Basics of Metadata Creation And Research Tracking Database System (RTDS)

Ethiopian Public health Institute, National Data Management Center (NDMC)

Arba Mich

Getaye w.

December, 2022

Presentation outlines

- What is Metadata?
- Rationale of the metadata
- Categories of metadata
- Metadata creation templates
- Uploading metadata to the research tracking database system (RTDS)
- Challenges in metadata creation & Research tracking database system
- Demonstration on uploading to the RTDS

Metadata

- Metadata is generally stated as ***"data about data"***.
- Metadata is a summary description about the datasets (the title, who collects the data, where, when, how, how often, and data quality level)
- It provides a structured reference that helps to sort and identify attributes of the information it describes.
- Metadata is a summary document that provides information on a dataset.

Metadata…

- It provides about the content, quality, type, creation of the datasets
- It **helps data users** (persons or organizations) identify the data importance to their purpose.
- It is the interoperability enabler providing information.

Rationale of metadata

- Facilitates resource discovery
- Facilitates the reuse of the data for further research and evidence synthesis
- Allows the users to understand and track important details of the work.
- Effective means of organizing electronic resources
- Facilitating interoperability and integrating resources
- Protect resources and their future accessibility
- Support archiving and preservation

Categories of metadata

- ***Metadata has three main categories:***

1. **Descriptive metadata**, enables discovery, identification, and selection of resources.

✓describes the content and context of the data;

✓Included the title, author, abstract and keywords ..e.t.c.

2. Structural metadata, about physical structure.

✓page order to chapters, chapters in a book, file format

Categories of metadata…

3. **Administrative metadata**

✓Information to help manage a resource, like a resource type, permissions, and when and how it was created

✓Has two categories;

- *Rights management metadata*
- *Preservation metadata*

Metadata template in NDMC, EPHI

- The repository of health data and research tracking system in EPHI has its own metadata template.
- It gives information for readers about the elements of the datasets archived in the repository.
- The metadata creation template includes the 5 major components

Metadata creation template

I) Data Contact information

Publisher

- It is an **institute or researcher** (if an individual) by whose name an item is published.
- Options **EPHI** Vs **Non-EPHI**.

Data Generating unit /work unit

- An organization or institution or research directorate, or a research unit that submitted the dataset to NDMC.
- E.g; EPHI directorates, Government organizations, NGOs, and any others…

I) Data Contact information…

Principal Investigator (PI)

- The PI is the main researcher leading a certain research project.
- The email address and phone number of the PI will be recorded .
- If not available; ***"default address"***

II) Data Contextual information

Date *data collection* started & end

- The data collection period in a format MM/DD/YYYY

❖Coverage Location

- Indicates sample representative of the study or a project.
- Commonly national, subnational, zonal & woreda level

Funding source

- Funding source is the name of an organization that contributed to this specific research /dataset.
- If not available --- not funded

III) Data Sharing or Access information

Level of access/ Restrictions

- It is a condition of whether and how a dataset or a resource can be obtained or accessed.
- Three options;

1. ***Access with request,***
2. *restricted access,*
3. *open access*

Source of data

- Internal vs **external** (website of that organization)

III) Data Sharing or Access information…

Microdata availability

✓disaggregated dataset

Suggested citation & Recommended citation

- ✓ Showed the preferable citation that publisher or producer of a dataset or publication would suggest/ acknowledged for citation.

Versioning and provenance

Date dataset archived

- ✓ Indicating the date of data archival to the NDMC; by MM/DD/YYY

Date data cataloged

- ✓ Data of completion of the metadata creation by MM/DD/YYY format

IV) Methodological information

❖Description of data set

- Indicated the tittle, design, when, list & name of variables, sampling, sample size, population coverage, analysis, and software used can be described.

IV) Methodological information…

❖Abstract

- Included the objectives, method, design, software used, study area and survey round, data content

❖Data type

- Indicated whether the data is census, survey, surveillance, routine report…

❖Age group

- It could be; Adolescents, Adults, women of reproductive age and all age groups.

IV) Methodological information…

❖Coverage sex

- This includes the coverage sex of the study included (“ M, F, Both, / N/A”)

❖Keyword

- Are words and phrases that users type into search engines to find information.

V) Data and file of the dataset

Dataset is collection of related, discrete items of related data that may be accessed individually or in combination

Dataset title

✓ Dataset title is the title of the research project (include the year of conducted)

Dataset Formats

✓ Software by which the data set stored; Excel, SPSS, Stata,…

Additional material

✓ **Codebook**, ethical clearance, and proposal if available

V) Data and file of the data set…

Geospatial data

- Geospatial data contains YES or NO options.

