OCTRI Research Forum: Data Management 101

Julie Mitchell
Oregon Clinical & Translational Research Institute
Data Coordinating Center
Data Management Lifecycle

Plan
Collect & Capture
Store & Archive
Publish & Share
Analyze
Process (Clean)
Reuse

https://utas.libguides.com/ResearchData/lifecycle
Clinical Research Study Timeline

**Pre-study**
- Study Planning / Protocol Development
- EDC Setup, Regulatory Submissions, Contracting
- Site Selection & Initiation

**Study Conduct**
- First Patient In (FPI)
- Last Patient Out (LPO)
- Analysis & Reporting

**Analysis**
- Data entry
- Medical coding
- CRF maintenance
- Data cleaning & processing
- DSMB/Safety reporting
- Regulatory reporting
- Monitoring
- Data integration & reconciliation
- Data extraction
- Finalize Statistical Analysis Plan (SAP)

**Study Closure**
- Database locking
- Data transformation
- Results (Tables, Figures & Listing) generation
- Data archival
- Data sharing

Additional tasks:
- Assisting with grant application
- Evaluating site feasibility
- Reviewing contracts & agreements
- Reviewing the Protocol and consent
- Case Report Form (CRF) design & annotation
- Electronic Data Capture (EDC) design, setup & maintenance
- Setting up data integrations
- Determining what Medical coding is needed
- Data Sharing Plan
- Manual of Operations and Procedures (MOP)
Data Management Lifecycle

https://utas.libguides.com/ResearchData/lifecycle
Plan

- Identify categories of data must be collected (e.g. medications, laboratory results, procedures, I/O, diagnoses)
- Level of detail required for each type of data
- Quantity of data that is needed to address the objectives
- Frequency of data collected
- Methods used to collect the data
- Format for collecting the data
- Method of data monitoring
- Data validation checks
- Determine when are calculations and interpretations needed
Case Report Forms (CRF)

- Organization of Data collection forms
- Closed vs. Open questions
- Amount of data to collect
- Coding Data
- Instructions for Participants, Investigators & Study Personnel
- Data missingness
- Requiredness
Common Issues to Address When Creating Fields

• More than one item per cell
• Inconsistent units for numbers
• For numbers, number of decimal places inconsistent
• Inconsistent data values in each column
• Date formatting inconsistent
• Date Comparisons
• Missingness convention
Additional data to consider collecting

• Screening & enrollment

• Compliance

• Early termination

• Safety
Case Report Forms (CRF)

Granularity of data collection should also be considered

**Race (select one) vs. Race (check all that apply)**

- American Indian or Alaskan Native
- Asian
- Black or African American
- Native Hawaiian or Other Pacific Islander
- White
- Multi-race
- Other
- Prefer not to answer

Plan

Collect & Capture

Process (Clean)

Analyze

Publish & Share

Store & Archive

Reuse
Designing & Annotating CRFs

Paper CRF

e-CRF
Tools for Managing Data

• Spreadsheets (Excel)
• Statistical Analysis Programs (R, SAS, SPSS, STATA, MATLAB)
• Desktop Database system (MS Access or Filemaker Pro)
• Relational Database Management Systems (RDBMS)
• Purpose-built electronic data collection (Qualtrics, REDCap)
• 21 CFR Part 11 systems (Forte EDC)
• Consent Only (Docusign)
Considerations for Selecting an Electronic Data Capture (EDC) System

- Participant completed surveys
- eConsent
- Multiple languages
- Randomization
- Reporting
- Data clarification
- Audit log
CRF Completion Guidelines

• Describes how case report forms should be completed
• May have an “overall” section that applies to common elements across all forms
• Form and/or timepoint-specific sections
• Expected completion timeframes
• Prescribes visit windows
Data Management Plans

• Cover the research data lifecycle
• Help you to properly manage your data for your use, meet funder needs & enable sharing
• They help researchers to:
  • Make informed decisions
  • Avoid duplication, data loss and security breaches
  • Develop procedures early on for consistency
  • Ensure data are accurate, complete, reliable and secure
  • Plan to share data and increase impact
Data Management Lifecycle

https://utas.libguides.com/ResearchData/lifecycle
Essentials for Collection and Data Storage

• Protect your data as it is collected
• Transform your data AFTER when possible
• Backup your data
• Version control
• Folder structure and file naming conventions
• Audit trails
User Access & Administration

• Permissions should be based on the Principle of Least Privilege
• User permissions should maintain blinding
• Minimize users with ability to modify the EDC
Data Management Lifecycle

Plan

Collect & Capture

Process (Clean)

Store & Archive

Publish & Share

Analyze

Reuse
Accuracy

Completeness

Consistency

Timeliness

Relevance

Validity

Uniformity

Plan

Collect & Capture

Process (Clean)

