **Ensuring a healthy future for scientific research through the Data Protection Regulation 2012/0011(COD).** 2015.
[Reference Source](#)
18. **Nick Thomas-Symonds MP asked the Secretary of State for Digital, Culture, Media and Sport what definition her Department uses for the phrase 'archiving in the public interest' (Written question – 111381. 13 November 2017).**
[Reference Source](#)
19. Information Commissioner's Office: **Guide to the General Data Protection Regulation, Principle (e): Storage Limitation.** 2018.
[Reference Source](#)
20. Carter P, Laurie GT, Dixon-Woods M: **The social licence for research: why care.data ran into trouble.** *J Med Ethics.* 2015; **41**(5): 404–9.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Open Peer Review

Current Referee Status:  

Version 1

Referee Report 02 October 2018

doi:[10.21956/wellcomeopenres.16124.r33871](https://doi.org/10.21956/wellcomeopenres.16124.r33871)



Elizabeth Lomas 

iSchool, Department of Information Studies, University College London, London, UK

This open letter is submitted by a group of academics with specific expertise in the use of data for longitudinal studies. As such they offer a valuable perspective on record retention decisions and perceived risks influencing these decisions. In particular, they highlight the dangers posed by misinterpretations of the requirements of the new General Data Protection Regulation (GDPR) and Data Protection Act 2018 which together legislate for the management regimes for personal data in the UK. The authors make a call for further guidance on the correct application of data protection law to take account of archival considerations. This review of the letter draws out some points for the authors further reflection. The review does not dispute the viewpoints and credibility of the piece but rather raises some issues for consideration. Overall it is the conclusion of the reviewer that it is important not only that this letter has been published but that the discussion contained within it is further publicly evolved beyond this piece. This review highlights some further guidance which is relevant to the discussion.

The focus of this letter relates to the historical research and archival value of holding records through time. The letter uses the emotive case of the destruction of the 'Windrush' disembarkation records as a lens to evidence key issues. The authors rightly state that the use of these records might have enabled the identification of a particular cohort with a range of challenges worthy of special examination, for example in terms of their social, economic and health outcomes through time. Whilst the initial case for the value and retention of the landing cards of the so called 'Windrush Generation' is made persuasively within the piece, it would have been beneficial to discuss this in greater detail to develop the case that these records did merit permanent archival preservation. The National Archives does hold Board of Trade passenger lists (BT26) which provide some of the same data covered by the landing cards. These are given limited discussion on the website of The National Archives by Kershaw (2018)¹. Significantly, the passenger lists span a much narrower period from 1947-1953. However, the question arises as to whether from a historical/archival perspective the landing cards from this same period would have added significantly to the data captured by the passenger lists. I would have liked some further perspectives on the value of the cards for the period. Where multiple data sets/records exist, with overlapping information, researchers need to argue persuasively for the retention of each data set/record series. All data retention comes at a cost to the public purse and not everything can be kept so we do need more ongoing debates around our expectations for archiving. The authors were well placed to develop and evidence some of these arguments.

In the Windrush case, the Home Office should have been best placed to comment on the current operational needs for retention. Government Department officials will have known, or certainly should have known if tasked with this responsibility, that the landing cards could have been retained for

operational purposes and as a separate consideration for permanent archival preservation. It appears that it was reasoned the cards were no longer required when considered in relation to the storage costs. A key issue in this example, from an operational perspective, is that it has been reported in the press that the records were still in current use when the disposal decision was made in 2009. There are balances to be struck in retention and destruction decisions. Any comment on the 2009 decision needs to be in the context of the operational requirements at the time of the evaluation. However, these records became of greater significance due to the Home Office later changing its rules in terms of the expectations of individuals to document their lives in the UK. The balance between State and individual recordkeeping responsibilities would merit a whole separate discussion. In regards to the archival selection processes it is important, and perhaps surprising to many, to note that the Home Office held the responsibility not only for the decisions concerning the current operational retention of these records but also the historical/archival selection decisions. This letter could have opened up this point to further consideration. The National Archives does work closely with Departments to determine the records which need to be kept permanently. Operational Selection Policies set out the selection criteria (see <http://www.nationalarchives.gov.uk/information-management/manage-information/selection-and-transfer/>) in conjunction with other overarching guidance which is in the public domain (e.g. TNA, 2016²). It is to be observed in terms of The National Archives' collection policies, that traditionally there has been a greater emphasis on focusing resources on archiving policy documentation rather than necessarily on taking complete personal data sets. The letter strongly makes the case for considering in greater detail the value of personal data sets/record series as significant historical resources. The authors bring into the discussion further case examples of records with personal information, which have in some instances survived by accident rather than design. These records have subsequently been used for significant research purposes. The inclusion of a table of examples is particularly beneficially in providing an overview of these records. The focus on the value of different types of records provides a significant academic perspective which needs to be captured and considered more extensively and this could be done in relation to the published Operational Selection Policies which are the subject of public consultation. Therefore, I would suggest that the authors could make a stronger call for academic engagement with the Operational Selection Policies published by The National Archives and the related consultations processes. Various groups of stakeholders do advise The National Archives but there might be merit in a working group or other actions to better surface the issues around selection choices for personal data.

