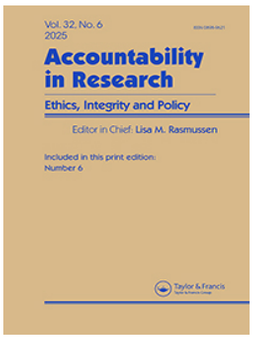

Taylor & Francis
Taylor & Francis Group

Accountability in Research
Ethics, Integrity and Policy

ISSN: 0898-9621 (Print) 1545-5815 (Online) Journal homepage: www.tandfonline.com/journals/gacr20

Disparate data retention standards in biomedical research

Carolyn T. Lye, Minal M. Caron, Lauren Walsh, Barbara E. Bierer & Mark Barnes

To cite this article: Carolyn T. Lye, Minal M. Caron, Lauren Walsh, Barbara E. Bierer & Mark Barnes (06 Aug 2025): Disparate data retention standards in biomedical research, Accountability in Research, DOI: 10.1080/08989621.2025.2543884

To link to this article: https://doi.org/10.1080/08989621.2025.2543884

Published online: 06 Aug 2025.

ACCOUNTABILITY IN RESEARCH, 2025
https://doi.org/10.1080/08989621.2025.2543884

Taylor & Francis
Taylor & Francis Group

Disparate data retention standards in biomedical research

Carolyn T. Lye [a], Minal M. Caron [a], Lauren Walsh [b], Barbara E. Bierer [c,d], and Mark Barnes [a,c]

[a]Ropes & Gray LLP, Boston, MA, USA; [b]Mass General Brigham, Boston, MA, USA; [c]Multi-Regional Clinical Trials Center of Brigham and Women's Hospital and Harvard, Boston, MA, USA; [d]Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA

ABSTRACT

The acquisition and retention of primary research data is fundamental to the reliability of and public trust in biomedical research. However, researchers are often unaware of or confused by applicable data retention requirements, in part due to the divergent requirements set forth by federal agencies, grant programs, and research institutions, as well as other applicable requirements under law, contract, and policy. This article summarizes current U.S. data retention standards applicable to biomedical research, including how institutional research retention practices have sought to reflect these standards, and discusses the importance of data retention in the context of research professionalism, data sharing efforts, intellectual property issues, and research integrity challenges, which increasingly have been the subject of much public interest. In conclusion, this article provides suggestions for both institutions and applicable federal agencies to streamline and clarify data retention standards, with the goal of improving research data retention practices.

ARTICLE HISTORY
Received 22 May 2025
Accepted 1 August 2025.

KEYWORDS
Data retention; biomedical research; research integrity

Introduction

In scientific research, the validity, reliability, and reproducibility of research findings depend upon the acquisition and retention of accurate, complete primary research data. Data retention is crucial to allow researchers to build upon previous research without duplicating that previous research. In addition, regulatory mandates, contract terms, and certain grant conditions may impose retention requirements on both the researchers who conduct a study and on the institutions at which the research takes place. In turn, research institutions often develop their own policies to govern research data storage and retention with which their

CONTACT Mark Barnes Mark.Barnes@ropesgray.com Ropes & Gray LLP, Boston, MA 02199, USA
*Dr. Bierer and Mr. Barnes serve as co-senior authors.

researchers must comply. Separately, researchers must preserve data in anticipation of defending themselves against challenges of research misconduct or other post-publication questions from scientific journals, institutions, or the media, and for supporting intellectual property and patent claims. Despite these compelling reasons to maintain rigorous data oversight and retention practices, researchers are often unaware of, confused by, or inattentive to data retention requirements applicable to their research, including those imposed by their research institutions. This may occur in part as a result of the divergent and confusing requirements set forth by various research institutions, federal agencies, and grant programs in the U.S. The lack of uniformity with respect to data retention standards is especially prominent in biomedical research in the U.S. context, as such research is often under the oversight of multiple federal agencies, while simultaneously subject to other requirements under law, contract, and applicable policies.

Given the recent attention that U.S. federal agencies and scientific journals have paid to making research data and publications publicly accessible, it is timely for both agencies and institutions to reconsider their data retention standards and coalesce around a single set of standards. In February 2023, the National Institutes of Health ("NIH"), the largest worldwide public funder of biomedical research, issued a notice requesting public input on its "Plan to Enhance Public Access to the Results of NIH-Supported Research" (NIH 2023b). NIH issued this plan in response to a request from the Office of Science and Technology Policy ("OSTP") for federal agencies to "develop new, or update existing, public access plans" to improve public access to research results while simultaneously employing accountability and validity measures to ensure scientific and research integrity. In the plan, NIH discusses its approach and timeline for data sharing by referencing its Data Management and Sharing Policy published and effective as of January 25, 2023, which directs NIH-supported investigators to consider "relevant requirements and expectations" for record retention. Subsequently, on December 17, 2024, NIH issued its 2024 NIH Public Access Policy, effective as of 1 July 2025, which requires that the final version of an accepted manuscript to be submitted to PubMed Central for public availability. While NIH has its own requirements for record retention, other government agencies that also have oversight of certain scientific research have set forth significantly different requirements regarding the time period within which researchers and institutions must retain research records and the types of data that must be retained (White House 2023). As a result of these multiple, inconsistent requirements, federal efforts to instigate broader data sharing and access may be stymied simply as a result of relevant data sets not being retained, or not being retained in a manner that preserves their provenance, accuracy, and reliability.

In this article, we first provide an overview of current U.S. data retention standards that apply to biomedical research, including data retention standards imposed by federal agencies and other non-research legal frameworks. Second, we review a sample of institutional research retention practices that seek to reflect these legal and regulatory standards. Third, we discuss the importance of data retention in the context of research integrity challenges, specifically in research misconduct proceedings, and more broadly as a key element of professional conduct in biomedical research. Finally, we provide recommendations for institutions and for OSTP, funding agencies, and other government stakeholders to streamline and clarify data retention standards.

Current landscape for data retention in biomedical research in the U.S.

