MEAL — MONITORING, EVALUATION, ACCOUNTABILITY AND LEARNING

CHAPTER 12

DATA MANAGEMENT AND PROTECTION

Secure collection, processing, storage, sharing, retention and disposal of programme data

Dr Abenet Yohannes, Ph.D.

16:9 Widescreen | Suggested training duration: 4–5 hours

Follow: Dr Abenet Yohannes, Ph.D. for more contents

Chapter Introduction

MEAL systems collect and process information about:

- Participants
- Households
- Communities
- Programme activities
- Service delivery
- Complaints and feedback
- Monitoring results
- Evaluations
- Partners and stakeholders

> *These data must be managed accurately and protected responsibly throughout their life cycle.*

Why Data Management Matters

Weak data-management practices may cause:

- Missing records
- Duplicate records
- Incorrect reports
- Unauthorized access
- Loss of confidential information
- Inability to verify results
- Noncompliance with donor requirements
- Harm to programme participants

Why Data Protection Matters

Sensitive information may concern

- Children
- Survivors of violence
- Persons with disabilities
- Refugees and displaced people
- Health conditions
- Household income
- Protection concerns
- Complaints and investigations

Poor protection may expose people to

- Discrimination
- Stigma
- Retaliation
- Exploitation
- Fraud
- Physical or psychological harm

Opening Case

A project stores its beneficiary database on a staff member’s personal laptop. The spreadsheet:

- Is not encrypted
- Uses no password
- Contains names and telephone numbers
- Is shared through personal email
- Has several versions
- Has no backup
- Remains accessible after the employee leaves

> *Question: What data-management and protection weaknesses are present?*

CENTRAL MEAL QUESTION

How can a programme ensure that data remain accurate, usable, traceable, confidential and secure from collection through final destruction?

Learning Objectives

By the end of this chapter, participants will be able to:

- Map a complete data-flow process.
- Apply data-coding and entry procedures.
- Conduct systematic data cleaning.
- Handle missing data appropriately.
- Identify and resolve duplicate records.
- Establish secure data storage and backup.
- Apply file-naming and version-control standards.
- Assign access permissions based on roles.
- Apply encryption and password-protection controls.
- Distinguish de-identification from anonymization.
- Develop data-retention and destruction procedures.
- Prepare data-sharing agreements.
- Respond to data incidents and breaches.
- Integrate organizational and donor requirements.
- Prepare a data-management and data-protection plan.

Chapter Roadmap

- Foundations of data management
- Data-flow mapping
- Data coding and entry
- Data cleaning
- Missing data
- Duplicate identification
- Storage and backup
- File naming and version control
- Access permissions
- Encryption and passwords
- De-identification and anonymization
- Retention and destruction
- Data sharing
- Incident reporting
- Compliance
- Practical planning exercise

Opening Reflection

A dataset can be accurate but still unsafe.
A dataset can also be secure but unusable.

Effective data management requires both:

- Data quality and data protection
- Accessibility and traceability
- Appropriate use and secure disposal

SECTION ONE

1 Foundations of Data Management

Definitions, principles, the data life cycle, classification and accountability roles

What Is Data Management?

Data management is the organized process of:

- Collecting
- Coding
- Entering
- Cleaning
- Storing
- Processing
- Analysing
- Sharing
- Archiving
- Destroying data

> *Its purpose is to keep data accurate, usable, traceable and secure.*

What Is Data Protection?

Data protection consists of the safeguards applied to prevent:

- Unauthorized access
- Unauthorized disclosure
- Accidental loss
- Improper alteration
- Misuse
- Unnecessary retention
- Insecure disposal

Data Management versus Data Protection

Data Management	Data Protection
Organizes the data life cycle	Protects people and information
Supports quality and usability	Supports confidentiality and security
Covers coding and cleaning	Covers access and encryption
Supports analysis and reporting	Prevents unauthorized use
Includes storage and archiving	Includes retention and destruction

The two areas are interconnected.

Data Life Cycle

The programme data life cycle includes:

1. **Planning**
2. Collection
3. Entry or import
4. Validation
5. Cleaning
6. Storage
7. Analysis
8. Reporting
9. Sharing
10. Archiving
11. Destruction

Data-Protection Principles

A responsible data-management system should apply:

- Lawful and authorized processing
- Clear purpose
- Data minimization
- Accuracy
- Confidentiality
- Security
- Limited retention
- Accountability
- Protection from harm

Purpose Limitation

Data should be collected and used only for clearly defined purposes.