Having made a number of points relating to the long-term value of a range of type of records, the letter raises concerns surrounding data protection misunderstandings. This is an important component of the piece. The full complexity and risks around this could have been further developed. The letter makes the case that there does need to be better understanding of the legal rights to retain information for archival/historical research purposes. Under the new General Data Protection Regulation (GDPR)/Data Protection Act 2018 regime there is additional complexity in the legislation as there is a recognition of the need to retain personal data for historical and research purposes and in addition recognition of the value of archiving in the public interest. This is an important new layer in the legislation which strengthens the case for the long-term retention of personal data, the value of historical research and the role of archives. It is right that the message that personal data can be kept does need to be communicated and disseminated as widely as possible. The current operational, archival, and historical considerations for retaining personal data do influence record storage and management plans through time. The authors highlight a very real concern that organisations may not properly balance and weigh the full range of considerations as to whether or not to retain personal data. Whilst a significant focus of this piece relates to UK Government records, in fact potentially the biggest risks that records will be lost is likely to be in contexts outside of the public authority domain where there is no requirement to have archival processes in place, e.g. in terms of business and charities where accountability and data use through the longer term

will often not be a fundamental consideration. As such, I agree with the authors that the risk of data destruction is 're-emerging' and has been heightened by the new legislation. Under data protection law, if an organization is the target of a security attack, it will subsequently be scrutinized not only as to whether or not it took appropriate security measures but in addition, whether any of the personal data compromised could have been destroyed at an earlier point in time thus minimizing the scale and impact of the attack. This position is evidenced in the recent case of the credit agency Equifax, which had a cyber security breach that compromised the details of millions of global citizens. This breach has been public knowledge for some time but the ICO has issued a £500,000 fine since the publication of this letter (ICO, 2018³). The ICO make the point that Equifax had retained personal data beyond its operational requirement for this data and as such it could have been deleted/purged. The security breach occurred before the introduction of new data protection laws. Had the same breach occurred under new the General Data Protection Regulation (GDPR)/Data Protection Act 2018 regime the financial penalty could have been far higher. Organisations will be taking note of this and other such decisions and reviewing pending retention/disposition schedules in the light of this ruling. From a data protection risk management perspective this will encourage organisations to destroy personal data far earlier. In addition, individuals now have some rights of control over their data as there is now a 'right to be forgotten' in certain specific circumstances although not if the record/data is 'archived'. As such accidental survival is less likely as rightly organisations must manage their personal data. Equifax are not likely to have considered long term historical/archival factors in terms of data retention, but credit data will have some longer term value. The question is how best we decide what is of value and encourage organisations to serve the interests of individuals/customers taking into account societal needs through time. Organisations do need to be made aware of the need to review the retention of data with longer term considerations in mind. Archivists and research communities must articulate stronger calls for archival retention targeting and explaining the value of particular personal data information far sooner, if this information is to be retained. The case for retaining medical records has been well evidenced and seems to be widely understood but the same case has not been made for many other personal records/data sets with valuable cultural, social and economic information.

In terms of providing guidance in the area of data protection law, archivists are taking action although more can be done. The National Archives does have an FAQs page on GDPR <http://www.nationalarchives.gov.uk/archives-sector/advice-and-guidance/managing-your-collection/archiv>. Stronger links could be made between the ICO and TNA guidance. Significantly The National Archives has produced a *Guide to archiving personal data* which clearly states:

"the law recognises there is a public interest in permitting the permanent preservation of personal data for the long-term benefit of society" (TNA, 2018, p.6⁴)

In addition, the *Guide* has been supported by the ICO and has a foreword by the Information Commissioner, Elizabeth Denham, supporting the place of archives:

"Archives are special places. They are our collective memory. They help us to understand the past, make sense of the present, and guide us for the future. And in an age of fake news, misinformation and opaque institutions, archives are more important than ever in helping to uphold democracy and hold power to account." (TNA, 2018, p.4⁴)

The Archives and Records Association (<https://www.archives.org.uk/>) is in the process of producing more detailed guidance on the application of data protection law.

In summary this letter does present a timely perspective on the issues arising from the implementation of

data protection law. It is to be hoped that authors will continue to advocate to raise awareness of the concerns around potential data loss and the value of personal data more generally. If there were any more specific recommendations that could be made within the letter, then this would be beneficial. However, whilst data protection law has brought risks one thing GDPR and global digital developments have achieved is to generate a better understanding of the value and potential uses of personal data which has perhaps not been fully recognised in archival collection strategies. I hope the authors will continue to evolve this discussion and applaud the timely publication of this piece.

