FEED THE FUTURE
The U.S. Government's Global Hunger & Food Security Initiative

Data Collection, Quality and Management

Robin Shrestha

USAID
FROM THE AMERICAN PEOPLE

Patan Academy of Health Sciences

JOHNS HOPKINS
BLOOMBERG SCHOOL
of PUBLIC HEALTH

Tufts
UNIVERSITY
Friedman School
of Nutrition Science
and Policy

TUSKEGEE
UNIVERSITY

U.S. GOVERNMENT PARTNERS

USAID
FROM THE AMERICAN PEOPLE

MILLENNIUM CHALLENGE CORPORATION
UNITED STATES OF AMERICA

USDA

OPIC

DEPARTMENT OF STATE
UNITED STATES OF AMERICA

USGS
science for a changing world

DEPARTMENT OF COMMERCE
UNITED STATES OF AMERICA

EXECUTIVE OFFICE OF THE PRESIDENT OF THE UNITED STATES
OFFICE OF SCIENCE AND TECHNOLOGY POLICY

PEACE CORPS

THE DEPARTMENT OF THE TREASURY
1789

UNITED STATES AFRICAN DEVELOPMENT FOUNDATION

OUTLINE

- Biological sample collection
- Data Management Plan

BIOLOGICAL SAMPLE COLLECTION

Why collect biological samples?

- Complement survey data, anthropometric and clinical data
- Enables us to find out things that the participants may not know
  - NCDs
  - Vitamin levels
  - Inflammatory markers
  - Minerals –
  - Blood Lipids
  - Toxins – Bacterial, viral and fungal