Example: a telephone number collected for programme follow-up should not automatically be used for:

- Marketing
- Political communication
- Unrelated research
- Personal contact by staff

Data Minimization

Collect only the data required to achieve the stated purpose. Ask:

- Is this field necessary?
- Will it be used?
- Could the purpose be achieved with less-sensitive data?
- Is an identifiable record required?
- How long must the information be retained?

Data Accuracy

Programmes should take reasonable steps to ensure that data are:

- Correct
- Complete
- Current
- Consistent
- Properly coded
- Updated when necessary

> *Incorrect personal data may directly harm programme participants.*

Confidentiality

Confidentiality means that information is accessed or disclosed only to authorized persons for approved purposes. It applies to:

- Paper documents
- Electronic files
- Conversations
- Reports
- Emails
- Mobile devices
- Photographs
- Audio recordings

Data Availability

Authorized users must be able to access required information when needed. Availability may be threatened by:

- Lost devices
- System failure
- File corruption
- Ransomware
- Inadequate backup
- Forgotten passwords
- Departure of key staff

Data Integrity

Data integrity means that information remains:

- Accurate
- Complete
- Consistent
- Traceable
- Protected from unauthorized alteration

> *A secure system should show who changed data, when and why.*

Data Classification

A programme may classify data by sensitivity, and the classification then drives every control applied to it.

Classification levels

- Public
- Internal
- Confidential
- Highly sensitive

The classification determines

- Storage requirements
- Access restrictions
- Sharing rules
- Encryption requirements
- Retention procedures

Examples of Data Classification

Data Type	Suggested Classification
Published annual report	Public
Internal activity plan	Internal
Beneficiary contact list	Confidential
PSEA allegation record	Highly sensitive
Staff payroll file	Highly sensitive
Anonymous survey summary	Internal or public, as approved

Data Ownership and Stewardship

Data owner — a senior role accountable for

- Authorizing use
- Approving access
- Determining retention
- Approving sharing

Data steward — responsible for

- Data quality
- Documentation
- Storage
- Access administration
- Routine management

Roles and Responsibilities

Data collectors

- Obtain data appropriately
- Follow approved tools
- Protect records in the field

MEAL staff

- Design data structures
- Clean and analyse data
- Maintain documentation

IT staff

- Manage system security
- Backups and user access
- Technical recovery

Managers

- Approve use, sharing and retention
- Provide oversight and resources

SECTION TWO

2 Data-Flow Process

Mapping how information moves from source to storage, reporting and destruction

What Is a Data-Flow Process?

A data-flow process shows how information moves:

- From the original source
- Through collection and processing
- To storage, analysis and reporting
- Until archiving or destruction

Basic Data-Flow Model

1. **Participant or event**
2. Data-collection tool
3. Data entry or synchronization
4. Validation and cleaning
5. Approved database
6. Analysis
7. Report or dashboard
8. Archive or destruction

Why Map Data Flow?

Data-flow mapping helps identify:

- Who handles the data
- Where data are stored
- How files are transferred
- Where errors may occur
- Where unauthorized access may occur
- Which approvals are required
- Which systems contain copies

Data-Flow Mapping Questions

Ask:

- What data are collected?
- Why are they collected?
- Who collects them?
- Which tool is used?
- Where are they transferred?
- Who enters or reviews them?
- Where are they stored?
- Who receives reports?
- When are they archived or destroyed?

Paper-Based Data Flow

1. **Paper questionnaire**
2. Supervisor review
3. Secure transportation
4. Data entry
5. Electronic validation
6. Paper filing
7. Secure archive or destruction

Electronic Data Flow

1. **Mobile data-collection device**
2. Encrypted transmission
3. Secure server
4. Validation
5. Clean database
6. Approved analysis
7. Restricted reporting output

Hybrid Data Flow

A hybrid system may combine:

- Paper registers
- Electronic spreadsheets
- Mobile applications
- Central databases
- Cloud-based dashboards

> *Risk: multiple systems may create inconsistent versions and uncontrolled copies.*

Data-Flow Risk Points

Common risk points include:

- Field collection
- Transportation
- Personal devices
- Email attachments
- Data entry
- Shared folders
- Cloud storage
- Report preparation
- External sharing
- Staff departure

Data-Flow Control Matrix

Stage	Main Risk	Control
Collection	Unauthorized viewing	Private interview and secure device
Transfer	Interception or loss	Encrypted transmission
Entry	Typing errors	Validation and double-entry checks
Storage	Unauthorized access	Role-based permissions
Analysis	Use of wrong version	Version control
Sharing	Excessive disclosure	De-identification and approval
Disposal	Recoverable deleted files	Secure destruction

Data Inventory

A data inventory lists all datasets held by the programme. It should include:

- Dataset name
- Purpose
- Data owner
- Storage location
- Sensitivity level
- Access roles
- Retention period
- Sharing arrangements
- Disposal method

SECTION THREE

3 Data Coding and Entry

Codebooks, variable naming, entry procedures and validation controls

What Is Data Coding?

