Position Statement on Data Retention Time of Studies Published Long Time Ago

Review Article

EBWHJ- Editorial Board

ABSTRACT

Data retention is vital for research integrity but how long should the raw data be kept and who is responsible for this task over years and most importantly what about raw data of research published long time ago even before evolution of research integrity offices. Evidence based women's health society has concerns regarding requesting raw data of papers published in the Middle East long time ago(more than 15 years ago) as universities and publishers have no regulations regarding data retention at that time and decided to make a position statement regarding this specific issue. EBHWHS has its clear position that Balancing data accessibility with pragmatic retention limits is critical to sustaining trust in science. While regulations and RIOs provide structure, excessive demands for antiquated data undermine research integrity. Institutions should adopt clear retention timelines (e.g., 10–15 years) and reject frivolous requests that serve no public interest and may have hidden agenda.

Key Words: Data retention, integrity, research.

Received: 31 May 2025, **Accepted:** 01 June 2025.

Corresponding Author: Ahmed Yousef, Faculty of Medicine, Benha University, Egypt. **Tel.:** + 01112220298, **E-mail:** aoy3030@yahoo.com

ISSN: 2090-7265, 2025, Vol. 15

INTRODUCTION

Research integrity is the cornerstone of scientific progress, ensuring that findings are reliable, reproducible, and ethically obtained. Central to this principle is the responsible management of research data, including its retention. However, the question of how long data should be retained remains contentious, intersecting with legal, ethical, and practical considerations. This position statement from EBWHS examines global regulations governing data retention periods, traces the evolution of research integrity offices (RIOs) in academia and publishing, and argues that requests for raw data from studies published over 15 years ago are often impractical and may conceal ulterior motives.

Global Regulations on Data Retention Periods:

Data retention policies vary widely across jurisdictions and disciplines. For example, European Union (EU) General Data Protection Regulation (GDPR) mandates data minimization and storage limitation, requiring data to be kept "no longer than necessary"[1]. Sector-specific guidelines often define retention periods. Example: Clinical trials under the EU Clinical Trials Regulation (No. 536/2014) require retention for 25 years after trial completion, While Horizon Europe Funding Rules: Require data to be stored for at least 10 years post-project[2].

In United States, Health Insurance Portability and Accountability Act (HIPAA) stipulates that medical records must be retained for 6 years from creation or last use[3]. Conversely, Australian Code for the Responsible Conduct of Research (2018) Mandates retention for 5–15 years, depending on the field[4].

As for Publisher policies, Springer Nature encourages transparency and mandates that datasets supporting the conclusions of research articles be made available, but the exact duration for which data must be retained post-publication is not clearly defined while PLOS ONE requires data availability for 10 years[5].

The Evolution of Research Integrity Offices (RIOs):

All RIOs emerged in response to high-profile misconduct cases and the growing complexity of research ethics. In 1981: The U.S. Office of Scientific Integrity (OSI) was established following the fraud case of William Summerlin, who falsified transplant data. In 1992: OSI reorganized into the Office of Research Integrity (ORI), formalizing oversight of Public Health Service-funded research[6]. The UK Research Integrity Office (UKRIO) launched to advise institutions by 2000. It took almost twenty years to release an international consensus : The Singapore Statement on Research Integrity catalyzed global standards, urging institutions to establish RIOs in 2010. And by 2012 The Netherlands established the National Board for Research Integrity (LOWI)[7].

DOI:10.21608/EBWHJ.2025.391032.1468

The Illogic and Risks of Requesting Data >15 Years Post-Publication:

While transparency is vital, demands for raw data from decades-old studies often lack merit and raise red flags as there are practical challenges as technological Obsolescence: Data stored on outdated media (e.g., floppy disks) may be irretrievable. In addition, original researchers may retire or lose institutional affiliations. Data storage also costs as maintaining decades-old data strains institutional resources.

There are ethical and legal concerns of storing research data as older datasets may lack modern anonymization safeguards. The logic behind requesting raw data long after a study has been published can be questioned both ethically and legally as scientific knowledge evolves, older datasets may become less relevant to current research questions. The context in which the data was collected may no longer apply, making it less useful for contemporary analyses.

There could be also a hidden Agendas for complainants requesting raw data after long time from publications as they may seek to discredit researchers[8] for example the Danish study published in 2002 that debunked the myth linking the MMR (measles, mumps, rubella) vaccine to autism. Despite its robust findings, the study faced repeated data requests, legal challenges, and harassment from anti-vaccine activists, particularly those associated with Andrew Wakefield's discredited 1998 Lancet study (which falsely claimed an MMR-autism link)[9].

CONCLUSION

Balancing data accessibility with pragmatic retention limits is critical to sustaining trust in science. While regulations and RIOs provide structure, excessive demands for antiquated data undermine research integrity. Institutions should adopt clear retention timelines (e.g., 10–15 years) and reject frivolous requests that serve no public interest.

CONFLICT OF INTERESTS

There are no conflicts of interest.

REFERENCES

1. European Union. (2016). General Data Protection Regulation (GDPR) https://gdpr-info.eu/.
2. Horizon Europe Funding Rules 2021 . Http://efaidnbmnnnibpcajpcglclefindmkaj/https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/horizon/guidance/programme-guide_horizon_v1.0_en.pdf.
3. Australian Code for the Responsible Conduct of Research https://www.nhmrc.gov.au/about-us/publications/australian-code-responsible-conduct-research-2018.
4. Reporting standards and availability of data, materials, code and protocols https://www.nature.com/nature-portfolio/editorial-policies/reporting-standards.
5. Plos One Data Availability : https://journals.plos.org/plosone/s/data-availability.
6. U.S. Department of Health and Human Services(1996). HIPAA Privacy Rulehttps://www.hhs.gov/hipaa/for-professionals/privacy/index.html.
7. COPE Data integrity issues 2022: https://publicationethics.org/guidance/case/data-integrity-issues.
8. History of ORI : https://www.encyclopedia.com/science/encyclopedias-almanacs-transcripts-and-maps/office-research-integrity#:~:text=Five%20months%20later%20in%20August,educational%20programs%2C%20and%20research%20projects.
9. Madsen KM, Hviid A, Vestergaard M, Schendel D, Wohlfahrt J, Thorsen P, Olsen J, Melbye M. A population-based study of measles, mumps, and rubella vaccination and autism. N Engl J Med. 2002 Nov 7;347(19):1477-82. doi: 10.1056/NEJMoa021134.