FEED THE FUTURE
The U.S. Government's Global Hunger & Food Security Initiative

OUR FIELD EXPERIENCE IN COLLECTING CLINICAL & BIOLOGICAL SAMPLES

USAID
FROM THE AMERICAN PEOPLE

Patan Academy of Health Sciences

JOHNS HOPKINS
BLOOMBERG SCHOOL
of PUBLIC HEALTH

Tufts
UNIVERSITY
Friedman School
of Nutrition Science
and Policy

TUSKEGEE
UNIVERSITY

FEED THE FUTURE
The U.S. Government's Global Hunger & Food Security Initiative

BIOLOGICAL SAMPLE COLLECTION IN INFANTS

USAID
FROM THE AMERICAN PEOPLE

Patan Academy of Health Sciences

JOHNS HOPKINS
BLOOMBERG SCHOOL
of PUBLIC HEALTH

Tufts
UNIVERSITY
Friedman School
of Nutrition Science
and Policy

TUSKEGEE
UNIVERSITY

FEED THE FUTURE
The U.S. Government's Global Hunger & Food Security Initiative

USAID
FROM THE AMERICAN PEOPLE

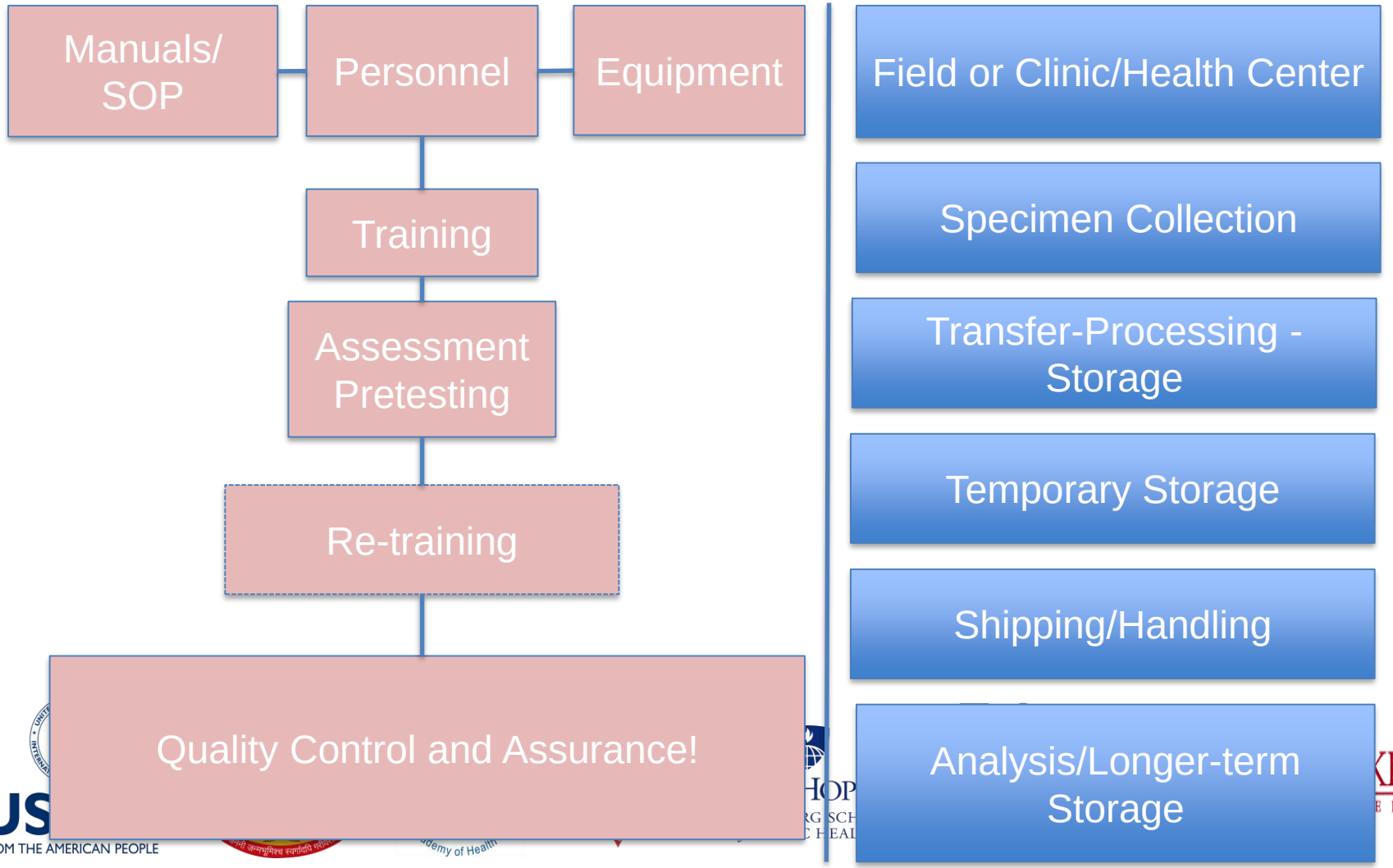
Patan Academy of Health Sciences

JOHNS HOPKINS
BLOOMBERG SCHOOL
of PUBLIC HEALTH

Tufts
UNIVERSITY
Friedman School
of Nutrition Science
and Policy

TUSKEGEE
UNIVERSITY

BIOLOGICAL SAMPLE COLLECTION



CHALLENGES

- Procurement – Supplies, Shipping
- Personnel – Cost, training and re-training
- Study site – Collection of samples in the field
- Collection –
  - Personnel - Quality assurance, techniques
  - Participants - reluctance, drop outs; difficult collection in infants and children
- Transfer and Processing – field level, cold chain maintenance
- Storage – Space, freezer (-80/-20 C)
- Shipping and Analysis – Resources, cold chain maintenance

THINGS TO CONSIDER

- It is possible to collect biological data within a study, but it is important to consider:
  - What you want to collect
  - Who will collect (interviewer, nurse, phlebotomist):
  - Ethical considerations
  - Training is vital – quality is important!
  - Resources: equipment - ease of use, portability, durability
  - Cold Chain - maintain quality of the samples!

