

RHoMIS

Rural Household Multi-Indicator Survey



A GUIDE TO THE DATASETS AND DATA-PROCESSING

v1.6

www.rhomis.org

Introduction

The RHoMIS datasets that we have made available contain information from 13,310 interviews conducted for the purposes of 34 different projects. The interviews took place in 21 countries across the world (Sub-saharan Africa, South America, Central America, South-East Asia and India).

This document aims to explain how to navigate the 'RHoMIS_Open_Data' filing structure in order to make use of the data available. This document also explains the steps involved in processing RHoMIS data and producing the RHoMIS indicator sheet.

Note:

RHoMIS is a research tool which consists of a suite of survey modules, a database of survey responses, data processing code, and the digital infrastructure necessary for the smooth implementation of the tool. The acronym RHoMIS stands for the Rural Household Multi-Indicator Survey.

Section 1: Datasets

All of the datasets available in the 'RHoMIS_Open_Data' folder are described below. In addition, the filepath for each dataset is included to make them as easy as possible to locate. Each dataset contains all the necessary IDs (household IDs, survey IDs, project IDs and country codes) to link information between datasets. It is important to note that outliers have not been removed in any of the datasets so some responses may represent error on the part of the enumerator or misreporting on the behalf of the respondent.

First dataset: RHoMIS Full Data

Filepath: "RHoMIS_Open_Data\Data_to_Share\RHoMIS_Full_Data.csv"

Summary: This dataset contains the responses to all the 'core' RHoMIS questions. These are the questions which are asked across almost every survey implementation. This dataset has been cleaned so that all 'crop_name' and all 'livestock_name' have been standardised (standardisation process outlined in section 2). For this dataset to be made public it has also gone through additional cleaning stages so that the households remain anonymous (see section 2).

Points to note:

- Responses to revenue and price questions are in local currency units (LCU).
- The indicator calculation scripts (see section 2) can be applied to this data set in order to calculate the indicators. However, it must be noted that the scripts provided were designed to calculate indicators on a project-by-project basis.

Second dataset: RHoMIS Indicator Sheet

Filepath: "RHoMIS_Open_Data\Data_to_Share\RHoMIS_Indicators.csv"

Summary: A sheet containing 42 indicators calculated from the 'RHoMIS Full Data' set described above (see section 2 for further detail on the indicator calculations). These indicators cover key areas such as household size, gender control, production value, food availability and dietary diversity.

Points to note:

- The indicator calculations are described in further detail in sections 2 and 3.

Third dataset: PRICES (general description)

These are general comments relating to the prices of farm produce (found in 'RHoMIS_Open_Data\Data_to_Share\Prices') more specific comments on each price sheet can be found below.

Summary: These are the prices used to calculate several of the RHoMIS indicators (further detail in sections 2 and 3). The prices are in international purchasing power parity dollars (int PPP\$) per unit (e.g. kg, litres or heads of livestock). The prices are listed for each individual crop or animal in each individual project. Where 'NA' is listed, this means that there was insufficient data to calculate that particular price for that project (e.g. no farmers interviewed for this project had sold this crop).

Points to note:

- In most cases, the prices in this dataset represent the median price received by farmers in the project. However, where the prices seemed unreasonable, it has been presumed that incomes or yields have been misreported. In these cases, prices have been obtained from FAOstat.

3a. Dataset: Crop prices for each project

Filepath: "RHoMIS_Open_Data\Data_to_Share\Prices\crop_prices_per_kg.csv"

Summary: Crop prices per kg for each individual crop

3b. Dataset: Egg prices for each project

Filepath: "RHoMIS_Open_Data\Data_to_Share\Prices\egg_price_per_egg.csv"

Summary: Price per egg (from each type of poultry).

3c. Dataset: Honey prices for each project

Filepath: "RHoMIS_Open_Data\Data_to_Share\Prices\honey_price_per_litre.csv"

Summary: Price per litre of honey

3d. Dataset: Meat prices for each project

Filepath: "RHoMIS_Open_Data\Data_to_Share\Prices\meat_prices_per_kg.csv"

Summary: Price per kg of meat for each type of livestock butchered

3e. Dataset: Milk prices for each project

Filepath: "RHoMIS_Open_Data\Data_to_Share\Prices\milk_prices_per_l.csv"

Summary: Price per litre of milk for each type of livestock milked.

3f. Dataset: Prices for livestock heads for each project

Filepath: "RHoMIS_Open_Data\Data_to_Share\Prices\whole_livestock_prices.csv"

Summary: Price per head of livestock sold.