Biomedical researchers and research institutions are typically required to comply with several different record retention obligations, including obligations of funders, such as NIH and National Science Foundation ("NSF"), and oversight agencies, such as the U.S. Food and Drug Administration ("FDA"), in addition to inferred data retention requirements under non-research legal frameworks, such as the Health Insurance Portability and Accountability Act of 1996 ("HIPAA"). In addition, research underlying intellectual property is subject to data retention periods that are effectively required by the term of patents, and research conducted at academic institutions is subject to institutional polices on data retention, which may or may not reflect data retention standards of federal agencies. Below, we summarize the key data retention standards in biomedical research in the U.S., which stem from: (i) funding and oversight agencies, (ii) institutional policy, and (iii) data retention requirements that are implied by HIPAA disclosure requirements, patent and license terms, and time limitations on the applicability of the research misconduct regulations.

Data retention standards of federal agencies

As a prolific biomedical research funder, NIH requirements impact many researchers and institutions. Institutions that receive research funding support from NIH must retain "financial and programmatic records, supporting documents, statistical records, and all other records that are required by the terms of the grant … , for a period of three years from the date the annual FFR [federal financial report] is submitted" (NIH 2024). For its own intramural research program, however, NIH requires that all research records must be maintained for at least 7 years after the completion of the project (*e.g.*, publication of the final results), and that research records that support intellectual property rights must be maintained for 30 years after the patent is

filed (NIH 2017). In short, even in regard to NIH's own research records, data retention standards are not uniform.

FDA, which regulates drugs and medical devices, among other products, has different data retention standards and requirements for investigational new drug ("IND") research and investigational device exemption ("IDE") research. IND research involves the clinical investigation of new drugs or biologics in human subjects. On the other hand, IDE research refers to the clinical investigation of medical devices. For IND research, under 21 C.F.R. § 312.62(c), FDA requires that records be maintained for 2 years after the marketing application approval date of the drug for the indication for which it is being investigated. For IDE research, under 21 C.F.R. § 812.140(d), FDA requires that records be maintained for 2 years after the latter of (i) the completion or termination date of the investigation or (ii) the date after which records are no longer needed to support any FDA submission. Industry sponsors typically include in their clinical trial agreements with hospitals and clinics a requirement that these study sites comply with these FDA records retention standards, even though the majority of sites may have limited visibility into some of these retention benchmarks, including, for example, the date on which a sponsor's investigation is completed or marketing application has been approved.

The record retention requirements of the Office of Research Integrity ("ORI"), which oversees research integrity issues for institutions that apply for or receive Public Health Service ("PHS") support for biomedical research, may be inferred from the ORI statute of limitations (which is, with exceptions, 6 years). The ORI statute of limitations requires that research data be retained for a longer time period than is required by the standards of other major federal agencies that fund or regulate research activities. Therefore, even if a researcher complies with the relevant funding agency's data retention requirements but disposes of the research data after that agency's retention period ends, the researcher may not have retained data for as long as is required for purposes of a research integrity proceeding. A failure to have retained data thus could prevent the relevant institution from fulfilling its legal obligations to the funding agency to investigate concerns about data integrity. Moreover, as further described below, failure to retain research records could also increase the risk that the researcher would be found to have committed misconduct if named as a respondent in a research misconduct matter. Finally, under 43 C.F.R. § 93.318(a), once a research misconduct proceeding has been completed, institutions must maintain the institutional record and all sequestered evidence for 7 years after completion of the proceeding.

Table 1 includes a summary of the applicable data retention standards for these and other major federal agencies and offices that fund and oversee biomedical research, demonstrating divergent requirements with respect to timeframes for data retention and the types or categories of data that must be retained.

Table 1. Data retention standards of select federal agencies and offices related to biomedical research.

Federal Agency or Office	Applicability of Retention Requirements	Required Data for Retention	Required Timeframe for Retention
National Institutes of Health	Recipients of NIH grant awards (NIH has separate guidelines for its Intramural Research Program not covered in Table 1)	Financial and programmatic records, supporting documents, statistical records, and all other records required by the terms of a grant, or may reasonably be pertinent to a grant. (NIH 2024).	A period of 3 years from the date the annual federal financial report is submitted to the NIH (NIH 2024). Exceptions: • 2 C.F.R. § 200.334(a): Records must be retained until all litigation, claims, or audit findings involving the records have been resolved and final action has been taken, if the litigation, claim, or audit began before the expiration of the 3-year period. • 2 C.F.R. § 200.334(b): NIH may otherwise require, through notice in writing, an extension of the retention period.
National Science Foundation	Recipients of NSF grant awards	Financial records, supporting documents, statistical records and all other records pertinent to the NSF grant, except as noted in 2 C.F.R. § 200.334 (Chapter VII.E., NSF, 2024).	A period of 3 years from award financial closeout (120 days after the award end date) (Chapters VII and VIII, NSF, 2024).

(*Continued*)

Table 1. (Continued).

Federal Agency or Office	Applicability of Retention Requirements	Required Data for Retention	Required Timeframe for Retention
U.S. Food and Drug Administration	Investigators of investigational new drugs; investigators and sponsors of investigational device exemptions	*Investigational New Drugs*: • Adequate records of the disposition of the drug, including data, quantity, and use by subjects; and • 21 C.F.R. § 312.62: Adequate and accurate case histories that record all observations and other data pertinent to the investigation on each individual administered the investigational drug or employed as a control in the investigation. *Investigational Device Exemption*: • 21 C.F.R. § 812.140: Any reports to FDA, records of receipt/use or disposition of the device, records of each subject's case history and exposure to the device, the protocol, and any other records required by FDA.	*Investigational New Drugs*: • 21 C.F.R. § 312.62: A period for 2 years following either: (1) the date the marketing application is approved for the drug for the specific indication for which it was investigated or (2) the date that the investigation was discontinued and FDA was notified. *Investigational Device Exemption*: • 21 C.F.R. § 812.140: A period of two after the latter of the following two dates: (1) the date on which the investigation is terminated or completed or (2) the date that the records are no longer required for purposes of supporting a premarket approval application, a notice of completion of a product development protocol, a humanitarian device exemption application, a premarket notification submission, or a request for de novo classification.

(*Continued*)

Table 1. (Continued).

