

# Ethiopian Public Health Institute



## *National Data Management Center for health (NDMC)*

*1<sup>st</sup> version*

Data Implementation Guideline

June, 2019  
Addis Ababa



## Contents

Acknowledgement.....	ii
Acronyms.....	iii
PART I: Database Information.....	1
Introduction.....	1
Sources of Health Data.....	1
Processes and Documentation.....	2
Metadata Standard.....	2
Terms and Definitions.....	2
Data Information Form (DIF).....	6
Codebook summary.....	7
PART II: Database organization plan.....	8
Dataset Categories.....	8
Classification by data Sources.....	8
Classification by Cause of death or injury.....	8
Classification by Study design.....	9
Classification by Demographic groups.....	9
Classification by Geospatial.....	9
Classification by Study Type.....	9
Sample Relational Tables.....	10
Sample Relational Model.....	12
PART III: Data Archival Processes.....	13
Existing/Retrospective data.....	13
EPHI dataset.....	13
Other datasets.....	14
Prospective data.....	15
PART IV: Data security.....	16
Establish a digital data security procedure.....	16
Data protection.....	17
PART V: Internal framework/procedure for sharing data to a third party.....	19
Flow chart for data sharing.....	20
Raw Data Access Request Form.....	21
References.....	26



## Acknowledgement

This Data Implementation Guideline has been prepared with full engagement of NDMC Staff members, and invaluable involvement of Fellows from Institute of Health Metrics and Evaluation, Washington University.



## Acronyms

AGLS	Australian Government Locator Service
CSA	Central Statistical Agency
DIF	Data Information Form
DDG	Deputy Director General
EPHI	Ethiopian Public Health Institute
GBD	Global Burden of Diseases
NDMC	National Data Management Center for health
PI	Principal Investigator



## PART I: Database Information

### Introduction

The quality and credibility of health evidence generated is mostly a portrayal of the quality and standard of the data used and partly the scientific methods applied to make sense out of them. Maintaining quality of data is vital. Provided that data submitted to central database has passed through prior Data Quality Assurance Mechanisms and procedures, the other most important component of quality data is a well-defined data dictionary.

Data dictionary is a structured data that stores metadata (data about data). This metadata, therefore, must be standardized by defining its characteristics and attributes best comprehensive and informative for any owner or user privileged to access conditional on institutional data submission, sharing and access policy settings.

Thus, it is demanding to standardize concepts and definitions and underpin the metadata that sustainably describe a certain data or dataset adhered to the standards. Metadata being based in its most efficient structural and content standard, facilitates interoperability to systems and reuse of the data for further research and evidence synthesis by making data more retrievable, identifiable and generally discoverable to researchers and institutions and raising motivation of any to make use of them and enhance health research.

The purpose of this database information is to present the Metadata Standard of the database organized by the National Health Data Management Center of EPHI. The standard contains a set of structured elements that support several functions such as identification, location, classification, indexing (attaching identifier protocol number), technical procedures of data production, management of user access and tracking of use of data.

### Sources of Health Data

In order to develop national health database, the collection and archival of the health and health related datasets is mainly consider sources from

- i. Datasets produced by EPHI
- ii. Datasets produced by academic institutions, health research partners and individual researchers

Integration of health research datasets and identifying theme specific, for instance, GBD specific datasets for disease burden estimation, are then worked through data management systems such as data warehousing and data mining procedures.

Thus, organization of EPHI's database takes those and other related considerations into account. This requires efficient, easy and sustainable database.



A step-by-step development and maintenance of the database would be helpful to have the intended quality and services of the database. To that effect, the database organization, collection and archival system will be maintained in two main phases. The first phase is Database for data produced in EPHI research projects and the second phase is working on that data produced by academic institutions and health system collaborators, institutions and individual researchers. Finally, the two databases will be integrated and archived under EPHI database.

## Processes and Documentation

The health database organization requires several processes and documentations. Specific deliverables that facilitate this task shall include, but not limited to:

- i. Mapping, collecting and archiving accordingly
- ii. Inventory of datasets produced overtime within EPHI and other health research institutions as well
- iii. Developing structure of the database to facilitate the collection and archival of datasets as well as enhance mapping and searching for datasets, and data identification, access and sharing for further analysis.
- iv. Derivation of standard and conventional naming and labeling of database items and components, and creation of relationships among several types and categories of datasets using relational keys or identification codes is also part in this phase.
- v. Abstraction of health research projects in EPHI and other health research institutions
- vi. Abstraction of detail metadata has to be done in times when datasets are being collected and archived in an agreed and standard file formats.