Fourth dataset: Crop details

Filepath: "RHoMIS_Open_Data\Data_to_Share\Extra_Outputs\crop_details.csv"

Summary: Details about the crops grown by each household over the 12 months before the interview took place. These details include crop yield, total crop harvested, income from crops sold, amount of crop sold and amount of crop consumed.

Points to note:

- All values relate to the crops grown within the past 12 months. For example, if there have been two harvests, the yield will be (total kg over two harvests)/(area planted).

Fifth dataset: Crop use

Filepath: "RHoMIS_Open_Data\Data_to_Share\Extra_Outputs\crop_use.csv"

Summary: The stated use for each crop (whether the crop was sold, consumed, or saved for seeds)

Sixth dataset: Crops Intercropped

Filepath: "RHoMIS_Open_Data\Data_to_Share\Extra_Outputs\Crops_Intercropped.csv"

Summary: Whether the crop grown was intercropped or not.

Seventh dataset: Crops Monocropped

Filepath: "RHoMIS_Open_Data\Data_to_Share\Extra_Outputs\Crops_Monocropped.csv"

Summary: Whether the crop grown was monocropped.

Eighth dataset: Livestock Details

Filepath: "RHoMIS_Open_Data\Data_to_Share\Extra_Outputs\livestock_details.csv"

Summary: Details on livestock and livestock produce. For each livestock product, the dataset includes the amount collected, the amount sold, the amount consumed and the income from the sale of the livestock product. For whole livestock, this dataset shows the number of livestock heads owned, the number that were sold and the income from the sale of the livestock.

Section 2: Data-Processing

All the datasets listed in section 1 have been processed in some way before they have been made available. This section describes how data has been processed, moving from raw data, to clean data to the indicators and extra outputs made available in the 'RHoMIS_Open_Data' folder. An overview of the processing steps is depicted in figure 1:

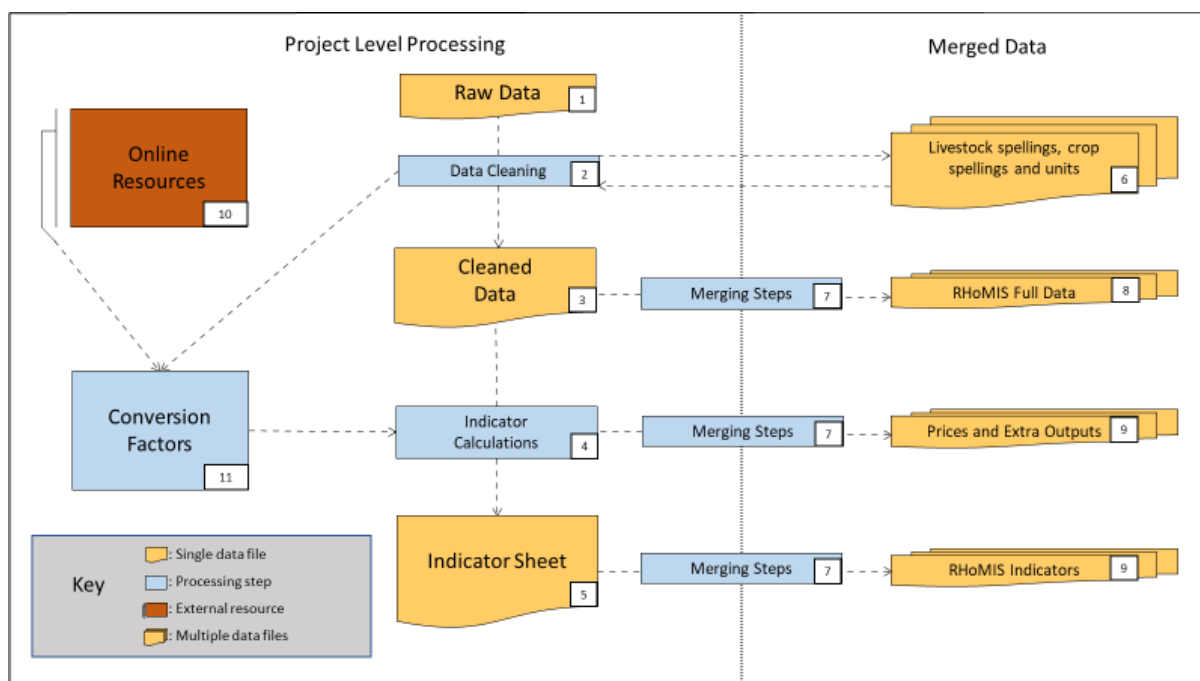


Fig 1. A schematic representing the data-processing steps for RHoMIS data. Project level processing (depicted in the left-hand portion of the partitioned schematic) occurs for each new use of the survey. The project level data is processed and then merged with previous data (depicted in the right-hand portion of the partitioned schematic).

