



Cornell University Policy Office

policy.cornell.edu

Policy 4.21 Research Data Retention

Responsible Executive: Vice President for Research and Innovation; Weill Cornell Medicine Senior Associate Dean for Research

Responsible Office(s): Research Integrity and Assurance; Weill Cornell Medicine Research Integrity and Compliance

Originally Issued: August 7, 2020

Current Version Approved: October 22, 2021

Last Updated: February 9, 2022

Table of Contents

1. Policy Statement and Requirements	2
1.1 Policy Statement	2
1.2 Scope	2
1.3 Policy Requirements.....	2
1.4 Collection and Retention of Information.....	8
1.5 Compliance	8
1.6 Resources.....	8
2. To Whom This Policy Applies	11
2.1 For Cornell community based out of:	11
2.2 Who should read this policy:	11
3. Definitions	11
4. Policy Administration	13
5. Revision History	13

1. Policy Statement and Requirements

1.1 Policy Statement

Accurate and detailed records of research data are an essential component of any research project. This policy defines the shared responsibilities of Cornell University (including Weill Cornell Medicine) and Cornell researchers in collecting, retaining, securing, accessing, publishing, and sharing research data.

1.2 Scope

The university is committed to the principles of open science and research integrity. The integrity of the record of research conducted requires that research data be preserved in sufficient detail and for an adequate period of time, to comply with sponsor requirements, federal, state, and local regulations, and inquiries governing the research.

Cornell is committed to maintaining the highest standards of research and complying with sponsor requirements and with federal and state regulations related to research integrity. This policy is intended to ensure a common understanding of the responsibilities of researchers with respect to maintaining research data.

This policy addresses retention of research data only. Retention and protection of all other Cornell Institutional data are addressed by [University Policy 4.7, Retention of University Records](#), and [University Policy 5.10, Information Security](#).

1.3 Policy Requirements

1.3.1. High-quality research and academic integrity: Sound stewardship of research data, which includes maintaining the highest standards in the generation, use, management, and retention of research data, is a fundamental requirement of ensuring high-quality research and academic integrity. In addition, public access to research data, within legal, ethical, and regulatory constraints, is an enabling factor in Cornell's mission to discover, preserve, and disseminate knowledge.

In collaborative research, whether with other Cornell researchers or with researchers at other institutions, lack of clarity concerning rights to data and means of sharing data can lead to disputes or allegations of misconduct. This policy does not address specific means of data sharing.

1.3.2. Reproducibility of research results: Reproducibility, which is the ability to verify research findings by other members of the scientific community or by using other methods, is essential to the advancement of science. This ability requires access to relevant research data, materials, documents, protocols, methods, and procedures.

1.3.3. Stewardship of private, confidential, or proprietary data: Cornell is committed to safeguarding the privacy and security of confidential, restricted, identifiable, or otherwise sensitive data entrusted to its care. The loss or release of such data can lead to significant harms such as privacy violations, identity theft, and financial liability for the university and, in some cases, individual liability for the person who was the custodian of the data. The university complies with external regulations that govern research data, such as the Health Insurance Portability and Accountability Act (HIPAA), Family Education Rights and Privacy Act (FERPA), and General Data Protection Regulation (GDPR).

Note for Ithaca-based campuses: While Cornell University is a single hybrid-covered entity under HIPAA, almost all Ithaca-based units, particularly those likely to be engaged in research, are not covered components within that entity. Cornell's Ithaca campus is not a "Covered Entity" under HIPAA and researchers therefore may not transmit, use, or store protected health information (PHI) unless it is in the form of a limited data set (LDS). Contact the Office of Research Integrity and Assurance for more information.

The principal investigator (PI) is the custodian of their research data and is responsible to Cornell for the proper use, access, security, and control of any research data under their management or supervision, including the use of data in scholarly publications and presentations.

To ensure the integrity of the research process and to comply with sponsor and/or federal regulations, the university must retain research data in sufficient detail and for an adequate period of time to enable appropriate responses to questions about accuracy, authenticity, primacy, and compliance with laws and regulations governing the research.

For research involving human participants, the Institutional Review Board (IRB), Cornell Compliance Office, and the Vice President for Information Technology and Chief Information Officer have the authority to require safeguards appropriate for the protection of those participants.

1.3.4. University ownership of research data: Cornell (including Weill Cornell Medicine) asserts ownership of research data and related property rights arising from the activities of its researchers and others who use university resources, including those provided through an externally funded grant, contract, or other type of award or gift to the university.