## Metadata Standard

Metadata Standard is “specifications for the minimum information that should be collected about research data in order for it to be re-used”. The metadata is prepared on the basis of three main types – descriptive metadata which contains elements that describe data such as title, author, subject, location; structural metadata which informs documents relationship within and among objects through elements such as links to other components; and administrative metadata for instance that mentioning version number, archiving date, other technical information that can be used for file and rights management and preservation.

## Terms and Definitions

**Data/Dataset** is an object that consists of values or observations defined by rows and columns where mostly the rows represent subjects of a study and columns define variables or measures of attributes of interest of a certain health study. In some instances, rows could define variables and columns may represent subjects under investigation.

**Variables** in a data dictionary are labeled with a unique name, and, that is no two variables in a single dataset can have the same variable name. All data elements or variables in a data, including, unique



identifiers or keys, must be defined clearly being informative and self-explanatory to owners and external users.

**Data elements** are those variables, fields, columns, attributes or characteristics that a data contains. In this document, “variables” might be frequently used interchangeably with data element in defining the standard of health metadata.

**Content Standard:** A Standard that defines elements users can expect to find in metadata and the names and meaning of those elements.

**Format Standard:** A Standard that defines the structures and formats used to represent or encode elements from a content standard.

**Metadata** and also **data dictionary** are as discussed in the introduction. Thus, a data dictionary contains tables about the name, meaning or semantics, relationships to other data, origin, usage, format and other attributes or data elements of a certain health research data.

The properties and standards reflected or set for a data are therefore equivalently interpreted to that of metadata or data dictionary.

**Codebook** is detail about variables names, variable label, variable values/codes, value labels, ranges of values and data types such as integer, byte, text/string, etc.

**Variable definition** gives clear definition about the name and meaning of a variable or data element and the definitions should follow the following principles.

- **Precise** - use words that have a precise unambiguous meaning that avoids words that lead to multiple meanings
- **Concise** - use the shortest and clear description possible
- **Non-Circular** - not use the term being defined in the definition itself
- **Distinct** - differentiate a variable from other variables
- **Unencumbered** - be free of embedded rationale, functional usage, domain information, or procedural information, acronyms
- **Consistent** – be related with concepts of a subject matter

**Variable names** are names given to an attribute in short hand term to identify an item or characteristics of an entity/or subject under study/or investigation.

**Variable label** is a short but precise description of a variable name.

**Variable values** are well defined codes representing categories of possible responses to a certain question or variable.

**Value labels** are brief descriptions or names of defined variable values or codes.



**Range of values** is a set of values with defined minimum and maximum values.

**Dataset name** is a file name given to a dataset as per EPHI/NDMC naming protocol.

**Dataset label** is a full name that describes a dataset file with a project name or shorter precise explanation.

**Date of release/publish** is the date that an item such as a dataset, metadata, report, article or any relevant document attached have been released/published and approved for use.

**Publisher** is an institute or researcher (if individual) by whose name an item is published.

**Principal Investigator (PI)** is the main researcher leading a certain research project.

**Co-Principal Investigator (Co-PI)** is a researcher leading a certain research project being responsible next to a PI.

**Contributors** are researchers or members of a certain research project team who participates in the production of a research outcomes, such as dataset, other publication, etc.

**Summary** is a brief description about objectives, method, design, study area and survey round, data content, and ethical considerations of a certain research project and/or dataset. A summary may follow a set of important terms in the sample below.

*“The dataset <<Datase Name>> is generated due to the project implementation of <<Complete project name>> in the year <<YEAR>>. The objective is <<OBJECTIVES>>. The <<METHOD>> is used and covers <<TARGET POPULATION>> and <<LOCATION>>. Major variables and indicators include <<VARIABLES and INDICATORS>>. <<Suggested level of and scope data analysis>>.”*

**Suggested citation** is preferable citation that publisher or producer of a dataset or publication would recommend.

**Recommended acknowledgement** is a statement that states and acknowledges contribution and supports of bodies and/or individuals who provide grant or a resource relevant for a study.

**Dataset Information** is a set of required details and descriptions that briefly explain and inform about a single final dataset.