DATA MANAGEMENT

WHY MANAGE RESEARCH DATA?

- **Research Perspective -**
  - Fulfill research's impact/objectives
    - Linking data in analysis and publication
    - Making data citable
  - Maintain research integrity
  - Regulatory requirements – Funding agencies, ethical committee, IRBs
  - Data re-use and replication of results

- **Researchers' perspective -**
  - Managing and sharing data is simply part of good research
  - Reputational risks if data management is not handled properly

WHAT IS A DATA MANAGEMENT PLAN (DMP)?

- Is a document that outlines a road map to manage research data that -
  - **Describes** the database design, data entry, cleaning and tracking guidelines, quality control measures, ethical and adverse events guidelines, data transfer/extraction, database sharing

WHILE CONSIDERING DATA MANAGEMENT PLANNING IN RESEARCH

- Know your **legal, ethical and other obligations** regarding research data, participants, researchers, collaborators, institutions and funders
- Design data management plans according to **needs of the research**
- Assign **roles and responsibilities** to relevant parties in research
- Implement **good practices** in a consistent manner
- **Implement and review** data management plans throughout the research process/cycle

FEED THE FUTURE
The U.S. Government's Global Hunger & Food Security Initiative

DATA MANAGEMENT PLAN MANUAL

Table of Contents

Background .....

Roles and Responsibilities of Data Management Team .....

Data Collection and Upload .....

Data Download and Organization .....

Data Versioning .....

Data Cleaning and Quality Assurance .....

- Error identification .....
- Recommended software for data cleaning .....
- Data error report and correction .....
- Correcting data after enumerators respond .....

Data Manipulation .....

Data Merging .....

Data Documentation, Storage and Backups .....

- Raw data .....
- Clean data .....

Reporting structure .....

APPENDIX .....

- General Data Cleaning and Analysis Guidelines .....

ELEMENTS OF DATA MANAGEMENT
ROLES AND RESPONSIBILITIES

- Roles and responsibilities are assigned and not simply assumed.
- People involved in data management:
  - Project Director/Principal Investigator – designing research
  - Research staffs (investigators, coordinators, research and data managers, analysts, involved throughout the research cycle)
  - External data centers or archives who facilitate data sharing

ELEMENTS OF DATA MANAGEMENT

DATA DOCUMENTATION

- Explains how data were **created**, what they mean, content and structure, how they are **organized** and **managed**
- **Good data documentation** includes:
  - Research Background – aims, objectives and hypotheses
  - Research design and methods
  - Data structure
  - Quality Assurance – Validation, checking, proofing and cleaning
  - Changes made to data (different versions)
  - Data confidentiality

DATA-LEVEL DOCUMENTATION

- Structure
  - Name, number of records, etc.
- Variables
  - Name, values, coding, etc.
- History
  - Creation, modification
- Storage information
  - Media, location, back-up, format – SPSS, Stata
- Additional Information – Creating Metadata that explains origin, purpose, time reference, geographic location, and terms of use of data collection. For search and bibliographic record for citation

STRUCTURE OF A DATABASE

- Line/Row represent records
- Column represent variables

	Identifier	Variable 1	Variable 2	Variable 3	…Variable n
Record 1					
Record 2					
Record 3					
..Record n					

IDENTIFIER IN THE DATABASE

- Unique
- May contain all information about that participant ID
- Maintained by computerized index

STRUCTURE OF THE VARIABLES

- Variables can be entered as
  - Integer – number of digits
  - Numeric – number of decimals
  - Alpha-numeric
  - Dates (specific format)

CREATING VARIABLE NAMES

- **Clear**
  - Need to refer to the questionnaire item
  - Understandable
- **Short, no space**
  - Some software may require less than 10 characters
- **Consistent**
  - "EXERPAST" for "Exercise daily in the past"
  - "EXERCURRDLY" for "Exercise daily in the current"
  - "VARIAB" for all crude variables (EXERCISE)
  - "VARIAB_12" for all dichotomized variables (EXERCISE_12)
- **No duplicates**
  - Research assigns variable names

