

**Research Data Management**  
**Principles and Guidance**  
**Yale University**

**Table of Contents**

1.	Introduction	p.1
2.	Why Manage Research Data?	p.2
3.	Research Data Responsibilities at Yale	p.3
4.	Research Data Principles and Guidance	p.4
4.1	Research Data Ownership and Rights	p.5
4.2	Research Data Retention	p.6
4.3	Research Data Preservation	p.7
4.4	Research Data Security	p.7
4.5	Research Data Sharing and Public Access	p.8
4.6	Research Data Plans and Documentation	p.9
5.	Appendix 1: Definitions	p.11

*This document is intended to provide guidance regarding principles to consider as research data management practices are developed at Yale University. It presents an overview of research data management, a description of relevant University policies, rules, and expectations, and guidance to Yale researchers about available university resources. It reflects the current state of data management and will be updated over time as conditions evolve.*

*This document should be read in conjunction with the University [Policy on Research Data and Materials](#), Yale [Information Technology Services \(ITS\) Security Policy](#), [Data and Information Classification Framework](#), [Copyright Policy](#), [Patent Policy](#), and [Faculty Handbook](#). In coordination with relevant policies, this document aims to assist Yale researchers in fulfilling their ethical and legal obligations while attempting to minimize the inherent burden associated with research data management.*

## **1. Introduction**

Two of the University's essential purposes are to impart knowledge and to enlarge humanity's store of knowledge. Specifically, the following principles of Yale University guide this document:

- a) [Yale is committed](#) to improving the world today and for future generations through outstanding research and scholarship, education, preservation, and practice. The University deems appropriate stewardship of research data as fundamental to both high-quality

research and academic integrity and therefore seeks to attain the highest standards in the generation, management, retention, and preservation of research data. Yale University regards making data resulting from academic research available to the public within regulatory and legal constraints as an extension of its mission.

- b) Yale University supports the primary right of [academic freedom](#) for researchers to choose the nature and direction of their research inquiries, to use research data generated to pursue future research and to publish related scholarly work. This freedom comes with the responsibility of researchers to disseminate their research findings to the scientific and academic community.
- c) Yale University supports the academic community's standard that the scientific principle of reproducibility is essential to the advancement of science. The ability to verify research findings by new investigators or by using independent methods relies upon access to relevant research data, materials, protocols, and documentation.

The University is primarily responsible for meeting all obligations concerning research data, and supporting an environment in which the objectives of its policies and principles are met. Yale University is committed to developing and supporting a collaborative environment in which the general principles and objectives of its research data policies can be achieved. Where research is resource-intensive and where possible researchers should seek to recover the direct costs of managing research data from the research funder. Research data management plans that involve significant commitment of University resources must be approved by the Office of the Provost or dean of self-supported schools.

Researchers are responsible for collecting, managing, retaining, and sharing research data and materials in compliance with legal requirements, Yale policy, and the terms and conditions of sponsored awards, and for assisting the University in fulfilling its obligations in this regard.

## **2. Why Manage Research Data?**

Research data are an increasingly important and valuable output of the scholarly research process, across all disciplines. They are an essential part of the evidence necessary to evaluate research results and to reconstruct the events and processes leading to them.

Research data management includes caring for, facilitating access to, and preserving research data. All research requires some level of data management before, during, and subsequent to the active research period. "Research data management" refers to the actions researchers take before, during, and after a research project that involve data – from planning, to creating, organizing, securing, sharing, and looking after the data – and the documentation they keep about these actions.