Federal Agency or Office	Applicability of Retention Requirements	Required Data for Retention	Required Timeframe for Retention
Office of Research Integrity	42 C.F.R. § 93.100: HHS and institutions that apply for or receive PHS support for biomedical or behavioral research, biomedical research training, or activities related to that research or research training.	*Research misconduct proceeding*: • 42 C.F.R. § 93.220(a): Records of research misconduct proceedings, including: ○ Documentation of the assessment as required by § 93.306(c). ○ If an inquiry is conducted, the inquiry report and all records (other than drafts of the report) considered or relied on during the inquiry, including, but not limited to, research records and the transcripts of any transcribed interviews conducted during the inquiry, information the respondent provided to the institution, and the documentation of any decision not to investigate as required by § 93.309(c). ○ If an investigation is conducted, the investigation report and all records (other than drafts of the report) considered or relied on during the investigation, including, but not limited to, research records, the transcripts of each interview conducted pursuant to § 93.310(g), and information the respondent provided to the institution. ○ Decision(s) by the Institutional Deciding Official, such as the written decision from the Institutional Deciding Official under § 93.314. ○ The complete record of any institutional appeal consistent with § 93.315.	*Research misconduct proceeding*: • 42 C.F.R. § 93.318(a): Institutions must maintain records of research misconduct proceedings in a secure manner for 7 years after completion of the research misconduct proceeding. • 42 C.F.R. § 93.104(b): The research misconduct regulations only apply to research misconduct occurring within 6 years of the date HHS or an institution receives an allegation of research misconduct, subject to certain exceptions.

(*Continued*)

Table 1. (Continued).

Federal Agency or Office	Applicability of Retention Requirements	Required Data for Retention	Required Timeframe for Retention
Office of Human Research Protections	45 C.F.R. § 46.101: All research involving human subjects conducted, supported, or otherwise subject to regulation by any Federal department or agency that takes appropriate administrative action to make the *HHS Policy for Protection of Human Research Subjects* applicable to such research.	45 C.F.R. § 46.115: An institution, or when appropriate an IRB, shall prepare and maintain adequate documentation of IRB activities, including the following: • Copies of all research proposals reviewed, scientific evaluations, if any, that accompany the proposals, approved sample consent forms, progress reports submitted by investigators, and reports of injuries to subjects. • Minutes of IRB meetings, which shall be in sufficient detail to show attendance at the meetings; actions taken by the IRB; the vote on these actions including the number of members voting for, against, and abstaining; the basis for requiring changes in or disapproving research; and a written summary of the discussion of controverted issues and their resolution. • Records of continuing review activities, including the rationale for conducting continuing review of research that otherwise would not require continuing review. • Copies of all correspondence between the IRB and the investigators. • A list of IRB members. • Written procedures for the IRB. • Statements of significant new findings provided to subjects. • The rationale for an expedited reviewer's determination that research appearing on the expedited review list is more than minimal risk. • Documentation specifying the responsibilities that an institution and an organization operating an IRB each will undertake to ensure compliance with the requirements of the agency or Department's policy.	45 C.F.R. § 46.115: At least 3 years, and records relating to research that is conducted shall be retained for at least 3 years after completion of the research.

(*Continued*)

Table 1. (Continued).

Federal Agency or Office	Applicability of Retention Requirements	Required Data for Retention	Required Timeframe for Retention
U.S. Department of Defense	32 C.F.R. § 219.101(a): All research involving human subjects conducted, supported, or otherwise subject to regulation by the Department of Defense.	32 C.F.R. § 219.115: Same requirements as for Office of Human Research Protections.	32 C.F.R. § 219.115: Same requirements as for Office of Human Research Protections.
U.S. Department of Veterans Affairs	38 C.F.R. § 16.101(a): All research involving human subjects conducted, supported, or otherwise subject to regulation by the Department of Veterans Affairs.	38 C.F.R. § 16.115: Same requirements as for Office of Human Research Protections.	38 C.F.R. § 16.115: Same requirements as for Office of Human Research Protections.

HIPAA

The disclosure requirements under HIPAA, although not specific to research, effectively impose data retention requirements that apply to human subjects research in which protected health information ("PHI") is used and disclosed. Specifically, the HIPAA Privacy Rule affords individuals the right to receive an accounting of disclosures of PHI made by a "covered entity" (including health plans, health care clearinghouses, and health care providers that transmit health information in covered transactions) during the 6 years prior to the date on which the accounting is requested, subject to certain exceptions provided in the regulations. These exceptions include, for example, disclosures made to carry out treatment, payment and health care operations and disclosures made pursuant to a valid authorization for the use or disclosure of PHI. While there are no specific exceptions to disclosures made in the setting of research, under 45 C.F.R. § 164.528(a)(1)(iv), there is an exception for disclosures of PHI made pursuant to a valid individual *authorization*. Thus, if the research is subject to HIPAA and the research subject for whom PHI was disclosed did not provide written authorization for the disclosure of their PHI for research purposes, study investigators and institutions must retain records of disclosures of such PHI for at least 6 years. Other data privacy laws at the state level also may effectively impose data retention requirements of which researchers and research institutions must be aware and must follow in their own jurisdictions.

Patents and licenses

Incentivized by the provisions of the Bayh-Dole Act to identify commercially viable intellectual property developed in federally-funded research, researchers seek to invent, and their institutions seek to assess for commercial promise and patenting, new technology. Under the Bayh-Dole provisions, researchers and their institutions essentially share income from out-licensing such inventions. The emergence of the technology transfer function at academic medical centers and other research institutions, and the prospect of deriving significant profits from inventions spawned by academic investigators, has created an additional set of data retention requirements. Specifically, a patent for an invention is the grant of a property right to the inventor, issued by the United States Patent and Trademark Office ("USPTO"), and its effective term generally is 20 years from the date on which the application for the patent was filed in the U.S., or in special cases, from the date an earlier related application was filed, subject to the payment of maintenance fees. To support potential questions or challenges regarding the patent, therefore, it is good practice to retain all data supporting a patent application for at least the 20-year life of the patent. Patent extensions, if sought and granted, may require even longer periods of data retention. Internal funding agency requirements can also result in longer data retention periods; for example,

as described above, research records maintained within the intramural research program at NIH that support intellectual property rights must be maintained for 30 years after the patent is filed (NIH 2023a). In short, the research data underlying intellectual property, and/or underpinning out-licensed technology, should be retained for at least as long as the life of the patent or license (*i.e.*, at least 20 years from the date on which the patent application was filed). Finally, and importantly, whether a research study will or could result in the invention of valuable intellectual property is most often not known or recognized during, or even immediately after, the research itself. Thus, researchers whose work has any possible commercial application should be aware of these lengthy data retention periods, and to the extent feasible, retain their data for the lengthy periods required in this patent context.