References

1. Kershaw R: Empire Windrush – 70 years on. *The National Archives*. 2018. [Reference Source](#)
2. Records collection policy. *The National Archives London: OPSI*. 2016. [Reference Source](#)
3. Equifax Ltd: monetary penalty notice, 19th September 2018. *ICO*. 2018. [Reference Source](#)
4. Guide to archiving personal data. *The National Archives London: OPSI*. 2018. [Reference Source](#)

Is the rationale for the Open Letter provided in sufficient detail?

Partly

Does the article adequately reference differing views and opinions?

Yes

Are all factual statements correct, and are statements and arguments made adequately supported by citations?

Yes

Is the Open Letter written in accessible language?

Yes

Where applicable, are recommendations and next steps explained clearly for others to follow?

Yes

Competing Interests: This review is written in a purely personal capacity. However, for transparency I wish to register that I am a Member of the Advisory Council on National Archives and Records as well as the Forum on Historical Manuscripts and Academic Research. In addition, I am a member of the Archives and Records Association's Legislation and Standards Working Group. I chaired the work on TNA's Guide to archiving personal data, which is cited in the review.

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Referee Report 24 September 2018

doi:[10.21956/wellcomeopenres.16124.r33873](https://doi.org/10.21956/wellcomeopenres.16124.r33873)



Natalie Banner 

Understanding Patient Data Lead, Wellcome, London, UK

This Open Letter raises a timely issue about the risk of historical records, that have high potential value for research, being destroyed by over-cautious interpretation of new data protection legislation. The

destruction of the ‘Windrush’ generation disembarkation cards is a clear case in point even though this occurred prior to the EU General Data Protection Regulation (GDPR) coming into force: these records could have provided a rich resource for retrospective research.

The introduction of the GDPR has thrown issues about data protection and retention into sharp relief among the research community, not least because of the high monetary penalties for failure to comply. It is also a complex piece of legislation, made more difficult to navigate in the UK context because of the UK’s domestic legislation having to perform a balancing act between pre-Brexit compliance and post-Brexit independence.

The authors are wise to focus in on one particular aspect of this complexity and to explore its implications for the specific but vitally important area of longitudinal research. The examples they draw on demonstrate clearly the value of using historical records for retrospective research, and set out why this kind of research may be at risk.

The letter needs updating in light of the UK Data Protection Act (2018) now being in force. As the GDPR Article 89(1) safeguards are derogated to Member States to specify, it would be helpful to include reference to the specific text used in the DPA (s.19) to describe these safeguards in the UK context. It is imperative that the conditions necessary to fulfil these safeguards are more widely understood among the research community as they provide:

1. the conditions under which data can be retained for archiving purposes, as the authors correctly note, but also
2. the only lawful route through which research with special categories of personal data (including data about health, racial or ethnic origin, genetics etc) can be conducted on a “public interests” lawful basis (via Article 9(j) of the GDPR)

In my view, disseminating understanding about the demands of the Article 89(1) safeguards needs to go beyond the audiences the authors identify: it is not only a matter for operational record managers and archivists, but for all researchers who use special category data in their research. I therefore consider that the authors have correctly diagnosed a significant concern but also underappreciated its potential impact beyond questions of long-term retention.

The recommendations are a helpful and concrete specification of actions needed from different stakeholders to rectify the lack of clarity about what the GDPR actually requires for long-term record retention. However, this is an area replete with guidance that is not always consistent and that can add to confusion, especially when compounded with differing institutional and funder policies. Rather than the ICO (the UK’s data protection regulator) issuing further guidance, further work to understand the specific areas of misunderstanding at a community and practice level, across research institutions, records offices, archives etc, may be a more fruitful starting point.

Overall, this Open Letter starkly illustrates a tangible risk emerging from misinterpretation of GDPR conditions for data retention, and one that should be drawn to wider attention among the longitudinal research community. The call for further public engagement and discussion about the value of historical records is encouraging, given the broader context of research participants and the public becoming increasingly aware of data about them being collected by various organisations for numerous purposes, and about their rights as data subjects. Research can and should be held to high standards of fairness, transparency and accountability and this letter is a step towards enabling a clear public rationale for why personal data should be retained for archiving purposes in the public interest.

Is the rationale for the Open Letter provided in sufficient detail?

Yes

Does the article adequately reference differing views and opinions?

Yes

Are all factual statements correct, and are statements and arguments made adequately supported by citations?

Partly

Is the Open Letter written in accessible language?

Yes

Where applicable, are recommendations and next steps explained clearly for others to follow?

Partly

Competing Interests: I lead the Understanding Patient Data initiative that is supportively referenced in the letter, and work for the Wellcome Trust (not in a funding division), which provided core funding for ALSPAC.

Referee Expertise: I work in the field of research data policy and public engagement on uses of health data for research. I led research-sector advocacy and engagement with the UK Parliament during the passage of the Data Protection Bill (2018).

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.