Analyze

Publish & Share

Store & Archive

Expand & Share

Analyze
Steps for Processing Data

- Inspection and Profiling
- Cleaning
- Verification
- Validation
- Reporting

Plan
Collect & Capture
Process (Clean)
Analyze
Publish & Share
Store & Archive
Forensic (Audit)
Analyze
Detail & Study
Store & Archive
Discrepancy Management Process

DATA VALIDATION

DISCREPANCY

NO
DCF
YES

CRA

DATA MANAGER

INVESTIGATOR

SDV

SEC

RESOLUTION

DATABASE

https://europepmc.org/article/pmc/pmc3326906

(DCF = Data clarification form (Query), CRA = Clinical Research Associate (aka monitor), SDV = Source document verification, SEC = Self-evident correction)
Data Management Lifecycle

Plan
Collect & Capture
Process (Clean)
Publish & Share
Store & Archive
Reuse
Analyze

https://utas.libguides.com/ResearchData/lifecycle
Monitoring Data

• Review data for discrepancies, trends, problems, or relevant areas to discuss with the investigator or research coordinator
• Verify data based on source documents and EMR access
• Issue and resolve queries
Types of Analyses

Highlight data trends and progress to help assist with study conduct.

• Aggregation
• Anonymization
• Augmentation
• Blending
• Decomposing
• Deletion
• Formatting
• Imputation
• Labelling
• Normalization
• Sampling
Other considerations

- Review the Protocol’s Statistical section & Data Safety Monitoring Plan
  - How many people are expected to be screened?
  - How many people need to be enrolled/randomized?
  - How many people need to complete the study?
  - When is an Interim analysis needed?
- DSMB reporting requirements
- Safety reporting (IRB, NIH, FDA, etc.)
Database Locking

• Final review of data should be performed prior to locking.
• Unblinding occurs **AFTER** all queries being resolved **and** database status has been ‘locked’.
• At database lock, user permissions/roles should be revoked, revised or restricted
• Most EDC’s have a setting for ‘Database lock’ that prevents data from being edited.
Data Management Lifecycle

https://utas.libguides.com/ResearchData/lifecycle
Data Sharing

• Pre-determine Access
  • Contracts, protocols & consent forms need to reflect this information

• Methods
  • Preservation, de-identification, transformation, standardization & coding

• Study Conduct Details
  • Data dictionary, operational data

• Timeline

• Dissemination
  • Data request/access process
Data Sharing

• Sufficient contextual information is required to make sense of the data, explaining:
  • how data were created or digitised
  • what data mean (metadata/operational)
  • what their content and structure are
  • any data manipulations that may have taken place

• Documents that help provide contextual information include:
  • Protocol (all approved versions)
  • Regulatory documents
  • CRFs
  • Data dictionary
  • Dataset – limited to restrict PHI
  • Manual of Operations and Procedures (MOP)

https://www.researchgate.net/figure/It-shows-Document-management-system_fig1_263657311
Data Management Lifecycle

Plan
Collect & Capture
Process (Clean)
Analyze
Publish & Share
Reuse
Store & Archive

https://utas.libguides.com/ResearchData/lifecycle
Data Storage

- University network drives
- Approved Cloud Storage
- Dedicated Research Data Storage Repositories
Data Storage & Archival

- Include:
  - Protocol (all approved versions)
  - Consents (all approved versions)
  - Regulatory documents
  - All signed consents
  - Source documents
  - Paper CRFs (all versions)
  - Copy of database (e.g. XML file, CSV files, etc.)
  - Data dictionary
  - Dataset
  - Manual of Operations and Procedures (MOP)

- Keep presentations
- Document decisions including during study conduct
- Record trainings
- Keep meeting minutes
Data Management Lifecycle

https://utas.libguides.com/ResearchData/lifecycle
Reuse

• Revise measures & processes for future data collection – based on learnings and analyses

• Create/modify data repositories to store and track data, specimens, files, and/or participants
Data Management Lifecycle

https://utas.libguides.com/ResearchData/lifecycle
Questions?

Thank you!
Document Retention Period

The length of time you are required to keep study records depends on the type of data and the terms of your grant or contract. Below are links to a summary and OHSU policy:

- OHSU Summary of Regulatory Retention Requirements For Records Associated with Research

- Records Retention and Destruction Policy number 07-90-010
  [https://o2.ohsu.edu/policies/records-retention-and-destruction](https://o2.ohsu.edu/policies/records-retention-and-destruction)

- OHSU Records Retention Schedule
  [https://o2.ohsu.edu/system/files/2023-08/ohsu-records-retention-schedule.pdf](https://o2.ohsu.edu/system/files/2023-08/ohsu-records-retention-schedule.pdf)