Academic institutions

Academic medical centers, universities, and other academic institutions at which biomedical research takes place set their own research data retention policies, and generally attempt to do so in compliance with the various regulatory requirements described above. As a result, and as discussed above and provided in Table 1, researchers within academic institutions may be subject to several different data retention standards, with different time periods during which data must be retained and different types of data that must be retained. Compounding these complications, institutional data retention policies may simply refer researchers to requirements imposed by applicable regulations or contract terms but without elaborating what those requirements are. Researchers at any given academic institution may not understand these federal standards and tend to expect that meeting their institution's own data retention policy will ensure compliance with applicable federal standards.

To demonstrate the range of data retention policies across institutions with medical schools, we compiled the publicly available institutional policies of the institutions with the top 10 medical schools in the United States, as ranked by U.S. News & World Report's in 2023 on the basis of their research enterprise (Table 2). The institutional policies vary widely in terms of the specific time periods during which data must be retained, as well as the scope of data-related topics covered by the policies. Further, many of these policies do not address the complexity created by the movement of researchers from one institution to another. Many researchers migrate to one or more institutions in their careers, and in doing so, often fail to retain or preserve their research data. While some institutions set forth clear expectations in this scenario, in many cases there is no formal communication between a researcher's prior and current institutions to discuss the allocation of data retention responsibilities. Policies that address this issue should be formalized, as discussed below.

Table 2. Data retention standards of institutions with the top 10 medical schools by 2023 U.S. News & World Report ranking for research.

Rank	Institution	Data Retention Standards
1.	**Harvard University**	"Research records should be retained, generally, for a period of no fewer than seven (7) years after the end of a research project or activity; for specific guidance on how long to retain different types of records or to obtain assistance with handling records that have met their mandated retention periods, contact Archives and Records Management." "Essential research records are those research records integral to: • substantiating grant applications or demonstrating compliance with contractual terms, if sponsored research, • substantiating published research and patents, whether or not the research is sponsored, • substantiating research described in grant proposals and other funding requests, • or records considered for permanent preservation and access by the Archives and Records Management Program at the Center for the History of Medicine. Essential research records also include any research data or materials designated as essential by the Schools, consistent with the best practices for the relevant discipline" Harvard Medical School (n.d.).
2.	**Johns Hopkins University**	"Research Data should be stored using a method that permits a complete retrospective audit if necessary. Unless ethical/professional/local or funding body guidance requires otherwise, Research Data should be archived in a durable form and in a secure location that is immune to subsequent tampering and falsification for a minimum period of 5 years after the date of any publication upon which it is based. It is recommended good practice that evidence for research based on clinical samples or relating to public health should be retained as required by the funding agency, federal laws, or other policies of the University." "**Time Minimums for Research Data Archival**: • Expired Grants and Contracts: OMB (3 years after completion of the entire research project); Federal (follows OMB); Private (varies; see specific policy). • Clinical Trials (all relevant records): At least 2 years after the last approval of a marketing application or at least 2 years after formal discontinuation of clinical development of the investigational product or longer if required by contract, but in no instance less than 3 years after the completion of the Clinical Trial. • Patent Files and Data in Support of Patent: 17 years from the date of the patent application. • Research Data which Supported Enactment of a Federal, State or Local Law: Indefinite" Johns Hopkins University (2008).

(*Continued*)

Table 2. (Continued).

Rank	Institution	Data Retention Standards
3.	**University of Pennsylvania**	**Research Administration Records** • Scientific records should be retained for "up to 7 years after completion of research." • Sponsored research should be retained for "up to 7 years unless a longer period is required by sponsor contract." • Research involving investigational drugs should be retained for "up to 8 years after completion of research or possibly longer (see requirements of 21 C.F.R. 312.62)." • Research involving medical devices should be retained for "up to 7 years after completion of research" University of Pennsylvania (2011).
4.	**Columbia University**	"Research Data should be retained, generally, for 3 years after the end of a research project, or if longer, the period required by the applicable sponsor, with original Data retained wherever possible. Individual schools, departments or centers may establish a longer, but not a shorter, period of retention. A research project should be regarded as having ended after the latest of: (1) final reporting to the research sponsor; (2) final financial close-out of a sponsored research award; (3) final publication of research results; or (4) cessation of an academic or research project, regardless of whether its results are published." "Circumstances that may require longer retention periods: • Data that must be kept for as long as necessary to protect intellectual property and complete patenting and licensing procedures for inventions resulting from University Research; • If any charges regarding the research arise, such as allegations of scientific misconduct or improper charging of costs, data must be retained at least until such charges are fully resolved or for such other period as may be required by University policy or regulation; and • If a student is involved in the research, data must be retained for at least three years after the student's degree is awarded or it is clear that the student has abandoned work on the project" Columbia University (n.d.).

(*Continued*)

Table 2. (Continued).

Rank	Institution	Data Retention Standards
5. (tied)	**Duke University**	"Duke University expects research personnel to retain, via archives and/or placement in established repositories, research data and outputs for a minimum of six years after the final reporting or publication of a project. The research data archived must be original and sufficiently document the methods and accuracy of research data generation as well as the methods and accuracy of research data analysis and interpretation. When determining the scope of research data to be retained, the PI should consider the standards of their discipline and practices of their organizational unit. The retention period and the research elements that are retained may be altered by funders, publishers, contractual arrangements, compliance or regulatory bodies, and applicable laws, and the below conditions: • **Translation & Commercialization**: Retain research data and outputs until any patentable invention resulting from the work is protected by the filing of a patent application or, if a decision is made by the University not to file for patent protection, until rights to the invention are returned to the inventor. See the Policy on Inventions, Patents, and Technology Transfer for further information. • **Investigations, Allegations, Litigation**: Research data must be retained if relevant to an investigation or legal proceeding, including, but not limited to, research data under a legal preservation notice or deemed relevant to Research Misconduct proceedings. See the Policy on Misconduct in Research for further information on impact to retention period. • **Trainee-Engagement in Research**: If, in the course of advancement to a degree, a trainee serves as the sole PI in the design, conduct or reporting of research, the research data and outputs connected to that research must be retained according to the practices of the organizational unit" Duke University (2025).

(*Continued*)

Table 2. (Continued).