Date of publication

- Date of release/publish

..\Metadata creation template sample.docx

Research tracking database system/ RTDS/

- Is an online and public database management system in EPHI that stored the data set archived from different data sources.
- This system created to upload the metadata of any health and health related dataset (research, survey, surveillance dataset e.t.c).
- It facilitate dataset search, data storage, retrieval, access, sharing and use
- Supports the data management
- https://rtds.ephi.gov.et/public/

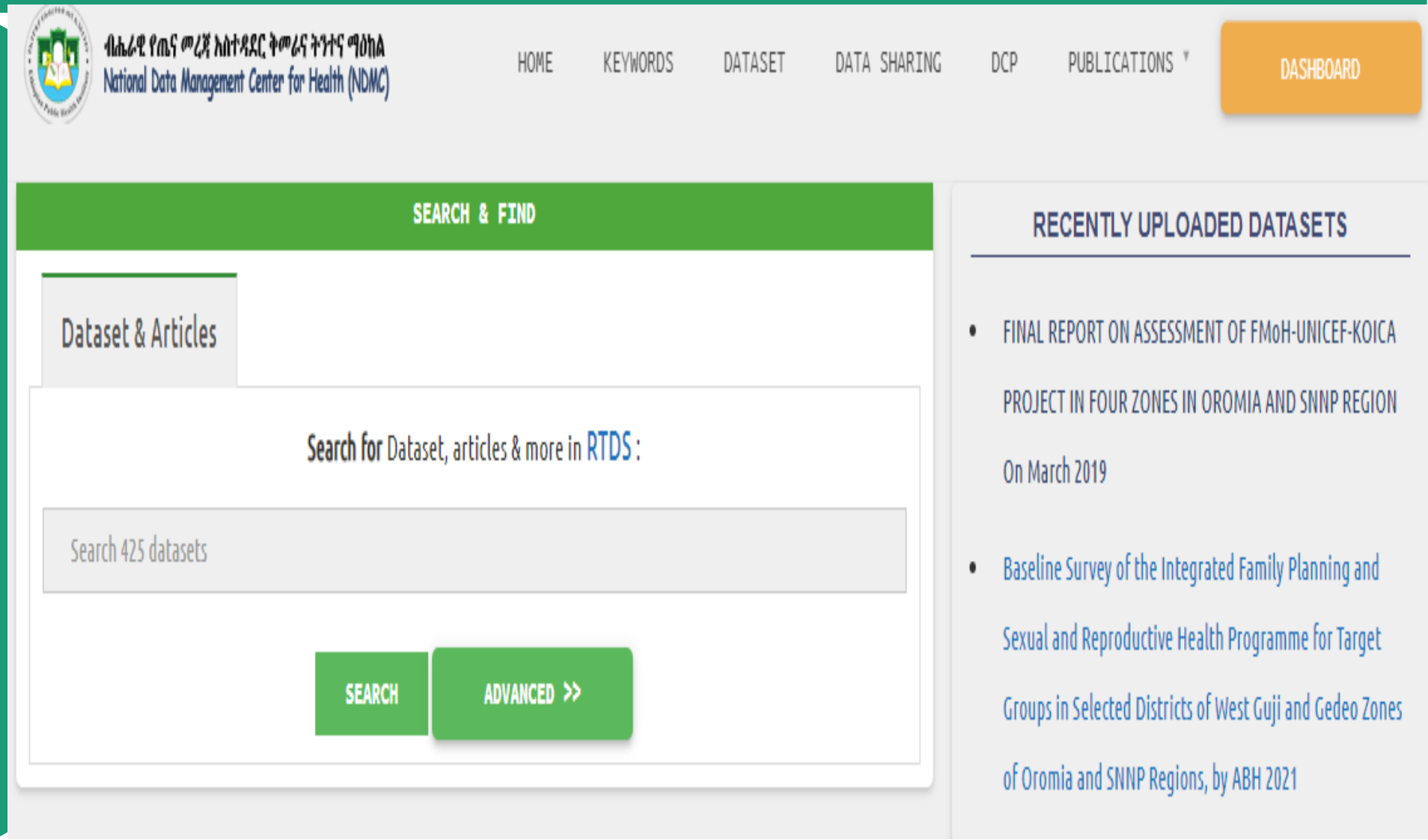
Research tracking database system/ RTDS…

