OCTRI Research Forum: Data Management 101

Julie Mitchell

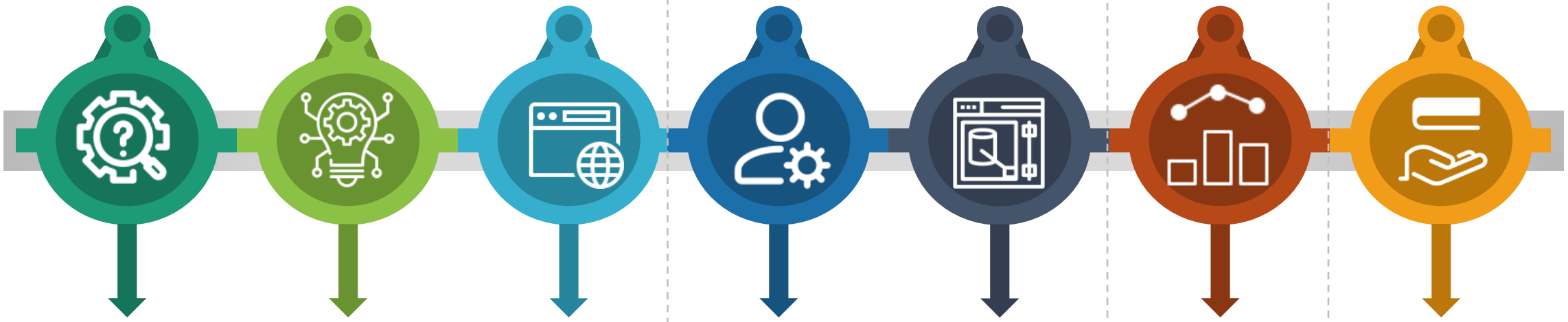
Oregon Clinical & Translational Research Institute

Data Coordinating Center

Data Management Lifecycle

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

Clinical Research Study Timeline



Pre-study

Study Planning / Protocol Development

Assisting with grant application
Evaluating site feasibility
Reviewing contracts & agreements
Reviewing the Protocol rand 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)

EDC Setup, Regulatory Submissions, Contracting

Site Selection & Initiation

Study Conduct

First Patient In (FPI)

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)

Last Patient Out (LPO)

Analysis

Analysis & Reporting

Database locking
Data transformation
Results (Tables, Figures & Listing) generation

Study Closure

Closeout

Data archival
Data sharing

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)

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

vs.

Race (check all that apply)

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

Designing & Annotating CRFs

Paper CRF

Screening ID *text (format = Site ID - XXX)* ___ - ___

Study Kit ID (if randomized) *text (6 characters)* ___

EMS HANDOFF REPORT (HAND ENTRY)

*Complete for all subjects screened. * Provide estimated time if exact time is unknown.*

1a. EMS Agency *text*

1b. EMS Rig No. *text*

Times*

2a. Injury *time (24hr)* H H : M M

2b. EMS Notification *time (24hr)* H H : M M

2c. Responding Unit Arrived at Patients Side *time (24hr)* H H : M M

2d. Study Drug Start *time (24hr)* H H : M M

2e. ED Handoff / Threshold Crossing Time *time (24hr)* H H : M M

Complete for all subjects randomized

Study Drug Administration

New Kit assigned to EMS rig? *radio*
- Yes
- No

If yes, new Kit ID *text*

3a. KCentra Intended Dose *radio*
- 2000 units (2 vials) BW<=75kg / 165 lbs
- 3000 units (3 vials) BW>75kg / 165 lbs

3b. Kit retrieved from EMS rig? *radio*
- Yes
- No

4a. No. of Vials Reconstituted *integer [0-3]*

Skip to 5a if answer matches Question 3

4b. If number of vials reconstituted doesn't match the intended dose – explain: *notes*

5a. No. of Vials Added to IV Bag *integer [0-3]*

5b. If number of vials reconstituted doesn't match the number of

e-CRF

EMS HAND-OFF REPORT

Complete for all subjects screened

Check if patient was randomized

SECTION 1. EMS AGENCY

EMS Agency: acme

EMS Rig No.: 22

SECTION 2. TIMES REPORTED BY EMS

Injury: 00:30

EMS Notification: 01:00

Responding Unit Arrived at Patients Side: 01:30

Study Drug Start: HH:MM

ED Handoff / Threshold Crossing Time: 01:58

SECTION 3. STUDY DRUG ADMINISTRATION

New Kit (study vials) Assigned to EMS Rig at ER?
- Yes
- No

If yes, new Study Kit ID

Kcentra Intended Dose
- 2000 units (2 vials) BW <= 75kg / 165lbs
- 3000 units (3 vials) BW > 75kg / 165lbs