Rank	Institution	Data Retention Standards
5. (tied)	**Stanford University**	"Research data must be archived for a minimum of three years after the final project close-out, with original data retained wherever possible. In addition, any of the following circumstances may justify longer periods of retention: (1) Data must be kept for as long as may be necessary to protect any intellectual property resulting from the work. (2) If any charges regarding the research arise, such as allegations of scientific misconduct or conflict of interest, data must be retained until such charges are fully resolved. (3) If a student is involved, data must be retained at least until the degree is awarded or it is clear that the student has abandoned the work. Beyond the period of retention specified here, the destruction of the research record is at the discretion of the PI and his or her department or laboratory." "When individuals involved in research projects at Stanford leave the University, they may take copies of research data for projects on which they have worked. Original data, however, must be retained at Stanford by the PI. If a PI leaves Stanford, and a project is to be moved to another institution, ownership of the data may be transferred with the approval of the Vice Provost and Dean of Research, and with written agreement from the PI's new institution that guarantees: 1) its acceptance of custodial responsibilities for the data, and 2) Stanford access to the data, should that become necessary" Stanford University (1997).
5. (tied)	**University of California, San Francisco**	The applicable retention period depends on the specific type of records, but generally, records must be retained "for 6 years after the expiration/termination of the sponsored activities; resolution of any litigation, claim, or audit; or the period stated in the award document – whichever is longer" University of California (2023).

(*Continued*)

Table 2. (Continued).

Rank	Institution	Data Retention Standards
5. (tied)	**Vanderbilt University**	"Vanderbilt requires that all Research Data relevant to sponsored research be retained for a minimum of three years following the completion of the sponsored project/expiration of the agreement with the funding source. In addition, any Research Data derived from projects involving post-docs or students must be retained until completion of the student's program or the post-docs' employment at Vanderbilt or until it becomes clear the student or post-doc has abandoned the project from which the Research Data is derived. Finally, any data relevant to a disclosure to the Office of Technology Transfer, in particular written records of early ideas or experiments, should be retained to support possible patent prosecution (up to 20 years from the initial disclosure to the Office of Technology Transfer). Beyond these retention periods, the retention or destruction of the research record is at the discretion of the PI." "Research Data includes any records that are necessary for the reconstruction and evaluation of reported results of research and the events and processes leading to those results, regardless of the form of media on which they may be recorded." "Ownership of Research Data belonging to Vanderbilt may be transferred to another institution when a PI moves to that institution with the approval of the Vice Provost for Research or his/her designee and with written agreement from the PI's new institution that guarantees (1) its acceptance of custodial responsibilities for the data and (2) that Vanderbilt may access the data should it become necessary" Vanderbilt University (2021).

(*Continued*)

Table 2. (Continued).

Rank	Institution	Data Retention Standards
10. (tied)	**Cornell University**	"**Ithaca-based faculty – collection and retention of data**: Research data is retained for a minimum of three years after the final project closeout. If the primary data and images are used in a subsequent publication, or the initial publication is citied in a subsequent publication or grant application by the faculty member, the data and images must be available for an additional six years. If specific software or code is required for the University to interpret the data, this software or code should also be deposited with the data, as long as license agreements permit. Data is retained unless the Vice President for Research and Innovation approves retaining summaries or other secondary data based on compelling justification in special cases. Different periods of retention are required in some circumstances . . . . Beyond any required period of retention, the destruction of research data is at the discretion of the principal investigator (PI) within limitations imposed by college or department expectations, the needs of collaborators or students, or the norms of their field. Records will normally be retained in the unit where they are produced. Research records must be retained in university facilities, or facilities mandated by the sponsor, or other facilities commonly used for the purpose by peers in the same field of study so long as the data is maintained with appropriate oversight and reasonable means are available for access by Cornell faculty and administrative personnel when needed." "**Weill Cornell Medicine (WCM) faculty – collection and retention of data**: Faculty are responsible for creating, abiding by, and funding a data management plan which satisfies all the expectations in this policy. In this plan, faculty must specify where they will deposit the data at the close out of the research . . . . Faculty must ensure that primary data and supporting images are available for the University for at least six years after publication. If the primary data and images are used in a subsequent publication, or the initial publication is citied in a subsequent publication or grant application by the faculty member, the data and images must be available for an additional six years. If specific software or code is required for the University to interpret the data, this software or code should also be deposited with the data, as long as license agreements permit. Beyond any required period of retention, the destruction of research data is at the discretion of the principal investigator (PI) within limitations imposed by college or department expectations, the needs of collaborators or students, or the norms of their field" Cornell (2022).

(*Continued*)

Table 2. (Continued).

Rank	Institution	Data Retention Standards
10. (tied)	**New York University**	The applicable retention period depends on the specific type of records, but generally, "[the l]onger of 3 years from submission of final expenditure report; (ii) 5 years after final reporting or publication of a project or (iii) period required by research sponsor." "Some common research circumstances where the University may require a longer retention period are: • if any intellectual property resulting from the work has been or is likely to be commercialized by NYU, Research Data must be kept for as long as may be necessary to protect it; • if any charge, audit, claim or litigation regarding the research arises, such as allegations of scientific misconduct or conflict of interest, data must be retained for seven (7) years after the completion of the proceeding adjudicating such charge, audit, claim or litigation is fully resolved and final action is taken; and • if a student is involved, data must be retained at least until the degree is awarded or it is clear that the student has abandoned the work" New York University (2010).

(*Continued*)

Table 2. (Continued).

Rank	Institution	Data Retention Standards
10. (tied)	**Yale University**	"Generally, Research Data and materials that are commonly accepted in the scientific community as necessary to validate research findings must be retained by Yale researchers for three (3) years after publication of the findings or all required final reports (e.g., progress and financial) for the project have been submitted to the sponsor unless a longer period is provisioned. Longer retention periods may be required by Yale Policy, publishers, sponsors, or applicable laws and regulatory requirements … Yale Researchers are responsible for consulting these requirements and Research Data must be retained in accordance with those more stringent requirements." Research data is defined as "[t]he recorded factual information associated with the Research, including, but not limited to, all records necessary for the reconstruction and evaluation of the results of Research, regardless of the form or medium on which the material is recorded (such as lab notebooks, photos, digital images, Research Data files, Research Data processing or computer programs (software), statistical records, etc.). Research Data does not include books, articles, papers, or other scholarly writings that are published or publicly presented; drafts of such scholarly writings; plans for future Research; peer reviews; or communications with colleagues." "When Yale Researchers who are PIs or co-PIs on Research at Yale leave the University, they may generally take copies of Research Data for projects on which they have worked … . Other Co-Investigators may take copies of Research Data for projects (or the portions of projects) on which they have worked only with the permission of the PI or PIs or, if the PI or PIs and the individual leaving cannot reach agreement, with the permission of the Office of the Provost. In all cases, the original Research Data and materials must be retained at Yale" Yale University (2024).