- Created by uploading the metadata
- Publications, reports, scanned old documents and other documents can be uploaded
- It has online data **request platform.**

Research tracking database system/ RTDS

Currently more than 425 metadata uploaded.

https://rtds.ephi.gov.et/public



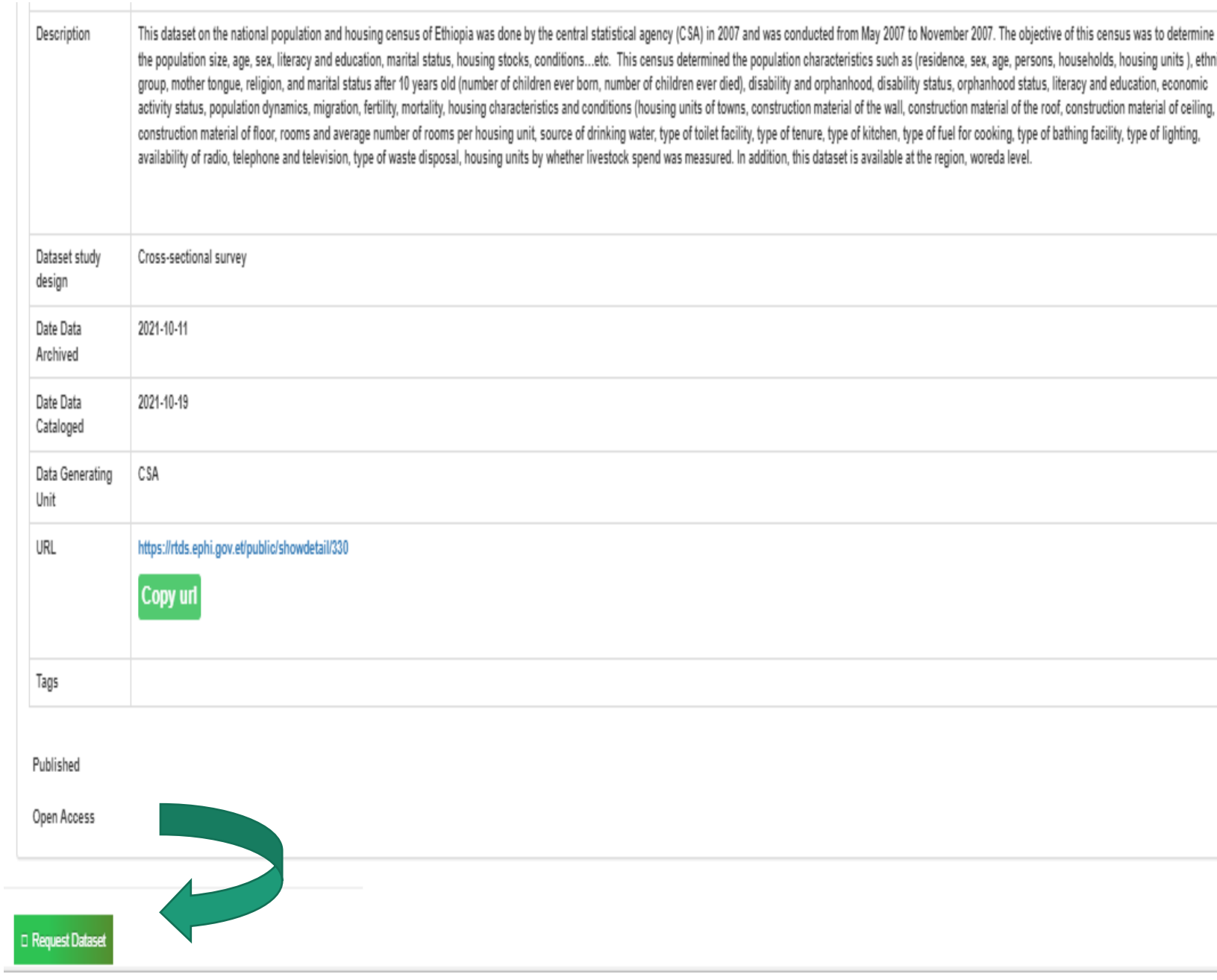
Research tracking database system/ RTDS

The URL created after complication of the uploading the metadata.