In addition to adhering to research data management [best practices](#), Yale researchers should engage in research data management for these reasons:

- a) Proper research data management aligns with Yale’s mission and key principles as stated in Section 1.0 above.
- b) Many public and private grant-funding organizations have policies on research data that include the expectation that data underpinning published research findings will be made openly available in a timely and responsible manner. Federal agencies, including the [National Institutes of Health](#) and the [National Science Foundation](#) among others, have expectations of grantees for providing public access to research data and materials, notably policies for data sharing, management and public accessibility to published results. For example: The [US Office of Science and Technology Policy \(OSTP\) 2013 memo](#), updated in [2014](#), directs federal agencies and offices to develop and submit plans to OSTP that ensure peer-reviewed publications and digital scientific data resulting from federally-funded scientific research are accessible to the public, the scientific community, and industry (to the extent feasible and consistent with applicable law and policy, agency mission, resource constraints, U.S. national, homeland, and economic security, and, the specific objectives of the memorandum). In response to the OSTP, federal agencies have begun issuing expectations of grantees for providing public access to research data and materials, while others (notably the National Institutes of Health and the National Science Foundation) have long standing policies for data sharing, management, and public accessibility to published results.<sup>1</sup>

Yale University and its researchers have legal, institutional, and ethical obligations to manage and retain records of research in sufficient detail and for an adequate period of time to enable appropriate responses to questions about accuracy, reproducibility, authenticity, primacy, and compliance with laws and regulations governing the conduct of the research.

### **3. Research Data Responsibilities at Yale**

- a) Yale University has legal, institutional, and ethical obligations to manage and retain records of research – including research data, materials and documents, as well as materials and information that relate to the administration and financial management of research, reporting of research results, sponsored award applications, and human research records – 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 conduct of the research. This responsibility continues even after researchers who originally collected those data and materials have left the University.
- b) Yale has a responsibility to protect the rights of Yale researchers as provided in its policies.
- c) Researcher, departments, faculties, divisions, central administrative units, service providers and, where appropriate, research sponsors and external collaborators should work in

---

<sup>1</sup> Additional information about federal mandates on access to research data can be found here: The [Association of Research Libraries](#); US Department of Health and Human Services [common guiding principles](#) for increasing public access to the results of federally funded research funded by HHS operating divisions; US National Institutes of Health February 2015 [report](#); US National Science Foundation data sharing [policy](#).

partnership to implement best practices and meet relevant legislative, research sponsor, and regulatory requirements.

- d) Researchers are responsible for assisting the University in fulfilling its obligations listed above. Researchers bear primary responsibility for the quality and integrity of their research, and for collecting, managing, retaining, and sharing research data, materials, and records according to legal requirements, Yale policy, and the terms and conditions of sponsored awards. This includes, but not limited to, the selection of an appropriate method of storing research data and materials, and identification of records related to the research process that should be retained to enable complete responses to questions about accuracy, authenticity, primacy, and compliance with laws and regulations governing the conduct of research. Researchers are expected to be cognizant of current best practices in their respective discipline.
- e) Principal Investigators (PIs) must abide by all relevant university and school policies (see especially [Faculty Handbook](#) Section XX.C). In addition, PIs are responsible for educating all participants in the research project of their obligations regarding research data and materials related to the research data. PIs should also consult with University resources regarding the development of any contingency plans for data storage in the event of unforeseen events. Particularly for long-term or longitudinal research projects, PIs should establish procedures for the protection of essential records in all forms (electronic and hardcopy) in the event of a natural disaster or other emergency.
- f) Researchers are responsible for specifying research data management plans in compliance with legal requirements, Yale policy, and the terms and conditions of sponsored awards. Best practice dictates that plans for storing, maintaining, and protecting research data and materials should be made prior to beginning data collection and should be reviewed on a continual basis. Where research is resource-intensive and where possible researchers should seek to recover the direct costs of managing research data from the research funder. Research data management plans that involve significant commitment of University resources must be approved by the Office of the Provost or dean of self-supported schools.

#### **4. Research Data Principles and Guidance**

In support of Yale's commitment to ethically responsible research, the University adheres to NIH and NSF requirements for training in the [Responsible Conduct of Research](#) (RCR).