Importance of data retention

Data retention in research misconduct proceedings

As noted above, institutions that apply for and receive PHS support for biomedical research must adhere to the research misconduct regulations under 42 C.F.R Part 93 ("Part 93"), which are promulgated by ORI. In a final rule issued on September 17, 2024, ORI made updates to Part 93 for the first time since it was codified almost 20 years ago (ORI, HHS 2024). These amendments, which include changes to the data retention requirement under Part 93, will become effective on 1 January 2026. When an institution receives an allegation of research misconduct, defined under Part 93 as fabrication, falsification, or plagiarism in proposing, performing, or reviewing research, or in reporting research results, the institution must initiate a research misconduct proceeding that follows the procedures required by Part 93. Data retention is vital to these federal research integrity obligations, as research records must be available to corroborate research results and to determine whether research misconduct has taken place. When relevant original data are lacking but a research misconduct allegation has been raised requiring access to those data, the institution's determinations in research misconduct proceedings are compromised.

The data retention requirement imposed by Part 93 is not explicitly set forth in the regulations, except for a required period of record retention after a research misconduct proceeding has been completed. However, under 42 C.F.R. § 93.104(a), research misconduct proceedings are required only for research that occurred within 6 years of the date HHS or a cognizant institution receives an allegation of research misconduct, subject to certain exceptions. This creates an implicit minimum 6-year data retention period (measured from the date of publication, and not the date the research was conducted). The exception that is invoked most frequently to extend this period is the "subsequent use exception" (42 C.F.R. § 93.104(b)), which "renews" an instance of alleged research misconduct occurring before the 6-year limitation if the respondent cites, reproduces, or otherwise uses the research for his or her potential benefit. Beginning on 1 January 2026, the subsequent use exception will be narrower in scope, applying only to the respondent's "use of, republication of, or citation to *the portion(s)* of the research record … alleged to have been fabricated, falsified, or plagiarized, *for the benefit of the respondent*" (ORI, HHS 2024, 76297). Historically, this exception in particular has been subject to interpretation, and ORI has previously indicated that it interprets the subsequent use exception broadly to include instances even in which an author within the past 6 years of the filing of a complaint of research misconduct has cited his or her research paper in his or her curriculum vitae, grant biographical sketch, or in another, later publication (NIH 2022; ORI, HHS 2023). Even though the scope of the

subsequent use exception soon will be narrowed, whether a "portion" of the research record has been reused, republished, or cited may be difficult to determine. Therefore, absent further guidance from ORI, all research data generally should be maintained for longer than 6 years, despite this policy change – namely, for as long as a research misconduct proceeding could be initiated, either based on the timing of the research itself or the author's subsequent re-citation of those research findings. Indeed, even 6-year retention periods do not accommodate the need for access to primary data if an author has re-cited his or her previous publication, however old it is. Moreover, many academic institutions do not have in their research misconduct policies any specific statute of limitations on research misconduct proceedings, and will take jurisdiction even of allegations of possible research misconduct in publications of distant vintage. This effectively requires researchers at those institutions to preserve their essential research records in perpetuity.

Further under 42 C.F.R. § 93.318, an institution has a continuing obligation to maintain adequate records for research misconduct proceedings. Specifically, institutions must maintain the institutional record and all sequestered evidence for 7 years after completion of a research misconduct proceeding. All of these research misconduct considerations – and in particular, the 6-year statute of limitations and the "subsequent use" exception – effectively impel institutions to require lengthy periods for data retention.

Perhaps in acknowledgment of the challenges that institutions face with respect to reaching a research misconduct finding in the absence of data, Part 93 allows an institution to consider missing research records as evidence against the respondent in certain instances. Specifically, under the previous version of Part 93 that is still in effect until 1 January 2026, the regulations provide that a respondent's inability to provide original research records to verify published findings can be regarded as evidence of the respondent's misconduct. This provision is understandable, because without it, researchers accused of misconduct would be able, once informed of an allegation, simply to destroy the relevant data, thus frustrating any attempt to investigate the concerns about their published research. Although this provision was omitted in the revision of Part 93, the principle underlying it survives in some institutional policies. Moreover, ORI has quietly suggested that this omission was an error that may be remedied in the future by issuance of new guidance on the issue of how failure to retain data should be interpreted in research misconduct proceedings. In any event, the revised Part 93 contains a new provision that a respondent's failure to provide research records documenting the questioned research is evidence of research misconduct if "the respondent claims to possess the records but refuses to provide them upon request" (ORI, HHS 2024 76,298). For all of these reasons, data retention to verify research findings and resolve any data integrity concerns

is essential, and the period of retention is effectively as long as the researcher seeks to preserve the ability to defend his or her published research in research misconduct proceedings.

Scholarly importance of data retention

In addition to being a regulatory compliance issue, the proper retention of research records is also a professionalism issue. To support the scientific rigor and integrity of research, and to support the conduct of future research, researchers should retain research records so that data may be shared and so that results that cannot be replicated by other researchers can be compared to past research. Moreover, without retaining the original research records that support a given publication, an author may not be able to validate and defend the underlying research if challenged by another researcher in the field or a media outlet.

As a matter of practice, researchers typically attest that they are agreeing to be held accountable according to these standards (ICMJE 2025). Under the International Committee of Medical Journal Editors' ("ICMJE") recommendations relating to the publication of scholarly work in medical journals, ICMJE encourages investigators to preserve "primary data" for at least 10 years and identifies the corresponding author as the author responsible for cooperating with journals' requests for data in the event that questions arise after publication (ICMJE). However, despite the ICMJE recommendations, corresponding authors often rely on first authors to store original data and typically argue during research misconduct proceedings that they are not primarily responsible for having retained the records of experiments performed by their junior colleagues – even if those junior colleagues worked under their supervision. Ultimate responsibility for data retention lies with the institution, although many institutions, as indicated by Table 2, do not currently require data retention for at least 10 years. However, in the absence of uniform regulatory requirements, the Cooperation & Liaison between Universities & Editors, an international working group focused on improving communication between journals and institutions, sets a reasonable professional standard that institutions should ensure that essential research data are retained for at least 10 years (Wager and Kleinert 2021). Institutions should set clear expectations for data retention, including who is responsible for retaining data and the appropriate manner in which data should be retained, and develop the infrastructure needed to support the preservation of data.