Data coding converts information into standardized values suitable for entry, processing and analysis. Examples:

- Female = 1
- Male = 2
- Yes = 1
- No = 0
- District codes
- Service-type codes
- Missing-value codes

Purpose of Coding

Coding supports:

- Consistent data entry
- Faster analysis
- Reduced spelling variation
- Automated validation
- Easier filtering
- Easier aggregation
- Improved comparability

Codebook

A codebook documents:

- Variable name
- Variable label
- Data type
- Permitted values
- Value labels
- Missing-data codes
- Validation rules
- Units of measurement

Sample Codebook

Variable	Label	Type	Coding
participant_id	Unique participant code	Text	Project-generated ID
sex	Sex of participant	Numeric	1 = Female; 2 = Male
age	Age in completed years	Numeric	0–120
disability	Disability status	Numeric	1 = Yes; 0 = No
district	Implementation district	Numeric	Approved district codes
service_date	Date of service	Date	DD/MM/YYYY

Good Variable Names

Variable names should be short, clear, unique, consistent, free from unnecessary spaces and understandable to authorized users.

```
participant_id
```

```
service_date
```

```
age_years
```

```
district_code
```

```
training_completed
```

Poor Coding Practices

Avoid:

- Different codes for the same response
- Unexplained abbreviations
- Using 0 for both “No” and “Missing”
- Changing codes between periods
- Mixing text and numbers
- Using colour as the only code
- Failing to maintain a codebook

Data-Entry Procedures

A standard data-entry procedure should define:

- Who enters data
- Which source documents are used
- Entry deadlines
- Validation rules
- Correction procedures
- Quality checks
- Approval requirements
- File-storage location

Manual Data Entry

Manual entry may involve:

- Spreadsheet entry
- Database entry
- Entry from paper forms
- Entry from registers
- Import from partner templates

> *Main risk: human transcription error.*

Electronic Data Collection

Electronic tools can reduce error through:

- Required fields
- Range checks
- Skip logic
- Automated calculations
- Duplicate alerts
- Time stamps
- Geographic validation
- Direct synchronization

> *Technology does not eliminate poor indicator definitions or unethical collection practices.*

Data-Entry Validation

Data-entry controls may include:

- Drop-down lists
- Range restrictions
- Required fields
- Date restrictions
- Duplicate warnings
- Cross-field checks
- Automatic totals
- Unique identification rules

Data-Entry Review

Before records are accepted:

- Compare with source documents.
- Check completeness.
- Confirm valid codes.
- Review unusual values.
- Confirm dates.
- Check duplicate identifiers.
- Verify calculated fields.
- Document corrections.

SECTION FOUR

4 Data Cleaning

Systematic identification and documented resolution of errors

What Is Data Cleaning?

Data cleaning is the systematic identification and resolution of errors, inconsistencies and incomplete records. It prepares data for:

- Analysis
- Reporting
- Visualization
- Sharing
- Archiving

Common Cleaning Tasks

- Correct invalid codes
- Standardize spelling
- Resolve missing values
- Remove duplicate records
- Verify outliers
- Correct date formats
- Recalculate derived fields
- Reconcile disaggregated totals
- Review inconsistent responses

Data-Cleaning Process

1. **Preserve the raw dataset.**
2. Create a working copy.
3. Apply automated checks.
4. Review flagged records.
5. Compare with source evidence.
6. Correct only with authorization.
7. Document every material change.
8. Save an approved clean version.

Preserve Raw Data

The original dataset should not be overwritten. Maintain:

- Raw data
- Cleaning log
- Cleaned data
- Analysis file
- Final approved dataset

> *This supports transparency and reproducibility.*

Data-Cleaning Log

Record ID	Variable	Original Value	Corrected Value	Reason	Evidence	By / Date

The log should explain significant changes made to the dataset.

Range Checks

A range check determines whether values fall within permitted limits.