Description	This dataset on the national population and housing census of Ethiopia was done by the central statistical agency (CSA) in 2007 and was conducted from May 2007 to November 2007. The objective of this census was to determine the population size, age, sex, literacy and education, marital status, housing stocks, conditions...etc. This census determined the population characteristics such as (residence, sex, age, persons, households, housing units), ethnic group, mother tongue, religion, and marital status after 10 years old (number of children ever born, number of children ever died), disability and orphanhood, disability status, orphanhood status, literacy and education, economic activity status, population dynamics, migration, fertility, mortality, housing characteristics and conditions (housing units of towns, construction material of the wall, construction material of the roof, construction material of ceiling, construction material of floor, rooms and average number of rooms per housing unit, source of drinking water, type of toilet facility, type of tenure, type of kitchen, type of fuel for cooking, type of bathing facility, type of lighting, availability of radio, telephone and television, type of waste disposal, housing units by whether livestock spend was measured. In addition, this dataset is available at the region, woreda level.
Dataset study design	Cross-sectional survey
Date Data Archived	2021-10-11
Date Data Cataloged	2021-10-19
Data Generating Unit	CSA
URL	https://rtds.ephi.gov.et/public/showdetail/330 Copy url

RTDS…

- The requesters for data sharing can request from it via online method.



Description	This dataset on the national population and housing census of Ethiopia was done by the central statistical agency (CSA) in 2007 and was conducted from May 2007 to November 2007. The objective of this census was to determine the population size, age, sex, literacy and education, marital status, housing stocks, conditions...etc. This census determined the population characteristics such as (residence, sex, age, persons, households, housing units), ethn group, mother tongue, religion, and marital status after 10 years old (number of children ever born, number of children ever died), disability and orphanhood, disability status, orphanhood status, literacy and education, economic activity status, population dynamics, migration, fertility, mortality, housing characteristics and conditions (housing units of towns, construction material of the wall, construction material of the roof, construction material of ceiling, construction material of floor, rooms and average number of rooms per housing unit, source of drinking water, type of toilet facility, type of tenure, type of kitchen, type of fuel for cooking, type of bathing facility, type of lighting, availability of radio, telephone and television, type of waste disposal, housing units by whether livestock spend was measured. In addition, this dataset is available at the region, woreda level.
Dataset study design	Cross-sectional survey
Date Data Archived	2021-10-11
Date Data Cataloged	2021-10-19
Data Generating Unit	CSA
URL	https://rtds.ephi.gov.et/public/showdetail/330 Copy url
Tags	

Published

Open Access

Request Dataset

Uploading to the RTDS

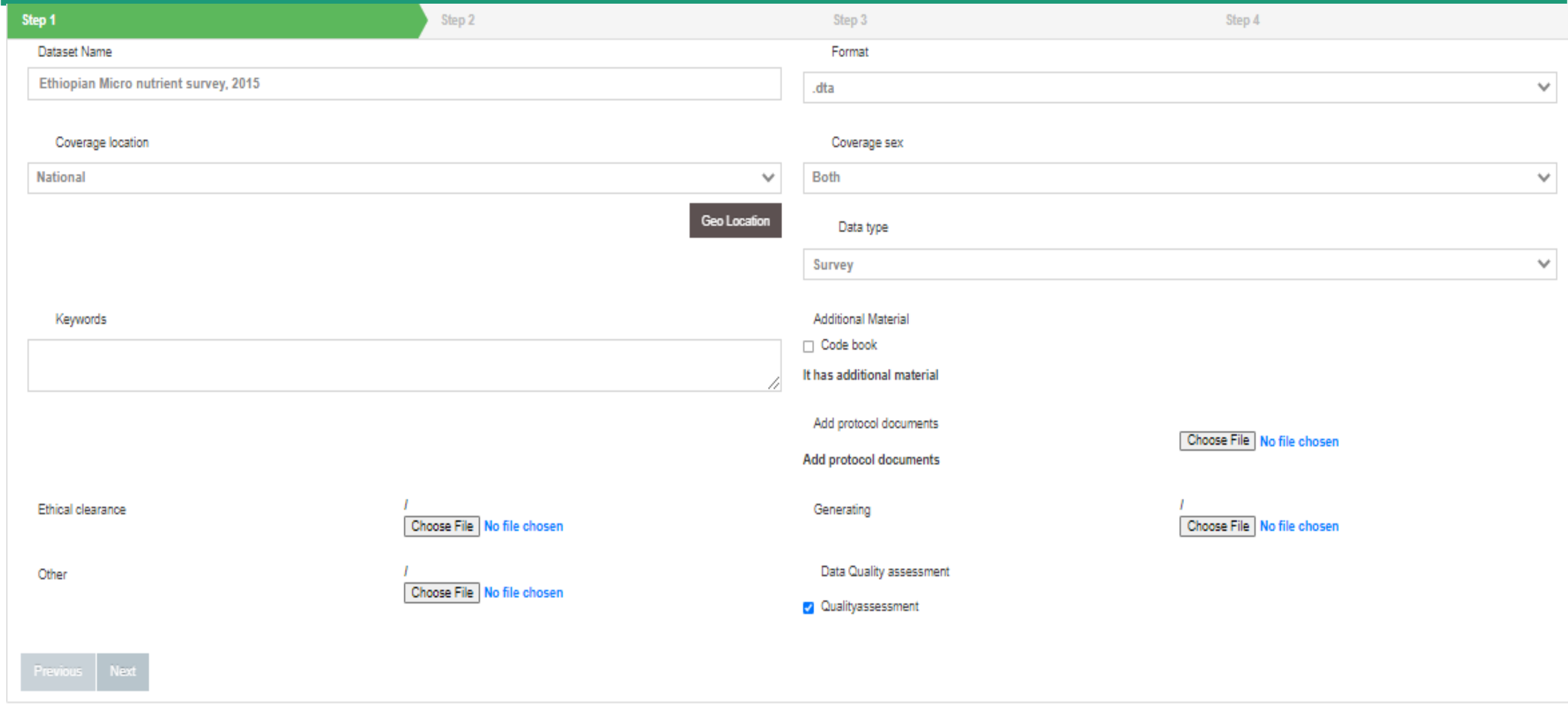
- It is storing/creating the metadata in standardized database system