*Data Information must be carefully completed by the PI/CoPI, Data Manager or any responsible research team member and submitted to the central database administrator attached to the dataset, research tools such as questionnaires and reports.*

**Index** is unique identifier given to a dataset and/or a research document.





**Last Update** is the date when a certain dataset has been modified.

**Submitting Organization** is an institution or research directorate, or a research unit that submitted a dataset.

**Creator** is responsible expert that produced the final clean dataset.

**Data quality** is a state where a dataset produced complies with standards as per NDMC data quality guidelines.

**Dataset category** is thematic group of a dataset that is mostly associated with.

**Dataset Format** is the type of software or program a dataset is stored

**Size** is the magnitude, in megabyte, of a dataset or a document produced, created and submitted.

**Availability** is a condition whether and how a dataset or a resource can be obtained or accessed, or including contact information.

**Study subjects** are units of measurement or entities under study, surveillance or investigation.

**Treatments** are items or systems used for intervention in a certain study design.

**Study design** is the type of study such as cross-sectional surveys, surveillance, cohort studies, case control, clinical trial.



## Data Information Form (DIF)

Information about datasets, metadata, reports, tools such as questionnaires, journal articles published from the datasets, etc are supposed to be provided by carefully completing the DIF given below.

Attribute	Description
Index	
Dataset name	
Dataset label	
Dataset format	
Dataset category	
Study design	
Sample size	
Study subjects/Sampling unit	
Treatments	
Disaggregation level	
Coverage (Age)	
Coverage (Geospatial)	
Coverage (Sex)	
Size	
Time period covered	
Keywords	
Directorate	
Departments/Case Team/Research	
Codebook file name	
Availability	
Name of survey/project	
Principal Investigator (PI)	
Co-Principal Investigator (Co-PI)	
Contributors	
Creator	
Publisher	
Date of release/publish	
Last Update	
Funding source	
Submitting Organization	
Suggested citation	
Recommended acknowledgement	



## Codebook summary

A summary of codebook might be presented using the template below.

Codebook name: \_\_\_\_\_ Dataset name \_\_\_\_\_ Index \_\_\_\_\_

Variable name	Variable label/description	Variable values/codes	Value labels	Ranges of values



## PART II: Database organization plan

### Dataset Categories

#### Classification by data Sources

1. EPHI – EPHI projects and Administrative sources<sup>1</sup>
  - Bacterial, Parasitic and Zoonotic Disease
  - Food Science and Nutrition Research
  - Health Systems and Reproductive Health, NCDs
  - HIV and TB
  - Preparedness, Early Warning & Response & Recovery
  - Traditional & Modern Medicine
  - Vaccine and Diagnostic Reagent production
2. Universities and health research institutions

#### Classification by Cause of death or injury

- Cardiovascular diseases
- Childhood illness
- Chronic respiratory diseases
- Diabetes and kidney diseases
- Digestive diseases
- Enteric infections
- HIV/AIDS and sexually transmitted infections
- Maternal and neonatal disorders
- Mental disorders
- Musculoskeletal disorders
- Neglected tropical diseases and malaria
- Neoplasms
- Neurological disorders
- Nutritional deficiencies
- Other infectious diseases
- Other non-communicable diseases
- Respiratory infections and tuberculosis
- Self-harm and interpersonal violence
- Sense organ diseases
- Skin and subcutaneous diseases
- Substance use disorders
- Transport injuries
- Unintentional injuries

<sup>1</sup> Based on current EPHI structure



### Classification by Study design

1. Cross-sectional
2. Surveillance
3. Cohort
4. Case control
5. Clinical trials
6. Longitudinal
7. Vital events registration
8. Census
9. Laboratory
10. Randomized controlled trial
11. Repeated cross-sectional
12. Retrospective
13. Routine

### Classification by Demographic groups

1. Children and Women
2. Under Five
3. Youth, adolescents
4. Other age groups
5. Common demographics: age, sex, other conditions

### Classification by Geospatial

Coordinates, Region/Zone/Woreda/Kebele. Something else useful to set up alongside this in terms of capturing geographic representativeness of sources is a location hierarchy for Ethiopia that can account for districting changes over time. One way to do this is location hierarchies by year, then some sort of map between most detailed locations in each of these years to current boundaries. This way any data from the past can be easily linked to the present.