Eligible age: $6 \leq \text{Age} \leq 17$

Flag records where:

- Age is below 6
- Age is above 17
- Age is negative
- Age is missing

Consistency Checks

Examples:

- End date must be after start date.
- Number completing cannot exceed number enrolled.
- Total participants must equal disaggregated categories.
- A child’s age should be consistent with date of birth.
- Service completion requires a service-start record.

Outlier Review

An outlier is a value that differs substantially from expected patterns. Examples:

- Age = 170
- Household size = 95
- Training attendance = 5,000 at a small venue
- Service duration = 900 hours

> *An outlier is not automatically an error. It must be investigated.*

Standardizing Text Data

Text values may appear in many forms for the same place:

```
Jijiga
```

```
jigjiga
```

```
JIGJIGA
```

```
Jig-Jiga
```

```
Jgj
```

Use a standard reference list to assign one approved value.

Date Cleaning

Common date problems include:

- Mixed formats
- Impossible dates
- Future dates
- Dates outside the reporting period
- Day and month reversal
- Missing year
- Text stored as a date

Derived Variables

Derived variables are calculated from other fields.

$$\text{Age} = \text{Interview Date} - \text{Date of Birth}$$
$$\text{Total Score} = \Sigma\ \text{Question Scores}$$

Derived variables should be generated using documented formulas rather than manually entered where possible.

SECTION FIVE

5 Handling Missing Data

Recognizing, coding, measuring and reporting missingness

What Are Missing Data?

Missing data occur when an expected value is not recorded. Missingness may result from:

- Refusal
- Nonresponse
- Inapplicability
- Enumerator error
- Data-entry error
- Lost documentation
- System failure
- Intentional omission

Types of Missing Values

Different missing values should be distinguished. Examples:

- Blank = not reviewed
- 97 = not applicable
- 98 = respondent does not know
- 99 = respondent refused

> *Codes must be documented and kept outside valid response ranges.*

Why Missing Data Matter

Missing data may:

- Reduce sample size
- Distort averages
- Bias conclusions
- Prevent disaggregation
- Affect indicator calculation
- Conceal vulnerable groups
- Reduce confidence in results

Missing-Data Rate

Missing-Data Rate = (Missing Required Values ÷ Expected Values) × 100

Example

- Missing disability-status values: 35
- Expected records: 500
- **(35 ÷ 500) × 100 = 7%**

Handling Missing Data

Possible responses include:

- Retrieve the value from source documents.
- Return the form for completion.
- Re-contact the respondent where ethical and feasible.
- Retain the value as missing.
- Exclude the record from a specific calculation.
- Apply an approved statistical method.
- Report the extent of missingness.

Do Not Invent Missing Values

Never fill missing data using:

- Personal assumptions
- Most convenient values
- Target values
- Prior-period values
- Unverified averages
- Informal guesses

> *Any imputation must follow an approved analytical method and be documented.*

Missing-Data Decision Table

Situation	Recommended Action
Value available in source document	Enter and document correction
Data collector omitted required field	Return for correction
Respondent refused	Code as refusal
Question not applicable	Code as not applicable
Value cannot be recovered	Retain as missing
Statistical imputation required	Use approved documented method

SECTION SIX

6 Duplicate Identification

Distinguishing genuine duplicates from legitimate repeat services

What Is a Duplicate Record?

A duplicate occurs when the same person, household, event or service is recorded more than once when the indicator requires unique counting. Duplicates may be:

- Exact
- Partial
- Possible
- Legitimate repeat events

Exact Duplicates

Exact duplicates contain identical values in key fields. Example:

- Same participant ID
- Same name
- Same date of birth
- Same service date
- Same location

> *These are easier to identify automatically.*

Possible Duplicates

Possible duplicates may contain small differences.

Name variations

- Ahmed Mohamed
- Ahmed Mohammed
- A. Mohamed

Other matching fields

- Telephone number
- Date of birth
- Household ID
- Community
- National identification number

Duplicate-Detection Rules

A duplicate rule may combine key fields:

Name + Date of Birth + Location
or
Participant ID + Service Type + Service Date

The rule should reflect the indicator’s unique-counting requirement.

Duplicate Review Process

1. **Run an automated duplicate check.**
2. Produce a list of possible matches.
3. Compare supporting records.
4. Determine whether the records are duplicates.
5. Retain legitimate repeat services.
6. Remove or merge confirmed duplicates.
7. Document the decision.

Duplicate versus Repeat Service

A person may legitimately receive several services — these are not automatically duplicate records.

