



Data Management Plan for Research Projects

Project title	A Comparative Food Consumption Analysis of 24-Hour Dietary Recall Data and Household Consumption and Expenditure Survey Data using FRESH baseline survey
Funder(s)	Not applicable
Principal investigator	Rie Goto
Contact email	Rie.Goto@lshtm.ac.uk
Date of last edit	26/07/24

Guidance on writing a Data Management Plan can be found at
<https://lshtm.sharepoint.com/Research/Research-data-management/>
and <http://servicedesk.lshtm.ac.uk>

Advice and feedback can be obtained from:
researchdatamanagement@lshtm.ac.uk

DESCRIBE YOUR RESEARCH

1. What digital resources – data, code, collection tools, etc. - will you collect/obtain and use?

Two data type will be analysed; 1) FRESH dataset is collected in a baseline survey of FRESH study by IFPRI (October 2023 – January 2024) in Tanzania (FRESH) and 2) Tanzania National Panel Survey datasets are open sourced national data and collected by Tanzania National Bureau of Statistics in 2014-15 (TNPSW4) and 2020-21 (TNPSW5)

Source and Format: 1) FRESH dataset will be provide by IFPRI and the data will be converted into csv format, 2) TNPS data were downloaded from World Bank Microdata Library (<https://microdata.worldbank.org/index.php/home>) in csv format.

FRESH baseline survey covered 2611 households in Northern Tanzania. TNPSW4 covered 3352 households and TNPSW5 covered 4709 households. FRESH baseline is conducted with a single interview method. TNPSW4 and TNPSW5 survey were designed panel survey and same households with extended households were interviewed.

2. What hardware and software will be used in your research?

Data analysis was conducted in the RStudio (RStudio 2023.06.0 Build 421, Posit Software, PBC).

3. What data-related activities will be performed during the research?

List key data-related activities that you and/or others will perform during the research. For instance, trial draft survey in month 6, collect data in month 8-10, clean and anonymise data in month 11, analyse data in month 12-18.

Task	Description
Preparation of the data analysis (Month 1, July 2024)	Write study protocol and analysis plan Write data management plan Complete and submit LSHTM ethics application Draft analysis plan (including tables) and methodology
Data analysis (Month 2, August)	Prepare R scripts for the analysis FRESH data and conduct analysis (after LSHTM ethical approval receives)
Draft data analysis and manuscript (Month 3, September)	1 st draft to organise data analysis and manuscript
Revisions of the manuscript # 1 (Month 4, October)	Revisions by small working team
Revisions of the manuscript # 2 (Month 5, November)	Revisions and check by all authors (at least two rounds)
Submit a manuscript for the publication (Month 6, December)	Submit paper to a peer-reviewed journal

4. What quality checks will you perform to ensure resources are fit for purpose?

Before the data analysis:

An initial assessment of the data will be performed to understand its structure, format, and any inherent limitations or biases.

During the data analysis:

The data will undergo a cleaning process to address any errors, outliers, or inconsistencies identified during analysis.

After the data analysis:

The documentation of the data analysis will accompany with the datasets (guidance_FRESH-TNPS.doc). It will include; a) person to create the data, b) date of creation, c) the purpose to create the file, d) methodology of the analysis, and e) lists of variables and its units.

5. How will you address ethical & legal issues within your research?

FRESH Initiative in Tanzania obtained the permission from;

- National Institute of Medical Research (NIMR) – NIMR/HQ/R.8a/Vol1X/4537
- IFPRI Institutional Review Board – 00007490
- UC Davis Institutional Review Board - 2066798-1
- Wageningen University Research Ethics Committee - 2066798-1

TNPS dataset is open sourced data and Tanzania Bureau of Statistics collected the data under the supervision of World Bank.

6. What documentation will be created to ensure resources can be understood?

The documentation of the data analysis will accompany with the datasets (guidance_FRESH-TNPS.doc). It will include; a) person to create the data, b) date of creation, c) the purpose to create the file, d) methodology of the analysis, and e) lists of variables and its units.

STORAGE AND SECURITY

7. Where will resources be stored at key stages of your research?

The data are confidential. The original data will be stored in LSHTM SharePoint and the temporary files will be kept in the Home Drive while it is analysing.