Kit retrieved from EMS rig?
- Yes
- No

Number of vials reconstituted

If the number of vials reconstituted doesn't match the intended dose - explain:
- Time to hospital was limited
- Device/drug malfunction
- Clinical care prioritized
- CPR Initiated
- Other
- Number of vials reconstituted is the same as the intended dose

Other, pl

Number of vials added to the IV bag

If the number of vials reconstituted doesn't match the number of vials added to the IV bag:
- Time to hospital was limited
- Device/drug malfunction
- Clinical care prioritized
- CPR Initiated
- Other

Other, pl

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

a.k.a. Manual of Operations & Procedures (MOP) / Standard Operating Procedures

- 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

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

Accuracy

Consistency

Timeliness

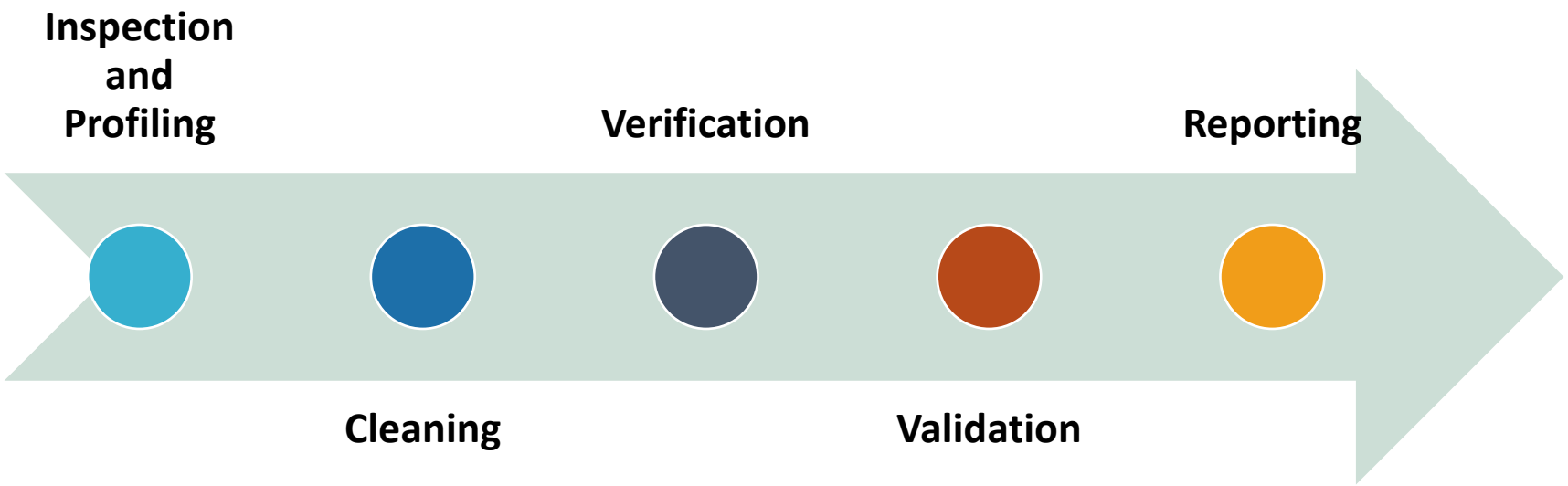
Relevance

Completeness

Validity

Uniformity

Steps for Processing Data



Scrub for Duplicate

Scrub for Irrelevant Data

Scrub for Incorrect Data

Fix Structural Errors

