

Guidelines for retention and disposal of data

This document offers guidance on data retention and disposal to address the common question of what to do with all of the accumulated research data collected during all or part of a career. Unfortunately, the rules governing data retention are subject to numerous conflicting regulations, definitions and requirements and individual decisions are often based on a combination of circumstances. Considerations about the types of issues that arise and how to address them are provided below.

How long should I keep original data?

Original research data are the foundation for all important research conclusions and are required by federal regulation to be preserved for a period of time to support the integrity of those conclusions and to provide sources for ongoing research endeavors. Original, or primary, data refers to data generated in its original format. However this standard is tempered by the nature of the original data which varies widely across disciplines. For example, data intensive methodologies, such as imaging, may be prohibitively large for permanent storage of primary data but in practice storage of samples or compressed formats may be standard to the field. To further complicate matters, different disciplines, and journals representing those disciplines, may have their own requirements, informal practices and/or expectations.

The [JHU Policy on Access and Retention of Research Data Materials](#) defines research data as occurring in any form or media and includes biological specimens, environmental samples, laboratory notebooks, photographs, films, digital images, and so on. **It requires that faculty must keep original research data at least five years after any publication on which it is based.** Although NIH generally has a 3 year record retention (financial, IRB and data records) from the official close out of the award (unless there is a misconduct or audit proceeding involved with the grant), JHU faculty are subject to the longer 5 year requirement (Appendix IV in above policy). But...the JHU policy is superseded by funding or regulatory agency requirements for maintaining records when they are longer. FDA sponsored investigational drug and device research, for example, has a separate set of requirements. For clinical data covered under HIPAA, adults have the right to an accounting of the data used for research through 7 years; for minors, the right extends until they are age 23. There are complexities even within these regulations. Note that for HIPAA covered data, the retention rule is based on either when the last participant has completed the study or the date of the last disclosure of identifiable health data, whichever comes later (hopkinsmedicine.org/institutionalreviewboard) but not from the date of publication as is the case with the JHU retention rule.

The exceptions provided here are not exhaustive, and subject to change over time, so it is incumbent on you to understand the requirements in your own particular research landscape.

For research involving human subjects, data retention or disposal should proceed as stated in the approved IRB protocol – as long as that plan is in accordance with the data retention timelines of the sponsor, regulators and JHU. While you may believe that the IRB prefers records to be destroyed sooner rather than later, the IRB's main concern is in the retention of

identifiers and the resultant loss of confidentiality should that information be inadvertently disclosed. If you wish to retain data but want to make sure that it is properly de-identified, contact the JHSPH IRB Office for instructions on how to proceed.

The federal Office of Research Integrity has a useful tutorial that discusses many issues related to data at all stages of the research cycle, including retention:

<https://ori.hhs.gov/education/products/rcradmin/topics/data/open.shtml>

Although JHU “owns” all of the data and specimens that are generated while you are a faculty member, investigators are stewards of the data. Technically, institutions are responsible for retaining records generated from PHS grants beyond the life of the award within their regulatory requirements. However, in reality costs are often passed to investigators as alternate sources for funding of physical storage of paper records or specimens are unclear. As digital data and digitized record-keeping become the norm, the cost burden will lessen.

In reality, faculty often want to retain data beyond the regulatory and/or JHU requirements. There are good reasons for doing so. Data are expensive to collect and often unique so there can be value in revisiting them at some unanticipated point in the future or passing the data on to others. Moreover, in the event that data in a publication are challenged or you are accused of research misconduct, it is far easier to support your work if you have the original records than if you have discarded them. Disputed manuscripts can be retracted if original data cannot be provided, regardless of the amount of time that has elapsed. This has resulted in the ambiguous but excellent advice to keep your data “a good long time” provided in a thoughtful blog post ([How long should you keep data?](#)).

Note that data generated by trainees (students and postdocs) who graduate, leave the school prior to their expected tenure, or are terminated also belong to the university. Decisions on whether to allow the trainee to retain copies of the data or materials are made at the principal investigator or departmental level.

How long should you save your original data? For non-HIPAA covered data, at least 5 years from the last publication but preferably for as long as you can afford or have access to physical or digital storage, properly deidentified, and in a format closest to the original as practicable.

What do I do with data as I wind down my scientific career or no longer need to store it in its original format?

Faculty members who are retiring or nearing retirement and/or have data from projects that were completed beyond the JHU or other regulatory retention requirement periods should consider which of the following makes the most sense:

1. Transferring original data to another JHU faculty member
2. Storing the data by digital means (e.g., scanning paper records) and maintaining on a personal computer, in secure cloud storage, and/or transferring those files to another faculty member

3. Physically storing records
4. Arranging for data disposal

Before embarking on any of these options, particularly data disposal, you should contact your collaborators to make sure they are supportive of your approach. Suggestions on how to implement these are provided below.