1. Rural/urban
2. Agro ecology zones (AEZ)
3. Other environmental characteristics

### Classification by Study Type

1. Qualitative
2. Quantitative
3. Mixed data

### **Notes:**

Data should be indexed with unique identifying number that can link metadata to collected data

Should include searching by



- i. disease/health category, i.e. Malaria or smoking
- ii. data sources/institutions/departments
- iii. categories/groups/spatio-temporal nature of data/type

Should automate use of mixed *Search queries*

### Sample Relational Tables

List of tables

Entities (Tables)	Attributes (columns/fields)	Relationships (keys)
Table 1: Institution -Departments/Research case teams/institutions	Table 1	DeptID
Table 2: Survey/Project	Table 2	IRB code
Table 3: Datasets	Table 4	IRB code
Table 4: Survey/Project classifications	Table 3	IRB code

Table 1: Institution - Departments/Research case teams/institutions

Names	DeptID	data sources
HIV and other Viral diseases		EPHI departments, research case teams
Malaria and other parasitic and vector borne diseases		EPHI departments, research case teams
Noninfectious diseases		EPHI departments, research case teams, Health research institutes
Planning, Monitoring and Evaluation		EPHI, Partners, MoH, MOA,
Preparedness, Early Warning & Response & Recovery		EPHI PHEM and International sources
TB and other Bacterial & Mycotic diseases		EPHI departments, research case teams
Traditional & Modern Medicine		EPHI departments, research case teams
Vaccine and Diagnostic Reagent production		EPHI departments, research case teams
Zoonotic diseases		EPHI departments, research case teams

Table 2: Survey/project

Project name	DeptID	IRB code	Project Investigator (PI)	Study design	Description
Performance evaluation of Genexpert HIV-1 Quantitative Viral Load Assay in Resource Limited Settings					
Unmet Health Needs Among Disabled Children in Rural Amhara Region					
Field Evaluation of HIV Point of Care Technology for Early Infant Diagnosis in Ethiopia					
Survival of Patients with Ovarian Cancer in Addis Ababa					
Evaluation of HIV Rapid Test Quality Improvement Initiative on HIV Testing Performance in Ethiopia (RTQII)					
Studies on Malaria Vector Behavior and Entomological Monitoring in Four Sentinel Sites in Ethiopia					



Assessment of Plasma levels of bone metabolism biomarkers among HIV infected adult patients in Ethiopia					
Health Facility Assessment in Ethiopia: Service Availability and Readiness Assessment (SARA)					

Table 3: Dataset names

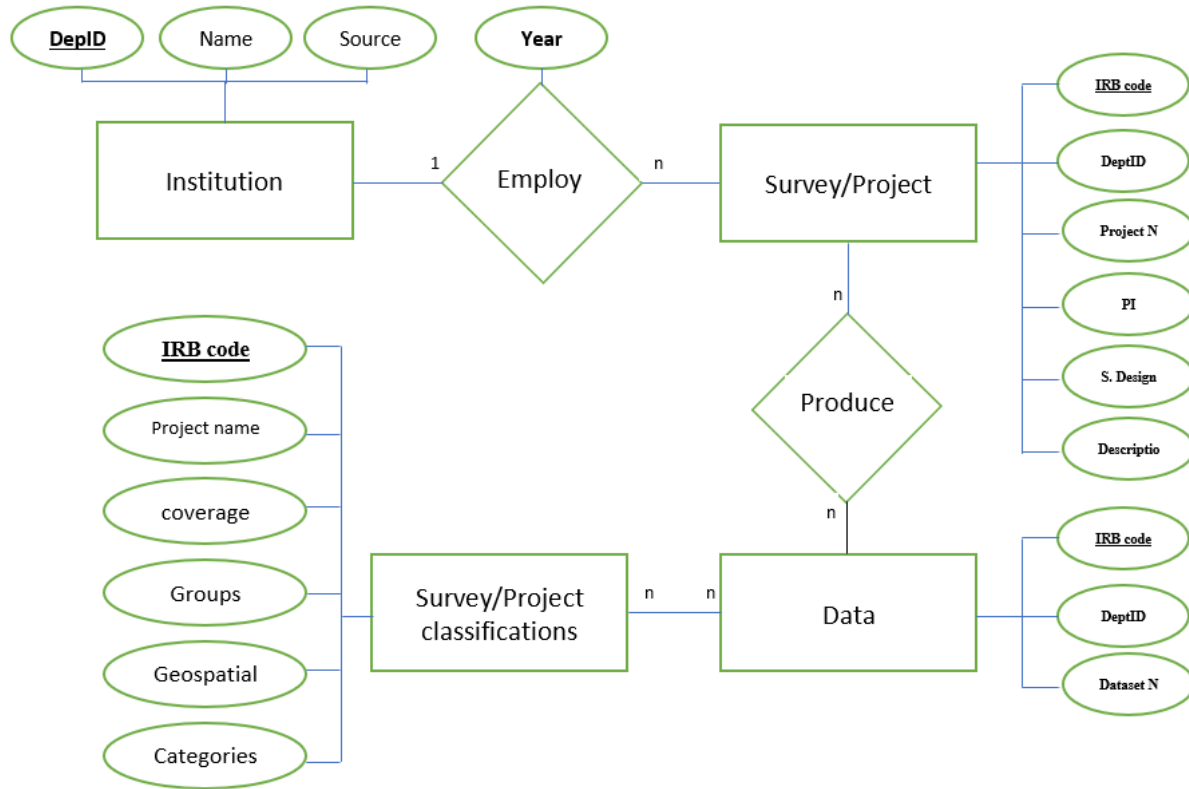
Dataset name	DeptID	IRB code
DatasetA		
DatasetB		
DatasetC		
DatasetD		