In the main folder titled "FRESH-TNPS" will be organised in LSHTM SharePoint. The folder will contain four sub-folders: Datasets, Outputs, and Manuscripts. The "Datasets" sub-folder will include the original datasets from IFPRI and the World Bank Microdata Library, organised into folders named "FRESH24hR," "FRESHHH," "TNPSW4," and "TNPSW5." All data will be formatted as CSV files, with DTA files converted to CSV. The "Outputs" sub-folder will store all outputs generated during the data cleaning ("cleaning" folder), and preliminary analysis stages including figures ("analysis" folder). This sub-folder will also house the datasets and R scripts. The "Manuscripts" sub-folder will contain all drafts of the manuscripts, figures and tables, supplementary materials ("manuscripts" folder), and documentation for publications, including cover letters, authorship agreements ("publications" folder), reviewers' comments, and revised manuscripts ("revising" folder).

8. What labelling conventions will you apply to manage your resources?

The working datasets will be named to reflect the original dataset and version, for example, "FRESH24hR_xxxx_v1.csv" with the version information (i.e. _v1). The variable names will be modified during the data analysis. The minor changes of the version information will be identified as _v.1.1, _v.1.2.

9. How will you keep data safe and secure? (choose one or more)

Only anonymised data will be used - personal, sensitive, or otherwise confidential data is not needed for the research	✓	Store personal details in a separate secure location & link it via an identifier		Delete personal & confidential details at earliest opportunity (specify when below)	
------------------------------------------------------------------------------------------------------------------------	---	----------------------------------------------------------------------------------	--	-------------------------------------------------------------------------------------	--

Use digital storage that require a username/password or other security feature		Physical security (such as locked cabinet or room)		Protect portable devices using security features, e.g. biometric	
Encrypt storage devices		Encrypt during transfer		Avoid cloud services located outside EU	
Take 'Information Security Awareness training'		Ensure backups are also held securely			
Notes:					
Identify additional steps you will take to avoid, reduce, or eliminate risks that may affect your resources.					

ARCHIVING & SHARING

10. What resources should be kept as evidence of your research?

The R scripts will be kept for a minimum of 10 years in LSHTM Data Compass.

11. Where will these resources be hosted?

The sub folder "Datasets" will be stored the original datasets and IFPRI can access the folder.

The working datasets and used R scripts will be kept in "Outputs" folder in LSHTM data repository and it will be open to the public when the manuscript is published.

12. When will the resources be made available? (choose one or more)

During the research life		At the same time as findings are published in an academic journal	✓	A set time after research end, e.g. 12 months. Specify below	
Resources already available (provide details below)		On completion of my thesis		Other (provide details below)	
Further information / Other					

13. How will you make other researchers aware that the resources exist?

Publish a metadata record describing the resources in a repository or other catalogue	✓	Obtain a Digital Object Identifier (DOI) or other permanent ID	
Cite resources in future research papers, e.g. in the data access statement or reference list	✓	Cite resources in project reports	
Publish a description for the project website		Write and publish a Data Paper	
Add resources to a list of your academic outputs			
Other measures / Further details			

--

14. What steps will you take to ensure resources are easy to analyse and use in future research?
(choose one or more)

Prepare a codebook or other documentation that provides an accurate description of content	✓	Store resources in open file formats such as CSV, Rich Text, etc. See https://www.ukdataservice.ac.uk/manage-data/format/recommended-formats	✓
Write a user guide that provides a high-level overview of research		Apply a standard licence that allows a broad range of uses (e.g. Creative Commons, Open Data Commons)	
Designate a corresponding author / data custodian who will handle data-related questions		Use domain-specific standards that make it easy to import and analyse data	
Other / Further information			

15. If resources can be made available, but not openly, what conditions on access/use must be met?
E.g. data can be used for specific types of research only. Leave blank if not applicable.

Requirement:	To be addressed by:

RESOURCING

16. What are the primary data management challenges in your research?
There is no challenges of data management.

17. How can LSHTM & others help you to better manage your data?
The consultation of Data Management Plan was very helpful.