1.3.5. University responsibilities: Cornell's responsibilities for research data, and the scientific record arising from it, are based on federal regulations (such as [2 CFR § 200.315 \(e\) \(3\)](#)), [University Policy 1.5, Inventions and Related Property Rights](#), and other university policies, sponsor requirements, and precedent. Responsibilities include, but are not limited to:

- Protecting the rights of faculty, staff, and students, including, but not limited to, their rights to access data from research in which they participated
- Complying with the terms of sponsored project agreements
- Ensuring the appropriate use of animals, human subjects, recombinant DNA, biological agents, and radioactive materials
- Securing the intellectual property rights of the university
- Supporting, through the university's Office of the Vice President for Research and Innovation and Weill Cornell Medicine's Office of the Research Dean, the research data management efforts of the respective principal investigators
- Resolving disputes among researchers over data control or access
- Approving any transfer of original research data off campus for archival or other purposes
- Facilitating the investigation of charges, such as research misconduct or conflict of interest
- Providing a process for making data retention attestations and a suitable repository for completed attestations

1.3.6. Principal Investigator (PI) responsibilities: Within the limits set by the superseding authority of Cornell, agreements with collaborators, and any applicable terms within sponsored

agreements, the principal investigator (PI) has the right and authority to control the use of, and access to, any research data conducted under their management or supervision, including the use of data in scholarly publications and presentations.

With support from the university, the PI is responsible for maintaining and retaining research data in accordance with this policy. Responsibilities include, but are not limited to:

- Collecting, maintaining, retaining, and providing access to research data for the periods required by this policy
- Determining, according to the sponsored research agreement, the data use agreement, or other requirements, whether research data is public, confidential, or otherwise restricted
- Ensuring compliance with any restrictions mandated by federal International Traffic in Arms Regulations or Export Administration Regulations, including restrictions on publication, or sharing with non-United States citizens
- Securing and controlling access to research data and ensuring that required protections are provided
- Determining, consistent with obligations to sponsors, collaborators, and students, how data will be published or presented
- Ensuring that data is available for review by the university, sponsors, journals, and others as described by this policy, journal policies, and sponsor requirements
- Informing the Center for Technology Licensing, through the Technology Transfer Process, of any data supporting a new invention
- Providing a research data retention attestation

1.3.7. 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 cited 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, including, but not limited to, when:

- The terms of a sponsored research agreement require a longer retention period (for example, New York State and the Public Health Services Office of Research Integrity typically require retention for six years after the close of a grant).
- The Center for Technology Licensing deems retention is required to protect intellectual property.
- Allegations regarding the research, such as research misconduct or conflict of interest, arise and remain unresolved; research data must be retained until such allegations are fully resolved.
- Legal action, investigation, or official inquiry related to the research is ongoing; research data must be retained until such issues are fully resolved.
- A data use agreement requires that the data be destroyed in less than three years or retained longer than three years for the data it controls.

Research data involving human subjects must comply with the Institutional Review Board (IRB) expectations. Any data that includes Protected Health Information must comply with HIPAA mandates and processes, including security standards. Any research involving vertebrate animals must follow with University Policy 1.4, Care and Use of Live Vertebrate Animals in Research and Teaching.

Research data owned by a government or other agency to which a researcher is granted limited access (for example, access through a data use or sponsored research agreement to government administrative data or health records) is not subject to this retention requirement but the steps by which the researcher gained access to such data must be made publicly available so others can apply for access to the same data.

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.

Other private or personal storage of data is not acceptable unless the Vice President for Research and Innovation grants specific permission based on a compelling justification and assurance that data will be maintained with appropriate oversight.

Although the principal investigator is expected to maintain the research data in a form reasonably well organized according to the norms of the field of research, this policy recognizes that researchers have many ways of organizing their data and labeling their files and does not impose any specific required organization.

1.3.8. 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. For funded research, “close out” of research means the end of the grant or contract agreement or 60 days prior to the faculty member leaving the institution, whichever comes first. Faculty must enter required metadata, data, and a method description into the WCM Institutional Data Repository for Research (WIDRR). Faculty may either:

- Deposit all relevant research data to an approved public data repository with a deposit of metadata and access instructions to that public site into WIDRR after publication or within three years after the final project closeout of all funded or unfunded research.
- Deposit directly into the WIDRR after publication or within three years after the final project closeout of all funded or unfunded research.