Example: one participant receives

- Registration
- Training
- Referral
- Follow-up

Key question — does the indicator count

- Unique people?
- Service contacts?
- Events?
- Items distributed?

Duplicate-Control Measures

- Use unique identifiers.
- Apply duplicate alerts.
- Search before creating a new record.
- Standardize names.
- Capture dates of birth where appropriate.
- Use household identifiers.
- Review cross-site duplication.
- Document merged records.

SECTION SEVEN

7 Data Storage and Backup

Protecting records from loss, theft, corruption and unauthorized access

Secure Data Storage

Storage arrangements should protect data from:

- Unauthorized access
- Device theft
- Physical damage
- File corruption
- Malware
- Accidental deletion
- Environmental damage
- Staff turnover

Paper-Record Storage

Paper records should be:

- Stored in locked cabinets
- Kept in restricted-access rooms
- Protected from fire and water
- Organized for retrieval
- Logged when removed
- Returned promptly
- Securely destroyed when no longer needed

Electronic Storage

Electronic data should be stored in:

- Approved organizational systems
- Secure servers
- Approved cloud environments
- Encrypted devices
- Access-controlled databases

> *Avoid long-term storage on personal devices.*

Backup Purpose

Backups support recovery after:

- Accidental deletion
- Device failure
- File corruption
- Cyberattack
- Theft
- Natural disaster
- Human error

> *A backup is useful only when it can be successfully restored.*

Backup Principles

A backup plan should specify:

- What is backed up
- How often
- Where copies are stored
- Who is responsible
- How backups are protected
- How long backups are kept
- How restoration is tested

Backup Frequency

Frequency should depend on

- Data sensitivity
- Volume of change
- Operational importance
- Recovery needs
- System capability
- Organizational policy

Examples

- Real-time synchronization
- Daily backup
- Weekly full backup
- Monthly archive

Backup Separation

Backups should not all be kept:

- On the same device
- In the same office
- Under the same user account
- In the same unprotected system

> *A system failure or incident should not destroy both the original and the backup.*

Backup Testing

Periodically test:

- Whether the backup opens
- Whether files are complete
- Whether passwords are available
- Whether the correct version is restored
- How long recovery takes
- Whether responsible staff know the procedure

SECTION EIGHT

8 File Naming and Version Control

Making the right dataset easy to find and the wrong one impossible to use

Why File Naming Matters

Weak file naming creates confusion such as:

```
final.xlsx
```

```
final2.xlsx
```

```
final_updated.xlsx
```

```
latest_final_revised.xlsx
```

```
final_final.xlsx
```

A standard naming convention improves traceability.

File-Naming Convention

A standard file name may include:

Project _ Dataset _ Period _ Version _ Date

Example

ECW_Attendance_Q2_2026_v03_20260715.xlsx

Good File-Name Characteristics

File names should be:

- Descriptive
- Consistent
- Short enough to manage
- Free from unclear abbreviations
- Free from confidential personal names
- Organized chronologically
- Linked to version status

Version-Control Stages

Possible version labels include RAW, WORKING, CLEAN, REVIEWED, APPROVED and ARCHIVED.

```
WFP_Distribution_July2026_APPROVED_v01.xlsx
```

The version label tells a user, at a glance, whether a file may be used for reporting.

Version-Control Register

File Name	Version	Status	Prepared By	Reviewed By	Approval Date

The register prevents use of outdated or unapproved files.

Version-Control Rules

- Do not overwrite raw data.
- Maintain one approved master file.
- Restrict master-file editing.
- Record material changes.
- Archive superseded versions.
- Clearly label draft reports.
- Remove obsolete copies from shared locations.
- Use system-based version history where available.

SECTION NINE

9 Access Permissions

Least-privilege access, role-based controls and safe staff offboarding

Access-Control Principle

LEAST-PRIVILEGE ACCESS

Users should receive only the access required to perform their assigned responsibilities — no more, and only for as long as the responsibility lasts.

Types of Access

A user may be authorized to:

- View
- Enter
- Edit
- Export
- Approve
- Share
- Delete
- Administer users

> *Not every user requires every permission.*

Role-Based Access

Role	Suggested Permission
Data collector	Enter assigned records
Supervisor	Review and approve site records
MEAL officer	Clean and analyse approved records
Programme manager	View reports and dashboards
Data administrator	Manage users and systems
External consultant	Time-limited approved access

User-Access Controls

- Use individual accounts.
- Prohibit shared passwords.
- Approve access formally.
- Review access periodically.
- Remove access when staff leave.
- Limit administrator privileges.
- Maintain access logs.
- Investigate unusual activity.