Recommendations

Institutions: Defining essential research records and setting retention records

We recommend that researchers retain "essential research records" and that those "essential records" be defined in an institutional policy. Essential records for this purpose may be defined in ways that are discipline-specific, and may not uniformly include, for example, all original samples, cell lines, specimens, and raw experimental results, but would include any research records that would be critical for substantiating published research, prior and active grant or contract applications, submissions to regulatory agencies, and any sponsored research or research supporting active patents. As the research records that would be considered "essential" may differ depending on the specific area of research and the methodology used, for an institution to form an exhaustive list of "essential research records" is not possible, and institutional policies could allow some local determinations by departments or schools as to what records are "essential." As an example, if a Western blot were used to support results in a research publication, then the original images acquired from the Western blots, but not the cell lines or lysates used in the experiments, should be retained as an "essential research record." As another example, if a publication relies on a histology image, then ideally the original glass microscope slide with the pathology specimen should be retained, or at the very least the original image that was acquired from the pathology slide. When possible, researchers should maintain the original data (*i.e.*, the source data), particularly as most data generated today can be stored electronically. Many data integrity issues cannot be resolved without the original data. Ultimately, the primary goals for data retention are to support data sharing within the scientific community and to ensure that if the integrity of the research is questioned, the research records will be sufficiently fulsome and accurately recorded to support the reported research results.

The NIH Policy for Data Management and Sharing, released on January 25, 2023, defines "scientific data" as:

> The recorded factual material commonly accepted in the scientific community as of sufficient quality to validate and replicate research findings, regardless of whether the data are used to support scholarly publications. Scientific data do not include laboratory notebooks, preliminary analyses, completed case report forms, drafts of scientific papers, plans for future research, peer reviews, communications with colleagues, or physical objects, such as laboratory specimens.

This definition of "scientific data" is a useful starting point for developing a universal definition applicable not only to data sharing, but also to data retention for regulatory compliance, research integrity, and professionalism

purposes. Institutional data retention policies should be consistent with expectations of retention for data sharing purposes; these policies should be reviewed and updated periodically and evolve as public policy pushes researchers and research institutions to move toward greater transparency and openness in data sharing.

In regard to the length of required retention of research data, we further recommend that institutions have in place policies that provide clear requirements rather than simply referring to the funding agency standards and applicable regulations. Specifically, we recommend that institutions establish a baseline policy requiring all researchers to retain data for *at least* 10 years after completion of a given research project, in line with ICMJE and CLUE recommendations, understanding that the competitive renewal of a grant funding a research project is considered a continuation of the original project, not a new one. Given the subsequent use exception of Part 93, we recommend resetting this 10-year minimum retention period after each citation or republication of the research record by the researcher. After such a 10-year minimum retention period, institutional policies could mandate a review process by which the institutional academic leadership works with the researcher to determine whether and which data should be retained for a longer period (*e.g.*, to support an active patent, to preserve data that are referenced in a later-in-time publication or grant application, or to comply with funding requirements). A minimum 10-year timeframe, together with a formal retention review process at the end of the 10-year period, will ensure that data are retained for a period sufficiently long to comply with the large majority of existing data retention standards under laws and regulations, while also minimizing confusion among researchers by applying a uniform standard for all research records. Institutional policies should, however, also have specific provisions for a longer, perhaps indefinite, period of retention for data that are so significant as to underlie patented inventions and/or be of historical or continuing scientific interest. Policies must also, of course, require longer data retention periods to comply with any specific provisions in sponsored research or collaborative research agreements, as well as to allow defense of patented or other intellectual property. Research records that relate to pediatric care also require special consideration, so that they are preserved for at least 3 years after a subject's age of majority, in order to anticipate any possible tort claims. Notwithstanding these obligations to retain data, institutions should nevertheless, in our judgment, encourage their researchers to retain data for as long as possible, given the unpredictability of when original data may be needed.

In addition, as researchers often move between institutions, and in so doing may neglect to consider the data rights and data ownership of the institutions from which they are departing, all institutions should have clear expectations and policies governing data retention and sharing so that essential research records are not lost or purloined when a researcher leaves one institution to join another. For example, institutional policies should

clarify how investigators may seek and gain permission to take with them copies of original data and where the original data should be stored after the researcher's departure. Before the researcher's departure, a plan or agreement should be developed with the researcher, and optimally shared or executed with the institution that the researcher is joining, to ensure that essential research records are stored at a designated physical or electronic location, with a designated custodian. In addition, we recommend that institutional policies include requirements for documenting the transfer of research records when a researcher moves to a new institution, signed by designated persons at both the prior and new institutions, to ensure adequate record-keeping of transferred research records to which both institutions may refer if such research records need to be assessed for any purpose, including the use and custody rights of the institutions involved.

Given the vital role that institutions play in the retention of biomedical research data, we also recommend that institutions have a designated person responsible for research data stewardship. Not only would this individual be responsible for developing and implementing institutional policies on data retention, but also for fielding questions about data retention from researchers at the institution and for updating policies as public policy on data retention and sharing evolve over time. While we acknowledge that some institutions may not have the resources to dedicate a single individual responsible for institution-wide oversight of data retention in biomedical research, all institutions must understand their duty to ensure that their researchers comply with applicable data retention standards.

Office of Science and Technology Policy

Given the wide variability in data retention standards across federal agencies that provide oversight and/or funding for research, it can be difficult for researchers and institutions to identify which data retention standards may apply to a given research project. For example, some of the data retention time periods required by federal funding agencies are shorter than 6 years, even when some such funded research would fall within the scope of the 6-year statute of limitations for research misconduct. We therefore suggest that OSTP expand its recent initiatives regarding data sharing to seek uniformity and harmonization of data retention requirements across all governmental stakeholders that fund or otherwise have oversight of research, and in doing so establish data retention standards that are concrete rather than vague expectations that are difficult, or impossible, to meet in practice. This includes, for example, endorsing a streamlined, easy-to-follow data retention framework for determining what data must be retained and for what time periods, which other federal agencies can adopt and institutions can implement readily at the local level.