Exceptions to this requirement can only be made with the prior approval of the Sr. Associate Dean for Research.

Faculty operating under a sponsored research agreement, data use agreement, or other agreement, must determine whether the research data is high, medium, or low risk (as per WCM IT Policy 11.03), and must document this determination in the data management plan maintained in the Weill Research Gateway system.

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 cited 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.

Research data owned by a government or other entity to which a researcher is granted limited access (for example, access through a data use or sponsored research agreement to government administrative data or health records) is not subject to this retention requirement, but the steps by which the researcher gained access and performed analysis must be compliant with section 1.3.9 described below so others can apply for access to the same data.

Research data involving human subjects must comply with the Institutional Review Board (IRB) expectations. Any data that includes Protected Health Information must comply with HIPAA mandates and processes, including security standards. Any research involving vertebrate animals must follow with University Policy 1.4, Care and Use of Live Vertebrate Animals in Research and Teaching.

1.3.9. Research data security: Research data that incorporates confidential information including, but not limited to, personally identifiable human participant data, protected health information (PHI), trade secrets, or export-controlled information, must have the same security protections and be treated in the same manner as institutional information classified as “high risk” in [University Policy 5.10, Information Security](#), or confidential (level 1) information in [WCM IT Policy 11.03, Data Classification](#).

Suspected or proven disclosure or exposure of confidential or otherwise restricted data must be immediately reported to the Vice President for Information Technology and Chief Information Officer or the WCM Information Technologies and Services Chief Information Officer. If the data includes data from human participants in research, the Institutional Research Board for the respective campus must also be informed.

If the data was made available by an external sponsor as part of a sponsored research agreement or a limited-use data provider via a data use agreement, the Office of Sponsored Programs or the Office of Sponsored Research Administration must be informed. If the exposed data was export controlled, the Export Control Officer must be informed.

Note for Ithaca faculty: Costs associated with the preservation and security of research data during the term of a sponsored award are typically allowable direct costs of conducting research. The university provides secure cloud storage at affordable rates. The [Data Storage Finder](#) provides a mechanism to compare options. Storage options provided to the Cornell community are likely to serve the needs of most researchers. Those faculty who work with extremely large data sets may want to take advantage of archives available to their discipline so long as such archives have privacy and security protocols sufficient to address any privacy and security requirements associated with the data.

Note for WCM faculty: WCM ITS provides secure cloud storage at affordable rates. Those faculty who work with extremely large data sets are strongly encouraged to take advantage of NIH-approved archives available to their discipline in accordance with section 1.3.8.

1.3.10. Research data access: Cornell (including Weill Cornell Medicine) has the right to access all research data generated under its auspices, supported by its administered funds, or conducted using its facilities. The university has the right to take custody of research data to ensure needed and appropriate access, for example, to facilitate a response to an allegation of research misconduct.

Access to data provided to researchers under a data use agreement, nondisclosure agreement, subaward, or other contractual agreement is governed by the terms of such agreement. Access to export-controlled data is limited to individuals approved by the Export Control Officer, in accordance with applicable laws and regulations.

Subawards follow the primary sponsor's terms and conditions governing data retention. Should data retention requirements not be addressed in the primary sponsor's terms and conditions, subrecipients will follow their institution's policies. In the absence of a policy, 2 CFR § 200.334 applies.

A principal investigator (PI) may grant university researchers and staff access to research data for research or administrative purposes, subject to all university rules, state and federal laws, and contractual obligations relevant to the data. Access to research data by researchers who are not university employees may be governed by additional agreements.

The PI is responsible, with support from the Export Control Officer, Office of Sponsored Programs, or Office of Sponsored Research Administration, for ensuring compliance with any restrictions mandated by the Office of Foreign Assets Control, International Traffic in Arms Regulations, or Export Administration Regulations. This includes any agreed-upon terms from sponsors and data providers such as publication and sharing with non-United States citizen collaborators and/or students.

Faculty and staff who have the authority to give researchers access to data must inform them, in writing, of any limitations or restrictions on the use or dissemination of the data. (For Weill Cornell Medicine faculty this should also be documented in the data use agreement and in the data management plan.) For sponsored research, faculty and staff may not allow access to any data obtained under a sponsored research or data use agreement without approval from the sponsor or data provider. The Office of Sponsored Programs or the Office of Sponsored Research Administration can assist in obtaining approval.