Staff Departure

When an employee leaves:

- Disable user accounts.
- Recover devices.
- Transfer approved files.
- Change shared credentials.
- Review recent downloads.
- Remove cloud access.
- Confirm return of paper records.
- Document completion of offboarding.

SECTION TEN

10

Encryption and Password Protection

Technical safeguards for data at rest, in transit and in transfer

What Is Encryption?

Encryption converts readable data into a protected format that requires an authorized key or credential to access. It may protect data:

- At rest
- During transmission
- On portable devices
- In backups
- In approved cloud systems

Encryption at Rest and in Transit

At rest — protects stored information on

- Servers
- Laptops
- Mobile devices
- External drives
- Backups

In transit — protects data moving through

- Internet connections
- File transfers
- Mobile synchronization
- Email systems
- Application interfaces

Password Standards

Strong passwords should:

- Be sufficiently long
- Be difficult to guess
- Avoid personal information
- Not be reused across systems
- Be changed when compromise is suspected
- Be stored using approved password-management tools

Multi-Factor Authentication

Multi-factor authentication requires more than one form of verification. Examples:

- Password plus authentication application
- Password plus security token
- Password plus biometric verification

> *It provides additional protection when a password is stolen.*

Password-Protected Files

Password protection may be useful for:

- Sensitive spreadsheets
- Reports containing personal data
- Files transferred to authorized partners
- Temporary offline datasets

> *Passwords should not be sent through the same channel as the protected file.*

SECTION ELEVEN

11 De-Identification and Anonymization

Reducing and removing the link between data and identifiable people

Identifiable Data

Direct identifiers include:

- Name
- National identification number
- Telephone number
- Email address
- Exact address
- Photograph
- Biometric information
- Case number linked to identity

> *Indirect identifiers may also identify someone when combined.*

What Is De-Identification?

De-identification reduces the ability to connect data with a particular person. Methods include:

- Removing direct identifiers
- Replacing names with codes
- Generalizing locations
- Grouping ages
- Separating the identification key
- Suppressing rare characteristics

> *De-identified data may still carry a re-identification risk.*

What Is Pseudonymization?

Pseudonymization replaces direct identifiers with a code, held against a separate secure linking file.

Name	Participant Code
Amina Hassan	P-000145

Because the link exists, pseudonymized data are not fully anonymous.

What Is Anonymization?

Anonymization removes or transforms identifiers so that individuals cannot reasonably be re-identified. Properly anonymized data:

- Cannot be linked back through an identification key
- Contain no unnecessary unique combinations
- Are usually used for aggregated analysis or public reporting

De-Identification versus Anonymization

De-Identification	Anonymization
Reduces identification risk	Intends to prevent re-identification
May retain a linking key	No usable linking key remains
May be reversible	Intended to be irreversible
Often used for controlled analysis	Often used for broad sharing
Still requires protection	Lower privacy risk, but must be assessed

De-Identification Techniques

- Remove direct identifiers.
- Replace identifiers with codes.
- Use age bands instead of exact age.
- Report district instead of exact address.
- Aggregate small categories.
- Suppress rare combinations.
- Remove free-text details.
- Review data before release.

SECTION TWELVE

12 Retention, Sharing, Incidents and Compliance

Closing the life cycle responsibly and meeting external obligations

Data Retention and Destruction

A retention schedule defines:

- Which data are retained
- Why they are retained
- How long they are retained
- Where they are archived
- Who approves destruction
- Which destruction method is used

> *Data should not be retained indefinitely without a justified purpose.*

Retention Considerations

- Programme needs
- Donor requirements
- Audit requirements
- Legal obligations
- Consent terms
- Protection risks
- Research value
- Contractual requirements

Secure Destruction

Paper records

- Cross-cut shredding
- Approved pulping
- Controlled incineration where permitted

Electronic records

- Secure erasure
- Cryptographic deletion
- Physical destruction of unusable media
- Verified deletion from backups where required

Destruction Log

Record what was destroyed, when, by whom and under whose authority.

Without a destruction log, a programme cannot demonstrate that expired personal data were disposed of securely and with proper approval.