Conclusion

Data retention is critical to maintaining trust in scientific research. Robust data retention practices are of obvious importance to data sharing and reproducibility initiatives, and necessary in many instances for compliance with existing federal obligations and institutional policies, including those pertaining to research misconduct. As discussed above, existing data retention standards vary widely across federal agencies and under applicable laws and regulations, and institutional policies often do not provide researchers with clarity regarding data retention responsibilities. We recommend that institutions invest in systems and infrastructure to support the institutional retention of primary data and establish uniform data retention policies that apply to all disciplines regardless of funding source and that address (1) what essential data should be retained, (2) the minimum retention period for all research, (3) special circumstances that may require longer retention periods, and (4) clear procedures for data retention or transfer when a researcher moves to a new institution. In addition, we recommend that institutions implement in their data retention policies a built-in 10-year (or other defined) checkpoint at which researchers must determine if any remaining data retention obligations exist, and also institute a designated person responsible for research data stewardship, who can develop and implement such data retention policies.

As described above, OSTP has the opportunity to play a critical role in the development of more consistent standards on these issues. If OSTP were to address data retention issues rigorously and systematically, the research community could coalesce around common-sense standards that apply widely across many different funding and oversight agencies, bringing a welcome uniformity to our current chaotic standards and practices.

Disclosure statement

Mark Barnes, J.D. LL.M., Minal Caron, J.D., and Carolyn Lye, M.D., J.D. are attorneys at Ropes and Gray LLP, an international law firm that represents many entities in health care and academic sectors, including advising academic medical centers, institutions of higher education, pharmaceutical and medical device companies, and other research institutions on research misconduct matters at all stages of review. Mark Barnes and Minal Caron regularly serve as acting research integrity officers for universities andacademic medical centers. Barbara Bierer, M.D. formerly served as Senior Vice President of Research and the Research Integrity Officer at the Brigham and Women's Hospital and often serves on peer review committees conducting research misconduct proceedings on behalf of research institutions.

Funding

The author(s) reported there is no funding associated with the work featured in this article.

ORCID

Carolyn T. Lye http://orcid.org/0000-0002-7869-6242
Minal M. Caron http://orcid.org/0000-0001-6290-3748
Lauren Walsh http://orcid.org/0009-0008-7225-0486
Barbara E. Bierer http://orcid.org/0000-0001-6448-8170
Mark Barnes http://orcid.org/0000-0001-8194-4668

References

Cornell University Policy Office. 2022. "Policy 4.21: Research Data Retention." https://policy.cornell.edu/sites/default/files/vol4_21.pdf .

ICMJE (International Committee of Medical Journal Editors). 2025. *Recommendations for the Conduct, Reporting, Editing, and Publication of Scholarly Work in Medical Journals*. https://www.icmje.org/icmje-recommendations.pdf .

Johns Hopkins University. 2008. "Policy on Access and Retention of Research Data and Materials." https://dataservices.library.jhu.edu/wp-content/uploads/sites/41/2019/02/JHUDataRetentionPolicy2008_WithAppendices.pdf .

Longwood Research Data Management, Harvard Medical School. n.d. "Data Retention." https://datamanagement.hms.harvard.edu/store-evaluate/data-retention .

New York University. 2010. "Retention of and Access to Research Data." https://www.nyu.edu/about/policies-guidelines-compliance/policies-and-guidelines/retention-of-and-access-to-research-data.html .

NIH. 2017. "NIH Intramural Records Retention Schedule Items." https://oma.od.nih.gov/DMS/Documents/Records/Record%20Schedules/NIH%20Intramural%20Research%20Records%20Schedule_12.18.2018.pdf .

NIH. 2023a. *Conduct of Research in the Intramural Research Program at NIH*. 8th ed.

NIH. 2023b. *Request for Information on the NIH Plan to Enhance Public Access to the Results of NIH-Supported Research*.

NIH. 2024. *NIH Grants Policy Statement*.

NIH Grants. 2022. *Research Misconduct and Detrimental Research Practices – Overview and Case Studies*.

NSF. 2024. *Proposal and Award Policies and Procedures Guide (NSF 24-1)*. https://nsf-gov-resources.nsf.gov/files/nsf24_1.pdf

Office of Research Integrity, HHS. 2023. "Public Health Service Policies on Research Misconduct: Notice of Proposed Rulemaking." *Federal Register* 88 (193): 69586–69604.

Office of Research Integrity, HHS. 2024. "Public Health Service Policies on Research Misconduct." *Federal Register* 89 (180): 76280–76309.

Office of the Executive Vice President for Research, Columbia University. n.d. Data Retention. https://research.columbia.edu/data-retention-0 .

Office of the President, University of California. 2023. "Administrative Records Relating to Research." https://www.ucop.edu/research-policy-analysis-coordination/policies-guidance/record-retention/tables.html .

Office of the Provost, Duke University. 2025. "5.5 Research Data Retention." https://policies.provost.duke.edu/docs/chapter-5-research-data#55-research-data-retention .

Stanford University. 1997. "Retention of and Access to Research Data." https://doresearch.stanford.edu/policies/research-policy-handbook/conduct-research/retention-and-access-research-data .

University Archives & Record Center, University of Pennsylvania. 2011. "University Records Center Research Administration Records." https://archives.upenn.edu/records-center/resources/retention-schedules/research-admin/ .

Vanderbilt University. 2021. "Research Data Governance Framework." https://www.vanderbilt.edu/datagovernance/wp-content/uploads/sites/352/2021/07/Research-Data-Governance-Framework-Final-June-13-2021.pdf .

Wager, E., and S. Kleinert. 2021. "Cooperation & Liaison Between Universities & Editors (CLUE): Recommendations on Best Practice." *Research Integrity and Peer Review* 6 (1): 6. https://doi.org/10.1186/s41073-021-00109-3 .

White House. 2023. *FACT SHEET*: *Biden-Harris Administration Announces New Actions to Advance Open and Equitable Research (2023)*.

Yale University. 2024. "6001 Research Data & Materials Policy." https://your.yale.edu/policies-procedures/policies/6001-research-data-materials-policy.