Table 4:

Project name	Dept ID	IRB code	coverage	Groups	Geospatial characteristics	Categories
Performance evaluation of Genexpert HIV-1 Quantitative Viral Load Assay in Resource Limited Settings			National	Children	Rural/urban	Laboratory
Unmet Health Needs Among Disabled Children in Rural Amhara Region			Regional	Women	Agro ecology zones (AEZ)	Survey
Field Evaluation of HIV Point of Care Technology for Early Infant Diagnosis in Ethiopia				Youth, adolescents	Environmental characteristics	Surveillance
Survival of Patients with Ovarian Cancer in Addis Ababa						Cohort
Evaluation of HIV Rapid Test Quality Improvement Initiative on HIV Testing Performance in Ethiopia (RTQII)						Case control
Studies on Malaria Vector Behavior and Entomological Monitoring in Four Sentinel Sites in Ethiopia						Clinical trials
Assessment of Plasma levels of bone metabolism biomarkers among HIV infected adult patients in Ethiopia						
Health Facility Assessment in Ethiopia: Service Availability and Readiness Assessment (SARA)						



Sample Relational Model







## PART III: Data Archival Processes

### Existing/Retrospective data

#### EPHI dataset

##### *Phase I: Collect and organize baseline information*

Sources: Planning directorate, EPHI IRB office, other directorates

- ✓ Prepare registry forms
- ✓ Collaborate with planning directorates if common and harmonized interests could be done
- ✓ Review and register research projects approved and implemented in EPHI so far
- ✓ Update registry forms, if required, to capture a more detail and comprehensive information

##### *Phase II: Collect and organize datasets*

Cleaned data are obtained from target sources.

There could be instances where data are supposed to be cleaned at NDMC on the basis of some requirements if a submitted data doesn't comply with and the sources and departments or researchers didn't go further.

A standardized metadata should be linked with the structure of the original cleaned data through the database.

The standardized metadata and codebooks should be stored in the database maintaining a clear identification matched with the main data in an efficient way.

Identify steps for awareness, collaboration and effort mapping, shared objectives and confusion clearance among stakeholders within EPHI.

- ✓ meetings with departments and PIs (*separately at different times, or by mass*)
- ✓ document meeting outcomes: *agreements inquired, collaborative efforts promised and resources mapped, cleared confusions and measures (to be) taken,*  
...
- ✓ quick plan and action as per the documented meeting outcomes

Collect or update research or project and data.

Do exhaustive clean data submission to the server.

##### *Phase III: Create and update database structure*

Comprehensive understanding of information from Phase I and Phase II.



Construct conventional framework for the structure of the database. (Constructing the database framework should probably be piloted so that it doesn't have to change frequently)

Derive conventional dataset name labelling and description items.

*Phase IV: Publication*

Create web page

- ✓ available datasets with brief descriptions about the datasets (study objectives, design, methodology, sample size, time and duration, access and sharing conditions, variables...)
- ✓ related documentations (reports, briefs, visualizations, ...)

**Other datasets**

*Phase I: Collect and organize baseline information and datasets*

Sources: Universities, Health research institutes

Identify steps for awareness, seek collaboration and effort mapping, and shared objectives among stakeholders outside EPHI

Identify/classify dataset categories or types

*Phase II: Create and update database structure*

To have EPHI datasets and datasets from elsewhere be stored combined into the same structure, but tagged with an EPHI-specific label

Comprehensive understanding of information from Phase I and Phase II

Construct conventional framework for the structure of the database

Derive conventional dataset name labelling and description items

*Phase III: Publication*

Update NDMC web page with information about available data at EPHI data portal.



