NIPN Data Repository Metadata Checklist

Addis Ababa, Ethiopia

October 29, 2023

NIPN Data Repository Metadata Checklist

Background

The National Information Platform for Nutrition (NIPN) is a global initiative launched by the European Commission to support Scaling up Nutrition countries that have a high malnutrition burden. It supports the generation of evidence that is used by nutrition stakeholders to develop policy, design programs and to allocate investments. The NIPN initiative is currently implemented in nine countries around the world. The second phase of the project is financially supported by the European Union (EU) and the Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ).

The NIPN in Ethiopia was launched in 2018. It is hosted at the Ethiopian Public Health Institute (EPHI) within the Food Science and Nutrition Research Directorate (FSNRD). Technical assistance to the NIPN is provided by the International Food Policy Research Institute (IFPRI) and GIZ. The objective of the NIPN in Ethiopia is to strengthen national capacity to monitor progress towards under-nutrition reduction and to implement more cost-effective and evidence-based policies. The NIPN works under the umbrella of the Food and Nutrition Policy (FNP) and engages with multisectoral nutrition stakeholders.

The NIPN operational cycle consists of three elements that constantly revolve and feed into each other, namely: policy question formulation based on national priorities, analysis of data to inform the questions, and communication of the findings back to the government and stakeholders. To do data analysis, the data must be accessible. As such, the NIPN data repository is a central place where existing dataset descriptions are placed to be accessible to researchers, nutrition stakeholders, and policymakers, and includes information on how to access the actual datasets. For each existing dataset, a dataset description document (metadata), a codebook (or codebooks), data collection instruments/questionnaires, and relevant published articles related to the datasets, is included on the data repository. This allows data users to easily find and use the existing information to answer their own research questions. In addition, archiving descriptive information about the datasets will help prevent duplication of research and will optimize the use of existing resources.

To ensure continuity of uploading metadata documentation into the NIPN data repository, preparing a metadata checklist that describes the minimum requirements and components of the metadata documentation is necessary.

General Objective

The general objective of the metadata checklist is to set a standard content for each available multisectoral nutrition metadata that will be uploaded on the NIPN /National Data Management Center (NDMC) data repository. This will also help to improve the quality and completeness of the dataset.

Specific objective:

- To guide data providers to have a similar study level documentation/ meta data description and code book (see Annex 1 -2 on mandatory fields of metadata description and codebook),
- To facilitate data exchange and enable data sharing among different stakeholders,
- To promote the reuse of data by facilitating metadata discoverability and standardization, in an easily and known data repository,
- To ensure that metadata documents (study level description, codebook, and questionnaire) have the accurate and relevant information to enable users to make informed decisions about whether to use the dataset or not.

Metadata documentation

Metadata documents include study-level documentation or metadata descriptions, data-level documentation, or codebooks, and will include documentation resources such as existing questionnaires, reports, and publications. These documents will help researchers or data users to understand what the data are measuring and how the data were created, to assess the quality of the data, and to provide information on how to access the data.

Uploading metadata into NIPN/NDMC repository

To upload a new metadata into NIPN repository, two ways will be used. 1) NIPN data analyst team will prepare all the metadata documentations   2) NIPN stakeholders who are willing to use NIPN data repository will complete and submit a final metadata documentation along with the necessary steps to obtain the data.

In both cases, the potential dataset should meet the below list of minimum requirements. With out these requirements it will be difficult to prepare a metadata documentation. The checklist will be shared with NIPN stakeholders incase of any interest.

No	Requirements	Yes/No	Remark
1	Subject -specific dataset (Nutrition and Nutrition related) after 18 months of data collection		
2	Dataset should be complete, consistent, accurate, timely and unique.		
3	Depending on the data sharing policy of the Institute, either a cleaned dataset or metadata description and codebook should be shared. Please see Annex Table 1 and 2 for information that should be included in the metadata description and codebook.		
4	The researcher/ Principal investigator must have the final edited tools (such as questionnaires), which are aligned with the collected data. If a separate questionnaire is not available, then it should be included in the survey report.		
5	Information on affiliation (name of organization), author of the survey or research, title and funder should be provided/included		
6	Keywords or phrases describing the subjects or content of the data must be included in the report.		
7	Key dates associated with the data, including: start and end date of the project; time period for collecting the data and other dates associated with the data must exist.		
8	Electronic data collection tool, software used to collect data, and any other relevant information should be included in the survey report.		
9	Explanation of codes or abbreviations used in either the file name or the variables in the data files.		
10	The final technical report after dissemination must be available.		