Data-Sharing Agreements

A data-sharing agreement establishes the conditions under which one organization provides data to another. It should define:

- Purpose of sharing
- Specific data fields
- Legal or organizational authority
- Permitted uses
- Prohibited uses
- Security requirements
- Authorized recipients
- Retention period
- Destruction requirements
- Incident-reporting responsibilities
- Audit and compliance rights

Data-Sharing Principle

Share the minimum information necessary. Before sharing, ask:

- Is sharing necessary?
- Can aggregated data meet the purpose?
- Can identifiers be removed?
- Is the recipient authorized?
- Is the transfer secure?
- Is a signed agreement in place?
- Is onward sharing prohibited or controlled?

Incident and Breach Reporting

A data incident is an event that threatens the confidentiality, integrity or availability of data. Examples:

- Lost laptop
- Email sent to the wrong recipient
- Unauthorized database access
- Stolen paper register
- Malware infection
- Accidental file deletion
- Public disclosure of personal data
- Weak password discovered
- Unapproved data transfer

Immediate Incident Response

1. **Contain the incident.**
2. Preserve evidence.
3. Inform the designated focal person.
4. Assess affected data and people.
5. Determine severity.
6. Recover or secure systems.
7. Notify required internal and external parties.
8. Document actions.
9. Prevent recurrence.

Incident Report Contents

- Date and time
- Person reporting
- Description
- Data affected
- Number of records
- Sensitivity level
- People potentially affected
- Immediate containment
- Root cause
- Notifications
- Corrective actions
- Closure approval

Reporting Culture

Staff should report suspected incidents immediately. Incident reporting should promote:

- Rapid containment
- Transparency
- Learning
- Accountability
- Prevention

> *It should not discourage reporting through unnecessary blame.*

Compliance Requirements

A data-management system should comply with applicable:

- Organizational policies
- Donor requirements
- Partnership agreements
- Consent commitments
- Research protocols
- National laws
- Sector standards
- Contractual obligations

Compliance Matrix

Requirement	Source	Responsible	Evidence	Review
Encryption	Organizational policy	IT Manager	System configuration	Quarterly
Retention period	Donor agreement	Project Manager	Retention schedule	Annually
Access review	Security procedure	Data Manager	Access register	Quarterly
Incident reporting	Organizational policy	Protection Focal Person	Incident log	Continuous
Secure destruction	Records policy	Administration	Destruction certificate	As scheduled

Avoiding Conflicting Requirements

When organizational, donor or legal requirements conflict:

1. **Identify the conflict.**
2. Seek legal, donor or management guidance.
3. Apply the most protective authorized approach.
4. Document the decision.
5. Update the data-management plan.

PRACTICAL EXERCISE

Prepare a Data-Management and Data-Protection Plan

Group work: apply the full chapter to a live education-programme scenario

Exercise Scenario

An education project will register and monitor 5,000 children across ten schools. It will collect:

- Participant identification code
- Child's name
- Age
- Sex
- Disability status
- School
- Attendance
- Learning-assessment results
- Parent or guardian telephone number
- Referral information

> *Data will be collected electronically by field officers and synchronized to a central database.*

Participant Tasks

Prepare a plan covering:

- Data purpose
- Data-flow process
- Data classification
- Coding and entry
- Cleaning and validation
- Missing-data handling
- Duplicate identification
- Storage and backup
- File naming and version control
- Access permissions
- Encryption and password protection
- De-identification and anonymization
- Retention and destruction
- Data sharing
- Incident reporting
- Compliance monitoring

Sample Plan: Dataset and Purpose

Item	Plan
Dataset	Education participant monitoring database
Purpose	Registration, attendance monitoring and learning-outcome measurement
Data owner	Project Director
Data steward	MEAL Manager
Sensitivity	Confidential; referral information highly sensitive

Sample Plan: Data Flow

1. **Child and guardian**
2. Encrypted mobile form
3. Secure central server
4. Automated validation
5. MEAL data review
6. Approved clean dataset
7. De-identified analysis
8. Aggregated donor report

Sample Plan: Coding and Entry

- Use unique participant identification codes.
- Apply standard school and district codes.
- Use controlled values for sex and disability status.
- Use date restrictions.
- Use required fields for critical variables.
- Use electronic synchronization rather than personal email.

Sample Plan: Cleaning and Validation

- Run weekly missing-value checks.
- Review ages outside the eligible range.
- Match totals with school attendance registers.
- Check duplicate participant codes.
- Validate assessment-score ranges.
- Record corrections in a cleaning log.
- Preserve the raw dataset.