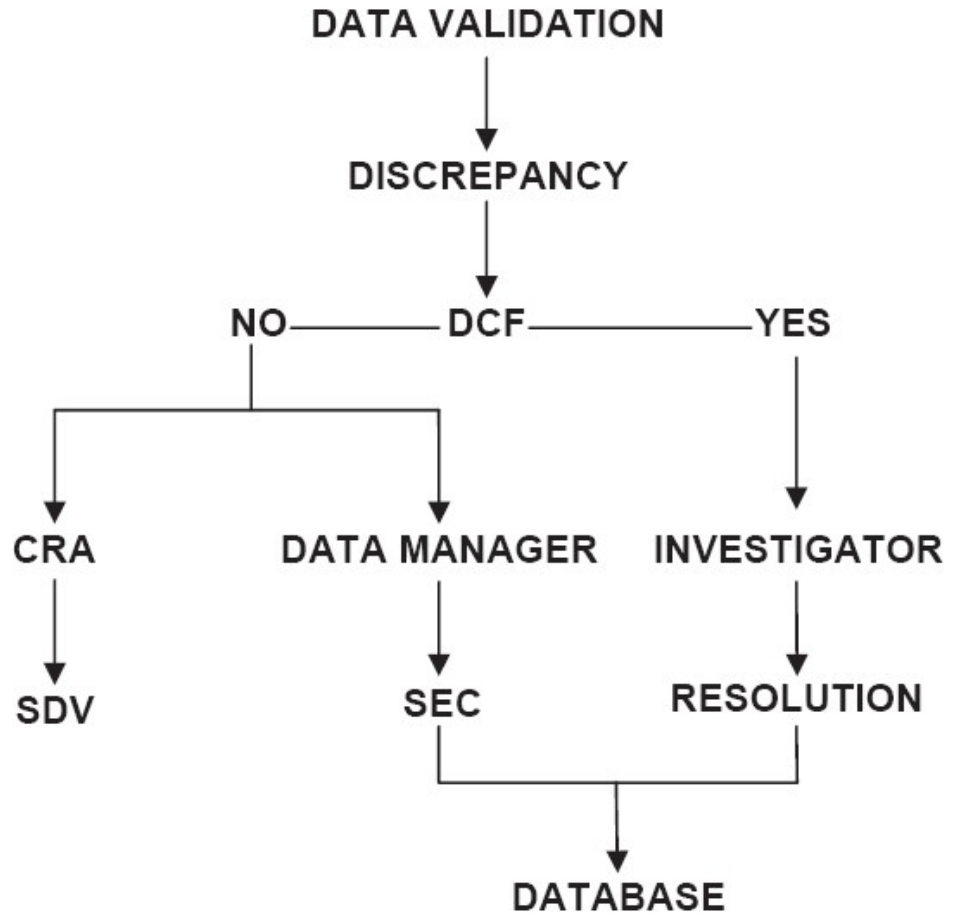
Handle Missing Data

Check the Outliers

Standardize

Normalize

Discrepancy Management Process



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

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.)

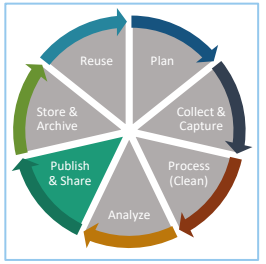
Database Locking

- 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.

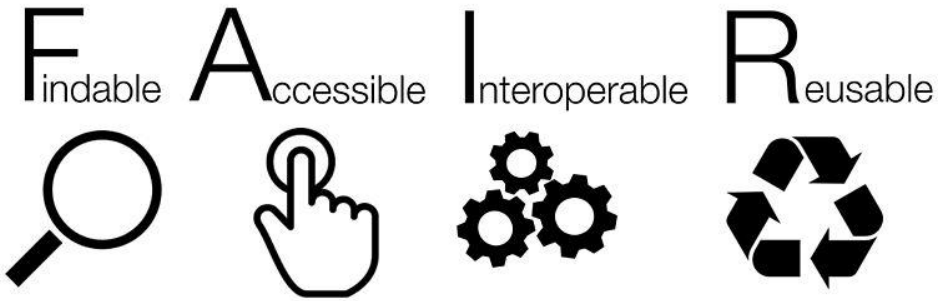
Data Management Lifecycle

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

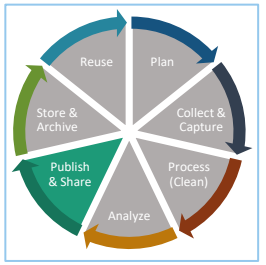
Data Sharing



- 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

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

Data Storage

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

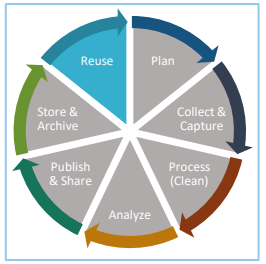
Data Storage & Archival

- 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 https://www.ohsu.edu/sites/default/files/2019-05/OHSU-Summary-of-Regulatory-Retention-Requirements-for-Records.pdf
- Records Retention and Destruction Policy number 07-90-010 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