For Ithaca-based faculty, a PI or other researcher who leaves the university may take copies of research data for projects on which they have worked, unless this is prohibited by a data use agreement, sponsor agreement, federal or state law, or other applicable prohibition. Taking copies of data supporting an invention disclosure or an unpublished patent application must be approved by the Center for Technology Licensing. Taking copies of any data covered by a data use agreement or sponsorship agreement must be approved by the Office of Sponsored Programs. The purposes for which such data may be used depend on agreement with the PI, or as formally agreed upon beforehand in a data use agreement. In all cases, the original research data must be retained at Cornell unless the Vice President for Research and Innovation specifically authorizes moving it to another institution.

For Weill Cornell Medicine (WCM) faculty, a PI or other researcher who leaves the university may request a copy of research data for projects on which they have worked. Requests are submitted to the Senior Associate Dean for Research. A review of the data is made and the request may be approved. The purposes for which such data may be used depend on agreement with the PI, or as formally agreed upon beforehand in a data use agreement. In all cases, the primary research data must be retained at WCM unless the Senior Associate Dean for Research specifically authorizes moving it to another institution. The PI or a delegated member of their research team must be available to assist anyone to access and understand the data for any WCM-approved purpose. This term remains a responsibility of the PI even after they leave WCM.

1.3.11. Publication: The principal investigator (PI) has the right and responsibility to ensure that research is accurately reported to the scientific and academic community, as well as to select the vehicle most appropriate for publication or presentation of research data and results.

The PI must ensure that any figures, tables, images, data, or assertions included in the publication can be defended with raw data, and demonstrably are not deceptively manipulated. For research conducted with co-PIs, the co-PIs jointly share this right and responsibility unless it is expressly written otherwise in the manuscripts.

1.4 Collection and Retention of Information

Records associated with this policy are retained or disposed of in accordance with [University Policy 4.7, Retention of University Records](#).

1.5 Compliance

The University Compliance Office, University Audit, and others may audit or investigate to assess compliance with this policy. Non-compliance with university policies is addressed in accordance with applicable policies and procedures, and is subject to progressive disciplinary action up to and including termination.

1.6 Resources

Topic	Contact	Phone/email	Website
Care and use of live vertebrate animals in research and training	Institutional Animal Care and Use Committee	iacuc@cornell.edu	University Policy 1.4, Care and Use of Live Vertebrate Animals in Research and Training
Data classification (Weill Cornell Medicine)	WCM-Information Technologies & Services	support@med.cornell.edu	WCM IT Policy 11.03, Data Classification
Data management planning	Research Data Management Service Group (RDMSG)	rdmsg-help@cornell.edu	Data Management

Data security, data classification	IT Security & Policy	itservicedesk@cornell.edu	University Policy 5.10, Information Security
Data storage and preservation systems	Research Data Management Service Group (RDMSG)	rdmsg-help@cornell.edu	Data Storage Finder Regulated Data Chart Data Storage and Backup
Data use agreements initiated outside Cornell	Office of Sponsored Programs	osp_dua@cornell.edu	Data Use Agreements
Data use agreements initiated outside Weill Cornell Medicine	Office of Sponsored Research Administration (OSRA)	grantsandcontracts@med.cornell.edu	Office of Sponsored Research Administration (OSRA)
Health Insurance Portability and Accountability Act (HIPAA) and Protected Health Information (PHI)	Weill Cornell Medicine-Research	support@med.cornell.edu	HIPAA and PHI information
Human participant research (Ithaca-based)	Office of Research Integrity and Assurance (ORIA)	irbhp@cornell.edu	Institutional Review Board (IRB) for Human Participant Research
Human participant research (Weill Cornell Medicine)	Weill Cornell Medicine Research	support@med.cornell.edu	Institutional Review Board (IRB) for Human Participant Research
Intellectual property rights	Center for Technology Licensing	ctl-connect@cornell.edu	University Policy 1.5, Inventions and Related Property Rights
Lab notebooks (Weill Cornell Medicine)	WCM Information Technology and Services	scisoft@med.cornell.edu	WCM Electronic Lab Notebooks