For transferring or storing, considerations include whether or not:

- the existing data format (e.g., 5¼ floppies) allows retrieval and transfer to current media
- there is a reasonable chance that the data will remain unique enough for others to want to use them (and go through the hassle of converting into newer technology formats)
- there are any on-going disputes about the validity of the findings generated by the data – in which case you are obligated to retain them
- the data currently include Protected Health Information (PHI) under HIPAA or personally identifiable information (PII) and, if so, whether this conforms to the IRB approval of storage of those data. If so, consideration of appropriate security measures to protect those data is paramount. Refer back to the original IRB documents to better understand the requirement and/or contact the IRB Office.

Note that data considerations are only one element in closing out research programs and laboratories when faculty retire or leave the School for other reasons. Here is a useful resource and checklist: [Guidance for departing investigators.](#)

How do I electronically archive data?

Data archiving is becoming more frequent for both required and volitional data sharing but also for retaining research data long past the life of the grant or active analyses to allow access for future investigators. In general, data archiving includes processes that enable access to the data by third parties. NIH has several data repositories, including *figshare* (general data), *dbGAP* (genomic data) and *Trace Archives* (sequencing data). This is a rapidly evolving service domain and there are a number of options offered by other academic institutions (e.g., *dataverse*, *ICPSR*) and consortiums (e.g., *Vivli*), with more to come.

JHU offers data archiving services through its Data Management Services, part of the Sheridan Library: <https://dataservices.library.jhu.edu/archiving/>

This service is generally intended for data with active access and reuse. However, faculty may want to consider this service if the data might be a valuable resource for other faculty members or trainees going forward. It is not typically used for conversion of older media (such as floppy disks), but the library does have access to equipment for these conversions if needed.

- Offers free archiving for projects under 1TB
- Accepts data from any research discipline and file format
- Provides each dataset with a permanent citation and DOI, facilitating both attribution for authors and linkage to research publications

- Preserves research data through regularly checking file integrity and retaining multiple copies

What about storing physical records?

First, here is what not to do: bring home boxes of file folders with participant names or other personally identifying information on them and store them in your attic or basement! Instead, determine whether your department has space to store your physical records (paper files, CDs, hard drives, etc.), or perhaps even a subset of your materials that are most sensitive or vulnerable. In the likely event that they do not, *Iron Mountain* (www.ironmountain.com/) is already contracted to the School and provides fee-based storage/archiving of paper files, data stored on electronic media, etc. They also offer scanning services to convert paper files to electronic records. **Never** store paper or electronic records that have PHI or PII in them in your home or on your personal laptop.

For biological specimens, **the Johns Hopkins Biological Repository (JHBR)** is a JHSPH service center for processing, storing, and retrieving samples. The JHBR is a large, state of the art facility that houses specimens in liquid nitrogen cryogenic units.

www.jhsph.edu/research/labs/biorepository/index.html

The **JHU Chesney Medical Archives** collects and preserves records and cultural material that have historical significance for the institution and may have value for use as primary resources in research and education. Chesney staff work with faculty members to assess materials and determine whether they have relevance to the goals of the Archives. Basic cataloguing and preservation of paper records can be done at no or minimal cost; digital data that would require conversion to a preservation format is possible. Note that the decision as to whether or not to archive a faculty member's materials is at the discretion of the [ChesneyArchives](#).

How do I dispose of data?

Once your data have been kept beyond the time limits described in the first section of this document and your collaborators have been appropriately informed, you can opt to dispose of it. There are various safe methods of disposal depending on the data format.

Paper: shredding and disposal services are available from companies such as *Shred-it* and there are disposal boxes located through the School. *Iron Mountain* provides boxes for long-term, fee-based storage at its facilities: <https://www.ironmountain.com/>. Should you wish to engage either company with your specific project, particularly if it is high volume, contact your departmental administrator.

Electronic media and hardware. There are specific ways to effectively destroy CD/DVDs, hard drives, flash drives, etc. The JHSPH Information Technology office can assist in proper disposal of media and hardware through recycling at no cost to faculty members.

<https://www.jhsph.edu/sustainability/electronic-hardware-recycling-faqs>

The program:

- Ensures proper data destruction and therefore prevents data leakage.
- Processes equipment with a goal of converting it first to a usable commodity.
- Operates with a zero landfill goal. No equipment is dumped overseas.

Digital data. You can use a variety of safe “erase” methods for data files on your personal computer or laptop. If you have software for encrypting files and file folders, (e.g., [VeraCrypt](#)), you can encrypt those files and then delete them rendering the files useless even if recovered. Other suggestions include:

PCs: www.pcworld.com/article/2464163
www.pcworld.com/article/209418

Macs: www.groovypost.com/howto/securely-delete-files-mac/

Need help? Contact JHSPH IT services for additional assistance.