Lab Data Management Protocols

## Downloads

* [Download as Word Document (DOCX)](/downloads/research/data-management.docx)

# Lab Data Management Protocols

## Introduction

Effective data management is critical for research integrity, reproducibility, and compliance, especially when incorporating AI tools into your research workflow. These protocols provide guidelines for managing research data in engineering labs that use AI tools.

## Data Collection and Storage

### Data Collection Practices

1. **Documentation Requirements**
   * Document all data collection procedures in detail
   * Record metadata about the collection process
   * Note any AI tools used in data collection
2. **Data Formats**
   * Use standardized, open formats when possible
   * Document proprietary formats and conversion processes
   * Maintain raw data in its original format
3. **Storage Infrastructure**
   * Use redundant storage systems (primary and backup)
   * Implement version control for datasets
   * Consider compute requirements for AI-based analysis

### Security and Access Control

1. **Access Levels**
   * Implement role-based access control
   * Maintain access logs
   * Review access permissions quarterly
2. **Security Measures**
   * Encrypt sensitive data
   * Secure authentication for access to data systems
   * Regular security audits

## Data Processing with AI Tools

### Preprocessing Protocols

1. **Data Cleaning**
   * Document all cleaning procedures
   * Preserve original data alongside cleaned versions
   * Record any AI-assisted cleaning methods
2. **Feature Engineering**
   * Document feature selection methods
   * Track transformations applied to data
   * Note any AI-generated features

### AI Processing Documentation

1. **Model Documentation**
   * Document all models used in analysis
   * Record model versions and parameters
   * Maintain logs of model training processes
2. **Output Validation**
   * Establish protocols for validating AI outputs
   * Document validation methods and results
   * Track cases of model failure or unexpected results

## Data Sharing and Publication

### Internal Sharing

1. **Lab Data Repository**
   * Organize data in a searchable repository
   * Include documentation and metadata
   * Track usage of datasets within the lab
2. **Collaboration Guidelines**
   * Establish protocols for sharing data with collaborators
   * Define access levels for different collaboration types
   * Document data exchange formats

### External Sharing and Publication

1. **Data Publication Standards**
   * Follow field-specific data sharing standards
   * Prepare data packages for repository submission
   * Include sufficient metadata for reuse
2. **Privacy and Ethical Considerations**
   * Review datasets for sensitive information before sharing
   * Ensure compliance with IRB and funding agency requirements
   * Document any data anonymization processes

## Data Archiving and Retention

1. **Archiving Timeline**
   * Archive completed project data within 3 months
   * Review archived data annually
   * Maintain data according to funder requirements
2. **Documentation Requirements**
   * Include comprehensive README files
   * Document directory structure and file naming conventions
   * Preserve software and environment specifications

## Compliance and Training

1. **Regulatory Compliance**
   * Maintain awareness of relevant regulations
   * Document compliance measures
   * Regular compliance reviews
2. **Training Requirements**
   * New lab members must complete data management training
   * Annual refresher training for all lab members
   * Specialized training for AI tool usage

## Implementation and Review

These protocols should be reviewed annually and updated as needed. All lab members are responsible for adhering to these guidelines and suggesting improvements to the data management process.