With respect to data acquisition, management, sharing and ownership, Yale's Office of Research Administration [RCR standards](#) state that,

“The integrity of research data and the usefulness of the research it supports depend on careful attention to detail from initial planning through final publication. While the scientific disciplines may differ in data management practices, there are generally accepted standards with which the University community should be aware and adhere to regarding data ownership, data collection, data protection and data sharing. Key considerations for data collection include using the appropriate method, providing

attention to detail, obtaining the appropriate permissions for use of certain categories of data and the accurate and secure recording of data. Data should be maintained and secured in such a way that permits confirmation of research findings, establishes priority, and can be reanalyzed by other researchers. Data should be stored in such a way that protects confidentiality, is secure from physical and electronic damage and destruction and can be maintained for the appropriate time frame dictated by sponsor and University policy. Conditions imposed by sponsors, the University, and data sources may affect data acquisition, management, sharing and ownership. NIH Policy, for example, includes specific criteria for sharing and publishing research data resulting from NIH-sponsored research.”

The following section describes relevant precepts as described in existing Yale policies in regard to six areas of research data management: ownership and rights; retention; and preservation; security; sharing and public access; and documentation and planning. Each section provides guidance to researchers in support of each of these precepts.

#### **4.1 Research Data Ownership and Rights**

- a) Yale [Policy on Research Data and Materials](#) states that “Yale’s ownership and stewardship of research data and materials for projects conducted at the University, under the auspices of the University, or with University resources are based on both regulation and sound management principles and apply in all cases except where precluded by the specific terms of sponsorship or other agreements.” The University may assert its formal ownership of all data and research materials created under its auspices regardless of funding source, unless specific terms of sponsorship, other agreements or University policy supersede these rights.
- b) Yale holds that researchers are custodians of their research data, responsible for the conduct of the research and for the appropriate management and retention of the resulting research materials and records.
- c) The University has an ownership interest in inventions or discoveries made under the auspices of the University, whether patentable or not, and certain copyrighted materials that result from the research activities of faculty, staff and students at Yale. Any intellectual property rights in research data and related materials shall remain with the University, and any dissemination and commercialization of knowledge resulting from research conducted by Yale researchers shall be done in accordance with University policy. The University may choose to license high-value data. See [Copyright Policy](#) and [Patent Policy](#).
- d) As stated in the [Policy on Research Data and Materials](#), ownership of the original data may be transferred from Yale to a Researchers’ new institution under some circumstances.
- e) Any research data that Yale University may not own, but that is used in the course of research at Yale, such as data obtained from a vendor through a license agreement, is

governed by applicable law and the specific conditions of any license agreement. Ownership of research data and materials generated at other institutions under an agreement with Yale should be defined in said agreement.

Guidance: Research Data Ownership and Rights
• Guidance regarding transferring an award is available at the <a href="#">Office for Sponsored Projects</a> .
• Guidance regarding data use agreement, material transfer agreements, and other contracts are available at the <a href="#">Office for Sponsored Projects</a> .
• Guidance regarding patents and copyrightable materials are available at the <a href="#">Office for Cooperative Research</a> .
• Guidance regarding metadata appropriate for information about data ownership and rights <a href="#">Research Data Consultation Group</a> .

#### 4.2 Research Data Retention

- a) Researchers have certain obligations to record, maintain and retain research records, and to make those records available for grant monitoring and auditing purposes, as well as to enable investigators and the institution to respond to questions of research integrity and stewardship (see, for example, [OMB Circular A-81: Uniform Administrative Requirements, Cost Principles, and Audit Requirements for Federal Awards](#)) and in order to support research findings, justify the uses of research funds and resources, and protect any resulting intellectual property for an appropriate period after the completion of the research and in accordance with Yale policy.
- b) As stated in the [Policy on Research Data and Materials](#), “Generally, research data and materials that are commonly accepted in the scientific community as necessary to validate research findings must be retained 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.” Longer retention periods may apply in accordance with applicable law or agreement and as described in the Office of General Counsel’s [record retention schedule](#). Longer retention may also be desired by researchers or expected by the broader scientific community for future scholarship.
- c) Yale expects researchers to retain all appropriate research data, whether resulting from federal sponsorship or not, in their laboratories or otherwise on University premises or in electronic computing systems maintained or sanctioned by the University. It is recommended that any research data retained in other bona fide research locations be recorded in research data management plans, lab notebooks, or other research records, and available for retrieval by the University if required.
- d) Researchers are responsible for planning the ongoing custodianship of their data during and subsequent to the active research period, and should have systems or practices for

maintaining the essential research records that they create. This information should be incorporated, where appropriate, into a formal research data management plan.