DESIGN DATA ENTRY-FRIENDLY DATA COLLECTION INSTRUMENT

- **Data collection instrument should facilitate data entry**
- Outline/Sections for variables
  - Identifiers (Personal information)
  - Demographics
  - Outcomes
  - Exposures (including confounders)
- Auto-coding function

CODING

- Prefer numerical coding
- Decide on
  - Missing values (. Or 0)
  - Not applicable
- Avoid cumbersome codes
- Uniform coding for dichotomized variables (1 for yes, 0 for no)

CONSTRUCTING A DATA DICTIONARY

- Each variable will have
  - Variable name
  - Description of questionnaire item
  - Various values of variable
  - Meaning of each value

- This is particularly useful when:
  - Researcher need to get back to the database later
  - When a database is shared with others

EXAMPLE OF DATA DICTIONARY

Number	Question	Variable name in Questionnaire	Value	Value label	Type of variable	Labeled?	N	# of "DK"	any ouside allowed range?
1.1.2	Date of interview (Day)	DAYU			numeric	no		0	no
1.1.2	Date of interview (Month)	MONTHU			string	no		0	no
1.1.2REC	Date of interview (Month) (recoded)	MONTHU_REC			numeric	yes		0	no
1.1.2	Date of interview (Year)	YEARU			numeric	no		0	no
1.1.2	Date of interview Gregorian calendar	DATEADTP1U							
1.1.3	VDC	HVDCU			string	no		0	no
1.1.6	Index Woman ID	HHWID			numeric	no		0	no
1.1.10	USG ID 1	HID1U			numeric	no		0	no
13.7.1	Ultrasound draw done?	ULTRA	0	No	numeric	yes		0	no
			1	Yes					

CHECK SPECIFICATIONS BEFORE DATA ENTRY

- Minimum and Maximum values
- Legal codes
- Skip Patterns
- Automatic coding
- Copying data from preceding record
- Calculations

DATA ENTRY

- Use as opportunity for partial data cleaning
  - Write comments
  - Seek clarification
- Use checks
- If using paper based, mark each paper as data entry is complete
- Validate after data entry

DATA STORAGE

Keep your data safe, secure and recoverable

- Making back-ups – protects against software faults, viruses and hacking, hardware, power failures, etc. Need to be careful with the number of back ups- minimal number of copies where data contain identifiable information
- Institutional Back up policy

DATA SECURITY

ENSURE…

- Physical security – authorized access to rooms and buildings
- Network security – not storing confidential data on computers, firewall protection to avoid viruses
- Computer system security – password protected, not sending data over email,

PREVENT…

- Unauthorized access
- Data disclosure
- Destruction of data – paper format

DATA SHARING

Sharing data within research members and collaborators can be challenging -

- Inform participants during consenting process
- Data Sharing Agreement with partners and collaborators
- De-identification – Anonymizing sensitive and personal information for ethical and legal reasons
- Encryption – to ensure security of personal data
  - Whole Disk Encryption – protects sensitive information in case a laptop is stolen or lost
- Cloud based data sharing (Dropbox/Tufts Box) – Transferring large files securely

DATA MANAGEMENT PLAN

- Example of Components of DMP –
  - Data generated by project – What?
  - Institution and contact person responsible for data – Who?
  - Data description
  - Data privacy and use restrictions
  - Pre-submission data processing
  - Final data deliverable
  - Timeline
  - Data repository and post-award curation
  - Responsible party
  - Target submission date
  - Associated costs

FEED THE FUTURE

The U.S. Government's Global Hunger & Food Security Initiative

www.feedthefuture.gov

USAID
FROM THE AMERICAN PEOPLE

Johns Hopkins
Bloomberg School
of Public Health

Tufts University
Friedman School
of Nutrition Science
and Policy