## Prospective data

### *Phase I: Database*

Develop a database system – enabling activities in “PART A” work for future

Develop integrated approach with planning directorate and EPHI-IRB for research projects clearance, approval, implementation and data quality controls

Data quality

- ✓ take part and collaborate with researchers, including during the process of ethical clearance
- ✓ data quality standards – detail data description including naming and labeling
- ✓ data cleaning and submission procedures to EPHI server
- ✓ preliminary and final report submission
- ✓ setting conditions and time/duration to open data for access
- ✓ involve PIs in the system and further researches

Trainings – EPHI researchers, Data managers, NDMC team, Health data producers

Maintain the EPHI Data Management System sustainability

- ✓ Upgrade servers
- ✓ Maintain applications’ interoperability
- ✓ Update database to the state-of-the-art technologies
- ✓ Update and enhance web-based data information and access
- ✓ Maintain efficient data integration, data federation, master data management and data governance
  - policies, strategies, guidelines, protocols, ... of sharing, access, use, confidentiality and privacy, data visualizations, ...

### *Phase II: Knowledgebase*

Develop system of knowledge repositories.

Collect and organize information and publications vis-à-vis main datasets.

### *Phase III: Data Science*

Preliminary introduction of activities involving

- ✓ big data storage, analytics, machine learning, intelligence, etc.
- ✓ processing, analyzing, and visualizing data



## PART IV: Data security

One of the main responsibilities of NDMC is to **protect data from theft, breach of confidentiality, premature and unauthorized release, and ensure secure disposal**. Security is a consideration at all stages of research, particularly if working with sensitive data. Data security refers to physical security which concerns with the status of devices on which data are stored and accessed.

NDMC put in place a controlling mechanism to have access to rooms, cupboards, and drawers where data is stored. To access data, responsible staff should sign a non-disclosure agreement outlining the nature of confidentiality, storage conditions, and data retention policies.

- Computers should be password protected, with file permissions controlled so users, depending on their status, can “read only”, “write”, or “execute” files.
- The center ensures computer firewalls are enabled and antimalware software are up-to-date and operational.
- Network connected personal computers should not store sensitive data, unless that data is encrypted, so to minimize network vulnerabilities.
- Any web applications originated from EPHI data center and needed to be accessed through Internet must be encrypted with latest technologies (SSL/TLS)

### Establish a digital data security procedure

#### *Passwords*

Passwords are a foundation of security. NDMC ensure having strong password that need to be changed at regular interval

#### *Encryption*

Encryption is a security technology tool to secure data at storage and during transmission through networking. Encryption maintains the security of data and documentation through an algorithm by transforming understandable information into something unreadable. A secret “key” is required to decrypt and return to clear text.

NDMC would consider having encryption/decryption technologies implemented on programs, backups and application software.

#### *Data and documentation destruction*

Data management responsibilities do continue after project ends. Disposing of paper or digital copies of data is not as simple as dumping paper into a waste paper basket or pressing the “Delete” key. If sensitive data is no longer needed, paper copies need to be shredded in a crosscut shredder, while digital storage devices need to be overwritten to scramble files. NDMC may consider using overwriting programs for deleting digital data

#### *Data back-up*



Due to system failure, NDMC's data can be lost or compromised. Identifying a data back-up strategy involving creating multiple copies of data and documentation having at least one external or off-site back up and updating data on a daily basis is essential.

#### *Secure working storage and data sharing*

NDMC's data management is collaborative and characterized by cross-institutional and cross-national collaboration. With the opportunity's technology allows for increased collaboration, the NDMC save data on EPHI's server dedicated for this purpose. The NDMC strives to make EPHI's server professionally backed-up, have reliable access, and secure. Storage on EPHI's servers can be problematic for collaborative research data management involving cross-institutional collaboration, as providing access to a network to users from other institutions is complicated. In this regard, NDMC will considering having its own "Cloud" based storage for open access data, for emergency, laboratory and research data. Cloud based storage are increasingly popular as they are easy to use, access, and often automate useful research data management practices.

### Data protection

In order to maintain the integrity of stored data, data should be protected from physical damage, loss, or theft by limiting access to it. The NDMC lead decides which staff members are authorized to access and manage the stored data. Privacy and anonymity can be assured by replacing names and other information with encoded identifiers, with the encoding key kept in a different secure location. To best protect data, NDMC will educate all staff members about data protection procedures. Theft and hacking are particular concerns with electronic data of human subjects and other confidential records. The center protects electronic data by taking the following precautions:

#### *Protecting access to data*

- Using unique user IDs and passwords that cannot be easily guessed.
- Changing passwords often to ensure that only current project members can access data.
- Provide access to data files through a centralized process.
- Evaluate and limit administrator access rights.
- Ensure that outside wireless devices cannot access our system's network.