Guidance: Research Data Retention

- [ITS](#) and the [Yale Center for Research Computing](#) are the relevant units for consultation about storage and the [Library](#) about preservation systems. They can provide advice and guidance to researchers on appropriate Yale technical resources that will help them comply with requirements.
- Guidance on data management planning is available at the [Library](#); The [DMPTool](#) is available to the Yale research community; login with Yale NetID. The Research Data Consultation Group (RDCG) assists with research data management planning, provides consultation about best practices, and ensures that researchers are taking full advantage of the many research services available on campus and that the relevant service providers at Yale know about planned use of systems and can plan accordingly. On the medical campus, additional guidance is available at the Yale Center for Clinical Investigators (YCCI), the Yale Center for Analytical Sciences (YCAS), and Medicine and University IT Partners. The Office of the Provost or deans of self-supported schools are the relevant units for approving research data management plans involving commitment of university resources. The Office of Research Administration and the Office of Sponsored Projects can provide advice and guidance to researchers as part of the grant submission process on how to seek to recover the direct costs of managing research data from the research funder where research is resource-intensive.

#### 4.3 Research Data Preservation

- a) Researchers should be aware of the relevant practices or codes within their research discipline that establish norms or best practices for the preservation of research data. In accordance with the standards of their field of inquiry and at their discretion, researchers may choose to deposit research data, samples, laboratory and research notebooks, correspondence and other relevant materials (such as software or source code) in archives for long-term preservation.

Guidance: Research Data Preservation

- Guidance regarding repositories approved by Yale for researchers to deposit their data, including any institutional partnerships Yale has with these repositories, metadata appropriate for long-term preservation, and data curation best practices to ensure persistent access to research data, materials, and records is available at the Research Data Consultation Group ([RDCG](#)).

#### 4.4 Research Data Security

- a) As stated in the Yale ITS requirements for [Security and Confidentiality](#), “appropriate protections (security controls) for the confidentiality, integrity and availability of data must be implemented to comply with regulations, contracts and other agreements.” The security status of research data is determined by the nature of the data and by the terms of any relevant legislation, regulation, agreement, or University policy. Yale’s [Data Classification Policy](#) provides guidance for classifying data according to their sensitivity,
- b) Yale is particularly focused on the protection of human subject research data that are confidential by reason of applicable law and regulation, agreements covering the acquisition and use of the data, and University policies. Researchers should consult the University’s [Human Research Protection Program](#) (HRPP) policies and procedures.
- c) Researchers are expected to understand Yale ITS security precepts and responsibilities well in advance of entering into an agreement, grant or contract. Researchers should consult ITS [policies](#) and [security standards and guidance](#).

Guidance: Research Data Security
<ul style="list-style-type: none"><li>• Guidance regarding what data Yale considers sensitive, confidential, or public data is available at the <a href="#">Human Research Protection Program</a>; <a href="#">ITS Data protection principles</a>; <a href="#">ITS Data and Information Classification Framework</a>. Guidance about levels of data security at Yale and approved services is available on It’s Your Yale: <a href="#">Protect Your Data</a>.</li></ul>
<ul style="list-style-type: none"><li>• Guidance regarding Yale security standards for data, application, and equipment (including storage, access management, file sharing and transfer, disposal of digital media which contain research data, and reporting security incidents for lost or stolen research data) is available at the Yale Center for Research Computing (<a href="#">YCRC</a>).</li></ul>
<ul style="list-style-type: none"><li>• Guidance regarding metadata appropriate for information about data security, including privacy and confidentiality is available at the Research Data Consultation Group (<a href="#">RDCG</a>). On the medical campus, additional guidance is available at the Yale Center for Clinical Investigators (<a href="#">YCCI</a>), the Yale Center for Analytical Sciences (<a href="#">YCAS</a>), and <a href="#">Medicine and University IT Partners</a>.</li></ul>