Lab notebooks and lab archives	Center for Teaching Innovation Research Data Management Service Group	Acadtech@cornell.edu rdmsg-help@cornell.edu	Electronic Lab Notebooks Lab Archives
Managing restricted datasets, data security plans and data provider agreements (Ithaca-based)	Cornell Restricted Access Data Center (CRADC) Office of Sponsored Programs	cradc@cornell.edu osp_dua@cornell.edu	Data Security Data analysis software applications
Managing restricted datasets, data security plans and data provider agreements (Weill Cornell Medicine)	WCM Data Core	dcore-ticket@med.cornell.edu	WCM Data Core
Retention of university records	General Counsel	607-255-5124	University Policy 4.7, Retention of University Records
Rigor and reproducibility in research repository	Office of Research Integrity and Assurance (ORIA)	crco_compliance@cornell.edu	Rigor and Reproducibility Planning
Standards on sharing data	The National Academies Press National Institutes of Health National Science Foundation	SciencePolicy@od.nih.gov	Sharing Publication-Related Data and Materials NIH Data Sharing Policy NSF Dissemination and Sharing Research Results
Technology transfer process	Center for Technology Licensing	ctl-connect@cornell.edu	New Invention Disclosure

2. To Whom This Policy Applies

2.1 For Cornell community based out of:

- Ithaca-based locations
- Weill Cornell Medicine – All Locations
- Cornell Tech – New York City

2.2 Who should read this policy:

All members of the university community, including faculty, staff, and students, who are involved in the design, conduct, or reporting of research at, or under the auspices of, Cornell University

3. Definitions

Term	Definition
Data management plan (DMP)	Description of the data to be collected or generated through a research project, including how the data will be managed, described, stored, and made available both during and at the conclusion of the project.
Research data retention attestation	Affirmation by a Cornell researcher that they understand and will abide by this policy.
Final project closeout	For data generated under a sponsored project when the award period has ended and all deliverables have been submitted. For data acquired through a data use agreement under a sponsored project, when the award period has ended, deliverables have been submitted, and data has been destroyed or returned to the data provider according to the terms of the agreement. For non-sponsored projects, when all work has ceased and no further publications related to the project are anticipated.
Metadata	Data that describes other data, for example title, abstract, author, and keywords (publications); organization and relationships of digital materials; and file types or modification dates.
Methods	Procedures, protocols, techniques, and related parameters and cutoffs associated with the processing and analysis of research data, which would facilitate a reasonable attempt to understand, validate, and replicate the findings.

Principal investigator (PI)	<p>Individual who is responsible for the overall direction of the project. In the case of research projects led by students who are supported by the university or require significant university resources, the faculty advisor.</p> <p><i>Note:</i> Students who are not supported by the university and who are not using significant university resources in their research are the owners of their data and any intellectual property resulting from the research project.</p>
Research data	<p>Information needed to evaluate reported results of research.</p> <p>This policy applies the Federal definition broadly to data related to research, without regard to how the research is funded or how the data was acquired. As defined in 2 CFR 200.315 (e) (3): “Research data means the recorded factual material commonly accepted in the scientific community as necessary to validate research findings, but not any of the following: preliminary analyses, drafts of scientific papers, plans for future research, peer reviews, or communications with colleagues. This ‘recorded’ material excludes physical objects (e.g., laboratory samples). Research data also do not include: (i) Trade secrets, commercial information, materials necessary to be held confidential by a researcher until they are published, or similar information which is protected under law; and (ii) Personnel and medical information and similar information the disclosure of which would constitute a clearly unwarranted invasion of personal privacy, such as information that could be used to identify a particular person in a research study.”</p> <p>Nothing in this definition is intended to supersede an agreement with a human research subject to code or otherwise de-identify personal information or specimens they provide for research purposes.</p>
Researcher	Any faculty or staff member, student, postdoctoral researcher, research associate or fellow, or other person involved in the design, conduct, or reporting of research.
University facilities	Facilities owned by the university or under contract by the university with a third party.
Weill Cornell Medicine Institutional Data Repository for Research(WIDRR)	Digital system established to register, submit, review, store, catalog, and provide archival access to Weill Cornell Medicine research data. Meets WCM data retention obligations and can be used to provide access to a subset of (published) research data to meet data sharing requirements. WIDRR is managed by WCM’s Office of Information Technology Services (ITS).

4. Policy Administration

Policy Clarification and Interpretation	Contact	Phone	Email/Web Address
Ithaca-based locations	Office of the Vice President for Research & Innovation	607-255-7200	vp_research@cornell.edu researchservices.cornell.edu/offices/ovpr
Weill Cornell Medicine – New York City	Office of the Senior Associate Dean for Research	212- 746 1361	ResearchDean@med.cornell.edu research.weill.cornell.edu

5. Revision History

Date	Summary of Revisions
02/09/2022	Interim Policy 4.20 replaced with Policy 4.21
08/07/2020	Originally issued