#### *Protecting EPHI/NDMC, ICT system*

- Keeping updated anti-virus protection on every computer.
- Maintaining up-to-date versions of all software and media storage devices.
- If the system is connected to the Internet, NDMC uses a firewall and intrusion detection software to monitor access.
- Implement physical security system (video surveillance, biometrics) to control entrance/exit to data center.

#### *Protecting data integrity*

- Record the original creation date and time for files on NDMC system.
- Use encryption or electronic signatures to keep track of authorship and changes made to data files.
- Regularly back up electronic data files (both on and offsite) and create both hard and soft copies.
- Ensure that data are properly destroyed.





## PART V: Internal framework/procedure for sharing data to a third party

As stipulated in article 12 of the 2016 EPHI's data sharing guideline, different units/individuals at EPHI play roles and are responsible for making decision for sharing EPHI's data to other party.

### ***EPHI's Director General***

- Receive the data sharing/access request
- Forward the data sharing request to NDMC for necessary action
- Approve the final decision based on the feedback from NDMC

### ***Director of the data generating directorate***

- Check the question on the data request can be addressed using the requested dataset
- Check the appropriateness and relevance of the request
- Check if there is no conflict of interest among users and principal investigator
- Provide feedback based on the finding from the discussion with the study team

### ***National Data Management Center for health (NDMC)***

- Assess the submitted "data sharing request" for its adherence to EPHI data sharing requirement
- Inform data generating directorate/unit about the "data sharing request" and get approval
- Following final approval by the director general, NDMC will provide access to the requested data (either with online data transfer or sharing file with CD, Memory sticks).

### **Data sharing/access fee**

As stipulated in article 13 of the 2016 EPHI's data sharing guideline, data can be shared without fee or with fee depending on whom the data is shared with.

Agencies or persons given access to data free of charge include

1. Policy-making Government Institutions/Organizations;
2. Students of higher education institutions who request access to raw data as part of fulfillment of their degree
3. Those Organizations/Institutions that have financially and technically supported the operation of the survey/census from which the raw data has been generated.

Agencies or persons given access to data with fee include all except those mentioned in article 13.1, others shall be charged a certain amount of fee that would be estimated by EPHI Supply and Finance Directorate and approved by the DG.



### Flow chart for data sharing

#### Director General

- Asses the data request for its appropriateness
- Send a copy of accepted request to NDMC and respective DDGs

Request rejected

#### National Data Management Center

- Vetting data request in line with 2016 EPHI data sharing guideline
- Consult directorate of data source for accepted request
- Communicate DDG when there is discrepant decision b/n NDMC & directorate
- Provide data
- Communicate data request refusal to requester

Discrepant decision

#### Deputy Director Generals (RTT, PHEM and Laboratory)

- Follow up the process
- Reconcile decision discrepancies b/n NDMC and directorate

Request rejected

#### Data generating directorate

- Consulted to share data





Ethiopian Public Health Institute

Raw Data Access Request Form

Applicant/Requesting Institution/Person

Name: \_\_\_\_\_

Name of Organization: \_\_\_\_\_

Organization/Affiliation Type:

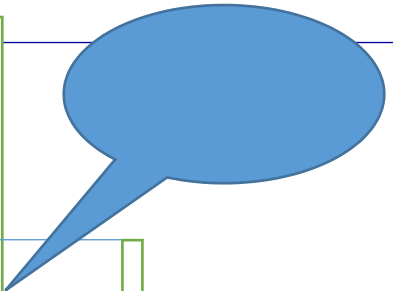
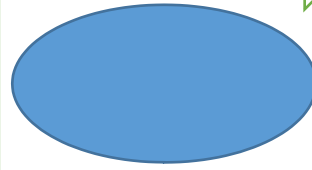
- Governmental Office
- Research Institution
- Higher Education Institution
- Researcher
- University Student
- Private /Consultant Organization
- NGO
- International /Foreign Organization

Funding fully/or partially

Technical Assistance

None

Signature:



Large empty rectangular box for providing details and contact information.

Bottom section containing a horizontal box and a vertical box with arrows indicating flow or relationship.