#### 4.5 Research Data Sharing and Public Access

- a) As stated in the [Policy on Research Data and Materials](#), “Research data and materials shall be made available to the extent feasible while minimizing harm to the legitimate interests of the University, to research subjects, and to other parties, subject to Yale policy and to legislative, regulatory, contractual, ethical or other obligations, including but not limited to whether providing such public access would be cost prohibitive.”
- b) Yale researchers shall comply with regulations, contracts and other agreements including any restrictions or waivers to ensure that research data are made available for access and re-use where appropriate and under appropriate safeguards. At all times,

public access to research data should be considered in light of any intellectual property rights that are elsewhere recognized and protected under the law.

- c) In addition to its obligation to provide public access to research data, Yale University supports the principle of openness in research, and considers making research output available to the public crucial to a vibrant and healthy academic environment. Public access to research data for academic purposes is defined as readily available access to research data for purposes of re-use, for example, verification, replication, and additional analyses. Researchers should strive to make research data available with as few restrictions as possible. Research data should not be licensed under terms of use that unreasonably restrict their management, sharing, or use by others and exclusive rights to re-use or publish research data should not be granted to other entities such as commercial publishers or agents without retaining the rights to make the data openly available for re-use, unless this is a condition of funding, or otherwise appropriate (e.g., requirement of the publisher, part of technology transfer, condition of use of equipment or software).
- d) Researchers should include information about public access to data in research data management plans and consult with the appropriate units on campus to ensure that they are meeting requirements, including the Office of Research Administration, Office of Cooperative Research, the Yale University Human Research Protection Program (HRPP), local and central Information Technology Services representatives, and the Research Data Consultation Group.
- e) Sufficient metadata should be recorded and made available such that other researchers can understand the process used for data collection and potentially re-use the data. This includes metadata specific to attribution and citation of the data upon re-use.

Guidance: Research Data Sharing and Public Access
<ul style="list-style-type: none"><li>• Guidance regarding appropriate access to and reuse of data is available at the Research Data Consultation Group (<a href="#">RDCG</a>).</li><li>• Guidance regarding metadata that describes access to and reuse of research data, including metadata specific to attribution and citation of the data upon re-use data is available at the Research Data Consultation Group (<a href="#">RDCG</a>).</li></ul>

#### 4.6 Research Data Plans and Documentation

- a) Researchers are expected to provide metadata (descriptive information) to ensure that research data are discoverable, citable, and understandable in the future by others. Good practice includes providing documentation of protocols used in data collection, generation, processing, or analysis where applicable; information about rights, ownership, licensing, and access, which are pertinent to the terms of future re-use of the data; proper de-identification of personally-identifying information (PII) or personal health information (PHI) where applicable; and information necessary for proper and

secure storage, retention, and preservation of research data. Where possible, metadata should be provided in machine-readable format.

- b) Researchers are obligated to comply with sponsored research agreements that may require research data management plans. In the event that a sponsor has no requirement or offers no guidelines, Yale encourages researchers to produce a data management plan for all research projects.
- c) Data management plans should identify the person responsible for data management and describe procedures and any delegated responsibilities for the collection, storage, security, use, re-use, sharing, access, retention, and preservation of the research data and records associated with the research in keeping with the principles herein and with relevant Yale Policies, and in accordance with best practices as defined by the respective discipline. As stated above, data management practices vary across scientific disciplines but there are generally accepted standards within some disciplines with which the University community should be familiar.
- d) Where appropriate, in a joint or multi-institution collaborative research project, associated research data management plans should define specific responsibilities of each institution. The PI has ultimate responsibility for management of research data that is under the University's control.
- e) All research data management plans should be shared with graduate students, postdoctoral associates and fellows and other members of the research team to ensure consistency and a common understanding and implementation of the plan.