Annex 1: List of contents for the study level description/ metadata description

Elements	Description
Title	Official title of the study / data (e.g., Ethiopia Demographic and Health Survey [EDHS] 2016.
Study Description	The major objectives, the aim of the study, reasons why data was collected.
Producer(s)	The person/organization who will act as responsible body for the data throughout the data life cycle. *Note:* Always put the name of the organization who will be responsible for the data throughout the data life cycle. However, if the study is owned by persons put the name of the person(s) who is/are the Principal Investigator(s) (e.g., data collected for a PhD dissertation).
Contributor	The organization or person responsible for ether collecting, managing, or otherwise contributing in some form to the development of the resource. (This includes data collectors, project leader, researchers, funding agency etc.,).
Geographic coverage	Information on the geographic coverage of the data (e.g., national, regional, program area).
Study design	Explanation on the type/design/kind of the data/study (i.e., survey, census, evaluation, follow-up/cohort, intervention study, case-control).
Sample and Sampling procedures	Detailed description of the target population and the sampling methods used (assuming the entire population is not studied).
Time dimensions:	Explanation of the frequency of the study and data collection (i.e., one time, annually, every five years).
Response rate	The number of people who answered the survey divided by the number of people in the sample. It is expressed in the form of a percentage or as a proportion.
Date of data collection	Information about when the data was collected (start date and end date).
Method of data collection	Lists all data collection methods used (e.g., face-to-face/key informant interviews, focus group discussions, observations, self-completion, clinical measurements, physical/anthropometric measurements).
Data collection material	Information on data collection material/device (paper, or electronic, or both). *If electronic*: the software used to collect the data.
Data quality control used during and before data collection	Lists all data quality control mechanisms used before and during data collection (e.g., pretesting of data collection instruments/questionnaires, training of data collectors, supervision during data collection, re-interview, refresher training).
Data source (s)	If applicable - when data are obtained from resources other than surveys - explain the original data sources or documents.
Study population	Describes which population group the data are collected for (e.g., children, women, men, households etc).
Number of observations/units	Number of total sample observation interviewed per unit (e.g., number of children 0-59 months, number of women aged 15-49).
Method of data entry	Explains the software used for data entry.

Weighting	If applicable, information on weight variables, how they were constructed, and how they should be used.
Representativeness of the findings	Describe at which level the data provide reliable estimates that reflects the characteristics of a larger group in a geographic location. For instance, the EDHS is designed to provide estimates of population and health indicators nationally, for urban and rural areas separately, and for each of the 11 regions. Lists all geographic areas for which findings from the data provide reliable estimates for a larger group in the geographic location (e.g., at national, regional, zonal, or woreda level, or urban/rural).
Data access	Explains how the data can be accessed (e.g., whether through an online form or through a paper data request form).

Annex 2: List of contents for data-level documentation/codebook.

Variable name	The name or number assigned to each variable in the dataset (i.e., Q004, Q005).
Variable label	A brief description which allows identifying the variable for the user (e.g., weight of a child in kilograms, sex of a child).
Values	The actual coded values in the data for this variable (e.g., 6, 1). Note: For all variables put the minimum and maximum values of the variables (e.g., for age of a child in months, the value will be minimum 0; maximum 59).
Value labels	The textual descriptions of the codes (e.g., for sex of the child: 1=male, 2= female). *Note:* For variables which do not need textual descriptions of the codes use "open ended" in the value label (e.g., for child age or child weight use "open ended" in this section of the codebook).
Variable type	Information on the format of the variable (e.g., numeric, string).