Data Requested:

No.	Name/Title of dataset	Year of data produced	Data Type (Laboratory, RTT, or PHEM)	Level of disaggregation of the data <ul style="list-style-type: none"><li>• Country level</li><li>• Regional level</li><li>• Zonal level</li><li>• Woreda level</li><li>• Kebele level</li><li>• Other /specify/</li></ul>	Remark
1.					
2.					
3.					
4.					
5.					
6.					
7.					
8.					
9.					
10.					
11.					
12.					
13.					
14.					
15.					
16.					
17.					
18.					
19.					
20.					
21.					
22.					
23.					
24.					
25.					



**Purpose of data use**

Objective

---

---

---

---

---

---

Research Questions

---

---

---

---

---

---

---

---

---

---

Data use/Study period

---



**For EPHI use only**

**Office of the Director General**

Data Request Received

Date \_\_\_\_\_

Decision \_\_\_\_\_ (Can be list of options of decision type as a check list to be marked; like for instance, Directed to NDMC for review of request and appropriate procedures; Refused because of ....some concerns might arise, etc)

Signature \_\_\_\_\_

**The National Data Management Center for health (NDMC)**

Data Request Process (*According to Internal procedure for sharing data to a third party*)

1. Review of the request is done at NDMC \_\_\_\_\_

Comments by NDMC Officer (Forward to Directorate (NAME); Request Incomplete)

\_\_\_\_\_

2. Review of the request is done at Directorate/Department (NAME)

\_\_\_\_\_

[

*The question can be addressed using the requested data set \_\_\_\_\_*

*The request is appropriate and relevant \_\_\_\_\_*

*There is no conflict of interest among users and principal investigator \_\_\_\_\_*

*(May provide feedback based on the finding from the discussion with the study team)*

]



**Directorate**

Consent to share data has been obtained from the Directorate as below.

(Data can be shared; Data should not be shared; If refusal, mention reasons/protocols as per EPHI’s data sharing policy about how, when and why data should/not be shared)

Directorate Head

Name: \_\_\_\_\_

Date: \_\_\_\_\_

Signature: \_\_\_\_\_

3. Final Decision

Head of NDMC

(Share data/ Access is denied for reason #X/Request is incomplete and suggest the applicant to complete/)

Share data

Free of charge \_\_\_\_\_

With Charge \_\_\_\_\_

Name: \_\_\_\_\_

Date: \_\_\_\_\_

Signature: \_\_\_\_\_

Data Sharing Process

1. NDMC Data Governor/Manager/Administrator

The final approval is given by Head of NDMC \_\_\_\_\_

Copy of Receipt of payment provided, if access is with charge \_\_\_\_\_

Data shared: Yes/No (If no, reason)

Name: \_\_\_\_\_

Date: \_\_\_\_\_

Signature: \_\_\_\_\_

Applicant has Agreed with the Terms and References of EPHI’s Data use, access and sharing policy by signing the “Data sharing agreement” as per article the 2016 EPHI’s data sharing guideline \_\_\_\_\_.

Applicant Received the Data Requested

Name: \_\_\_\_\_

Date: \_\_\_\_\_



Signature: \_\_\_\_\_

## References

- Central Statistical Agency (2004). DIRECTIVE No. 1 /2004. Directive to Establish Procedures for Accessing Raw Data to Users. Addis Ababa, Ethiopia.
- Central Statistical Agency (\_\_\_\_) Form for Requesting Access to Raw Data. Addis Ababa, Ethiopia.
- Ethiopian Public Health Institute (EPHI). (2016). Ethiopian Public Health Institute Guideline for Data Management and Sharing. Addis Ababa. Ethiopia.
- Healthcare Information Technology Standards Panel (HITSP). (2010). Data Dictionary Component. Review Copy. 20100131 V1.0.1.
- J. Paul Getty Trust. (2008). *Introduction to Metadata*. Second Edition. Edited by Murtha Baca. The Getty Research Institute.
- National Archives of Australia. (2010). AGLS Metadata Standard Part 1: Reference Description. Version 2.0. July 2010. (<http://www.agls.gov.au/>).
- National Archives of Australia. (2010). AGLS Metadata Standard Part 2 – Usage Guide. Version 2.0. July 2010.
- NIH Common Fund Metabolomics Data Sharing Plan.  
<http://www.metabolomicsworkbench.org/nihmetabolomics/datasharing.html>.
- OCLC and RLG (2005). Data Dictionary for Preservation Metadata. Preservation Metadata: Implementation Strategies (PREMIS). A working group jointly sponsored by OCLC and RLG. Final Report of the PREMIS Working Group. May 2005.