Guidance: Research Data Plans and Documentation
<ul style="list-style-type: none"><li>• General guidance regarding research integrity and responsible conduct of research (<a href="#">RCR</a>).</li></ul>
<ul style="list-style-type: none"><li>• Guidance regarding research data documentation and metadata appropriate for information about data collection, transformation, analysis is available at the Research Data Consultation Group (<a href="#">RDCG</a>) .</li></ul>
<ul style="list-style-type: none"><li>• A data management planning guide is available at the <a href="#">Library</a>; The <a href="#">DMPTool</a> is available to the Yale research community (login with Yale NetID). The <a href="#">RDCG</a> assists with research data management planning, provides consultation about best practices, and ensures that researchers are taking full advantage of the many research services available on campus and that the relevant service providers at Yale know about planned use of systems and can plan accordingly.</li></ul>
<ul style="list-style-type: none"><li>• The Office of the Provost or Deans of self-support schools are the relevant units for approving research data management plans involving commitment of university resources. The <a href="#">Office of Research Administration</a> and the <a href="#">Office of Sponsored Projects</a> can provide advice and guidance to researchers as part of the grant submission process on how to seek to recover the direct costs of managing research data from the research funder where research is resource-intensive.</li></ul>
<ul style="list-style-type: none"><li>• <a href="#">ITS</a> and the <a href="#">Yale Center for Research Computing</a> are the relevant units for consultation about storage and the <a href="#">Library</a> about preservation systems. They can provide advice and</li></ul>

guidance to researchers on appropriate Yale technical resources that will help them comply with requirements.

- Guidance regarding Yale security standards for data, application, and equipment (including storage, access management, file sharing and transfer, disposal of digital media which contain research data, and reporting security incidents for lost or stolen research data) is available at the [Yale Center for Research Computing](#).

## **5. Appendix 1: Definitions**

### **Co-Investigator**

This title designates key personnel for a project, but without the oversight responsibility of a PI.

### **Co-Principal Investigator**

This designation refers to individuals who share the responsibility for the project with the PI.

### **Principal Investigator**

The principal investigator (PI) is the individual (or in the case of co-principal investigators, the individuals) directly responsible for the overall design and conduct of a research project and is accountable to the University, and to external sponsors if any, for the proper programmatic, scientific, or technical conduct of the project, and its financial and day-to-day management. For sponsored projects, this role is formally defined by University Policy to be the individual designated by the University and approved by the sponsor to direct a project funded by the external sponsor. For research conducted by students, trainees, or postdoctoral fellows, this role is generally vested in the faculty advisor.

- The PI is ultimately responsible for the data used and generated in the project, and is responsible for managing research data and materials in accordance with the principles and requirements outlined in this policy. The PI may choose to delegate the management responsibility of the data and materials to a co-investigator of the project.
- Where research is conducted in collaboration with external research partners, responsibility for the management of data and materials for the project is still held by a Yale University PI, co-investigator, or delegate.
- The PI is responsible for ensuring continuity of the data and materials management and, should the PI leave the university or end association with the project, is responsible for making appropriate arrangements for the data and materials at the end of the project.

### **Research**

A systematic study intended to increase knowledge or understanding of the subject studied; a systematic study specifically directed toward applying new knowledge to meet a recognized need; or a systematic application of knowledge to the production of useful materials, devices, and systems or methods.

### **Research Data**

The 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, data files, 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.

### **Research Materials**

Tangible items that are the product of research or that are used to conduct research. Examples of research materials include reagents, cell lines, plasmids, vectors, chemical compounds, and some kinds of devices and software.

### **Yale Researcher**

A Yale faculty member, staff member, post-doctoral appointee, trainee, or student who generates research data or materials with the support of Yale resources.

### **Yale Resources**

Yale funds (including, for example, sponsored awards to Yale, salaries, gifts, endowment funds, and travel grants), facilities, staff, or equipment.