

# Data Retention Policy - Statistics & Data Science Department at Carnegie Mellon University

## 1. Introduction

This policy outlines the procedures for retaining and disposing of data collected, generated, or used by the Statistics and Data Science Department (StatsDS) at Carnegie Mellon University. It ensures responsible data management, compliance with legal and ethical requirements, and facilitates access to valuable research data for future projects.

## 2. Definitions

- **Data:** Any information collected, generated, or used for research purposes within the StatsDS department, including raw data, processed data, statistical models, and analysis results on departmental servers.
- **Data Subject:** The individual or entity to whom the data pertains (if applicable).
- **Metadata:** Information about the data, such as collection methods, format, and variables.
- **Research Project:** A defined study conducted by faculty, staff, or students within the StatsDS department.

## 3. Data Classification

Data will be classified based on its sensitivity and legal requirements. The following categories are examples:

- **Unrestricted Data:** Freely available datasets with no privacy concerns.
- **Restricted Data:** Data with confidentiality restrictions due to ethical considerations or data use agreements.
- **Highly Restricted Data:** Data subject to strict regulations (e.g., HIPAA).

## 4. Retention Periods

The minimum retention period for data will be determined by the following factors:

- **Legal Requirements:** Retention periods mandated by federal or state laws, funding agencies, or ethical research protocols will be followed.
- **Research Value:** Data with potential for future research will be retained for a longer period, considering its likely relevance to future studies.
- **Data Sensitivity:** Highly restricted data will have the longest retention period to ensure responsible use and compliance.

## 5. Retention Schedule

The following table provides a general guideline for minimum retention periods based on data classification. Data may be deleted sooner if deemed unnecessary. If data requires indefinite access, users can submit requests to the IT staff.

Data Classification	Minimum Retention Period
Unrestricted Data	3 years (may be deleted sooner if no longer used)
Restricted Data	5 years after project completion
Highly Restricted Data	As mandated by regulations (e.g., HIPAA)

## 6. Data Disposal

Data reaching the end of its retention period will be disposed of securely. This may involve:

- **De-identification:** Removing personally identifiable information (PII) before deletion.
- **Aggregation:** Combining data into a less detailed format that protects individual privacy.
- **Overwriting:** Securely overwriting electronic data to prevent recovery.
- **Physical Destruction:** Securely shredding physical data storage media.

## 7. Responsibilities

- **Principal Investigators (PIs):** Responsible for developing data management plans for their projects, specifying data collection, storage, and retention procedures.
- **Department Head:** Oversees the implementation of this policy and ensures adequate resources for data storage and security.
- **Information Technology (IT):** Provides secure storage solutions and data disposal methods in accordance with the policy.

## 8. Review and Updates

This policy will be reviewed periodically (at least every three years) to ensure its effectiveness and compliance with evolving legal and ethical requirements.

## 9. Communication

This policy will be provided to all new Statistics & Data Science account holders and will be available on the department website.

## **10. Retention Guidelines**

Data generated from research computing can grow rapidly. Users of StatDS computing assets must understand that computing storage resources are shared. Users are expected to monitor their storage usage and clean any unused or unrequired data.

Answering “Yes” to the following questions may help determine if data can be deleted.

- Is the data generated from a test?
- Is the data duplicated or stored in another location?

Answering “Yes” to the following questions may help determine if data should be retained.

- Will the data be used in future research efforts?
- Is this data containing final results?

## **11. Exceptions**

Exceptions to this policy may be granted on a case-by-case basis with written approval from the Department Head and justification provided.

This policy serves as a framework for data retention within the StatsDS department. By following these guidelines, the department can ensure responsible data management, protect sensitive information, and facilitate valuable research endeavors.

## **12. Active Usage Monitoring**

StatsDS Information Technology staff will actively monitor storage resource usage by users. IT staff may reach out to users who are consuming a significant amount of data in order to better understand their usage requirements and to provide guidance on the retention or deletion of that data.