Steps for uploading to RTDS

✓Creating the RDTS has the following steps

✓Search in to the webs site https://ephi.gov.et/ → https://ndmc.ephi.gov.et/ → https://rtds.ephi.gov.et/public/ →...

Steps for uploading to RTDS



Dataset

Step 1 | Step 2 | Step 3 | Step 4

Recommended Citations

Source | Styles | Format

NO

Cleaned

Raw format

. csdb

Date Data Collection Started

03/01/2015

Date Data Collection End

06/30/2015

Previous Next

Dataset

Step 1 | Step 2 | Step 3 | Step 4

Title

Ethiopian Micro nutrient survey, 2015

☐ Published

Sugested citation

Source | Styles | Normal

No

body p

Description

Source | Styles | Format

This large population-based cross-sectional survey was conducted by EPHI with a representative sample from nine regions and two city administrations in Ethiopia. The objective of this study was to estimate the prevalence of anemia, iron deficiency, vitamin A and B 12 deficiency, folate deficiency, zinc deficiency, iodine deficiency, and adequacy of iodized salt in Ethiopia. The findings of this survey revealed the following.

1. In Ethiopia, the prevalence of inflammation measured by CRP and AGP among under-five children, school children, and non-pregnant women of reproductive age was 44 %, 31.6 %, and 27.3% respectively. The prevalence of anemia adjusted for altitude among preschool children, school-age and non-pregnant women of reproductive age was 34.4, 25.6, and 17.7 %, respectively. The prevalence of Iron deficiency among preschool children, school-age children, and women of reproductive age by ferritin and adjusted for inflammation, was 17.8,

Dataset

Step 1 | Step 2 | Step 3 | Step 4

Study design: Cross-sectional survey

Age Group: All age

Geospatial: No

Microdata availability: Yes

Publisher: EPHI

Record type: Microdata

Data Generating Unit: Food Science and Nutrition Research Directorate

PI Name: Dilnesaw Zerfu

PI Phone: +251112786340

PI Email: null@gmail.com

Funding Source: EPHI

Level Of Access: Access with request

Restricted

☐ Is restricted

Is external ☐

Previous | Finish

Challenges

Incomplete datasets to create the complete meta data & RTDS

Unavailability of the data collection period start date & end date, sampling methods, design for the study.

Lack of complete address of the data owner.

Summary

The repository of health data and research tracking system in EPHI has its own metadata template.

Total metadata created & uploaded to the RTDS of EPHI are 425 data sets

Total metadata created & uploaded to the RTDS in the last year were 170+

Exercise

- Go to “https://ephi.gov.et/”
- Sample. ..\Metadata Sample.docx

❑Thank You