Sample Plan: Missing Data and Duplicates

Missing data

- Retrieve values from source documents where possible.
- Code refusal separately from “not applicable.”
- Do not invent missing values.
- Report missingness in analytical outputs.
- Escalate missing critical fields to supervisors.

Duplicate identification — match on

- Participant ID
- Name
- Date of birth
- School
- Guardian telephone number
- Review possible matches before merging or removing.

Sample Plan: Storage and Backup

- Store the master database on an approved secure server.
- Prohibit storage on personal laptops.
- Apply daily automated backup.
- Maintain a separate encrypted backup.
- Test restoration quarterly.
- Lock paper consent forms in restricted cabinets.

Sample Plan: File Naming and Version Control

Maintain separate raw, working, clean, approved and archived versions.

```
Education_Attendance_Q2_2026_CLEAN_v02_20260715.xlsx
```

The status label and version number together identify which file may be used for reporting.

Sample Plan: Access Permissions

Role	Access
Field officer	Enter assigned-school data
Supervisor	Review and approve school records
MEAL Officer	Clean and analyse
MEAL Manager	Approve final dataset
Project Director	View aggregate reports
IT Administrator	Technical administration only

Sample Plan: Security

- Encrypt devices and databases.
- Require strong unique passwords.
- Apply multi-factor authentication.
- Disable inactive accounts.
- Record user-access logs.
- Prohibit sharing of passwords.
- Use secure approved transfer channels.

Sample Plan: De-Identification

Before analysis or sharing:

- Remove names and telephone numbers.
- Use participant codes.
- Generalize small geographic categories.
- Remove free-text identifying information.
- Keep the code key in a separate encrypted location.
- Share aggregated data whenever possible.

Sample Plan: Retention and Data Sharing

Retention and destruction

- Retain identifiable records for the approved organizational or donor period.
- Review retention annually.
- Destroy records after the approved period.
- Securely erase electronic data and shred paper.
- Maintain a destruction register.

Data will be shared only when

- The purpose is documented and approved.
- A sharing agreement is signed.
- Data are minimized and identifiers removed where possible.
- Transfer is encrypted.
- The recipient accepts retention and destruction conditions.

Sample Plan: Incidents and Compliance Monitoring

Report incidents immediately to

- MEAL Manager
- IT Manager
- Data-protection or safeguarding focal person
- Project Director, where required
- The team will contain, assess risk, notify, recover data and document corrective action.

The plan will be reviewed

- Before data collection
- Quarterly
- After major system changes
- Following a data incident
- When donor or organizational requirements change

Chapter Summary (1 of 2)

- Data management covers the full data life cycle.
- Data protection safeguards confidentiality, integrity and availability.
- Data-flow mapping identifies systems, users, transfers and risks.
- Coding creates standardized values for entry and analysis.
- A codebook documents variables and permitted values.
- Raw data must be preserved.
- Data cleaning should be systematic and documented.
- Missing data should never be replaced through unsupported assumptions.
- Duplicate detection must reflect the indicator’s counting rules.
- Storage should use approved organizational systems.
- Backups must be protected and tested.
- File naming and version control prevent use of incorrect datasets.

Chapter Summary (2 of 2)

- Access should follow least-privilege principles.
- Encryption protects data at rest and in transit.
- Passwords should be strong, unique and appropriately managed.
- De-identification reduces identification risk.
- Pseudonymized data may still be linked back to individuals.
- Anonymized data should not permit reasonable re-identification.
- Retention periods must be justified and documented.
- Secure destruction prevents recovery of expired information.
- Data-sharing agreements define permitted use and security obligations.
- Data incidents require immediate containment, reporting and follow-up.
- Compliance must integrate organizational, donor, contractual and applicable legal requirements.

Data-Management and Protection Cycle

1. **Define the purpose**
2. Minimize data collection
3. Map the data flow
4. Code and validate data
5. Clean and approve the dataset
6. Store and back up securely
7. Control access
8. De-identify before sharing
9. Retain only as required
10. Destroy securely
11. Review compliance and incidents

KEY MESSAGE

Responsible data management does not end when a report is produced. Every dataset must remain accurate, traceable and appropriately protected from the moment it is collected until its authorized destruction.

THANK YOU

Questions and Discussion

Dr Abenet Yohannes, Ph.D.

MEAL — Monitoring, Evaluation, Accountability and Learning

Ethical MEAL protects both the quality of evidence and the dignity, privacy and safety of the people represented by the data.

Follow: Dr Abenet Yohannes, Ph.D. for more contents