Each point in the data processing schematic (see Fig 1.) is labelled with a number, the remainder of this section will describe the data processing steps with reference to these numbers. The process is as follows:

- Processing begins at point 1, the 'Raw Data'.
- The raw data is cleaned (2). During the cleaning process, new instances from each survey must be documented and put in a standard format. For example, misspelling of crop names e.g. 'sorgum' must be put into the standardised format, aligning with previous projects e.g. 'sorghum'. This standardisation considers crop names, livestock names and new units. The new units must be interpreted and given the appropriate conversion factor. For example, if crop amounts are given in pounds, the conversion factor of 1/(2.2) is stored to convert all crop amounts into kilograms. This conversion factor can then be applied to future projects (points 6 and 11). This cleaning stage results in a 'cleaned' data set (3).

- c) Clean data must go through the 'merging' process before it is ready to be made public. In these steps, household IDs, survey IDs, project IDs and project iteration are attached to the dataset. In addition, the dataset is modified so that the households can remain anonymous: GPS co-ordinates are rounded to 2 decimal places and village names, phone numbers and respondent names are removed. Finally, this processed data is merged with data from previous projects.
- d) The cleaned data (3) is processed, using a series of indicator calculation scripts (see section 3 for more details). Conversion factors (11) and information from online resources (FAOstat, World Bank PPP conversions and PPI conversion tables) are used to calculate a series of indicators, collated into a sheet (5).
- e) During these indicator calculations, there are important intermediate calculations which are made, the results of which are extracted (4->9). These include crop yields, amount of crop consumed, the income from each crop, the income from livestock products and the amounts of each livestock product collected.
- f) The indicator sheet and extra outputs generated are then processed using the same merging process (see point (c)).

Section 3: Indicator Calculations

In sections 1 and 2, the 'indicator calculation' process is briefly mentioned. These calculations, written in the 'R' programming language, are complicated, interlinked and rely on important assumptions which should be understood to gain a full picture of the dataset. As a result, it was more useful to describe the calculations in an excel workbook:

"RHoMIS_Open_Data\Explanation_of_Calculations_and_Outputs.xlsx"

This workbook goes through the calculations in detail and provides search references to follow the calculations in the original R-scripts. For those less interested in the finer details, the indicators used, the main principals of the indicator calculations and the major assumptions made are described below.

Section 3a: Indicators used

Topic	Indicators Used (Unit)	Column name
Household Basics	<ol style="list-style-type: none"> 1. Household size (Members) 2. Household size (Male adult equivalent) 3. Marital status of household heads 4. Education level of household heads 	<ol style="list-style-type: none"> 1. HHsizemembers 2. HHsizeMAE 3. HouseholdType 4. Head_EducationLevel
Land size	<ol style="list-style-type: none"> 1. Land Owned (ha) 2. Land Cultivated (ha) 	<ol style="list-style-type: none"> 1. LandOwned 2. LandCultivated
Food Security	<ol style="list-style-type: none"> 1. Worst food security month 2. Best food security method 3. Number of months food insecure 4. Household food insecurity access scale (HFIAS) 5. Food insecurity experience scale (FIES, asked in some cases in place of HFIAS) 6. Household dietary diversity score (HDDS, asked for good season and bad season, asked for farmbased and purchased foods) 7. Food Availability (kCal/MAE/day) 8. Food Self Sufficiency (kCal/MAE/day) 	<ol style="list-style-type: none"> 1. WorstFoodSecMonth 2. BestFoodSecMonth 3. NrofMonthsFoodInsecure 4. HFIAS_status 5. FIES_Score 6. (score_HDDS_GoodSeason, score_HDDS_farmbasedGoodSeason, score_HDDS_purchasedGoodSeason, score_HDDS_BadSeason, score_HDDS_farmbasedBadSeason, score_HDDS_purchasedBadSeason) 7. Food_Availability_kCal_MAE_day 8. Food_Self_Sufficiency_kCal_MAE_day

Wealth, value of production and incomes	<ol style="list-style-type: none"> 1. Progress out of poverty likelihood 2. Livestock Holdings 3. The total value of activities per male adult equivalent per day (int PPP\$/MAE/day) 4. Total household income (int PPP\$/year) 5. Total household off-farm income(int PPP\$/year) 6. Total household income from on-farm activities (int PPP\$/year) 7. Value of all household activities on-farm (int PPP\$/year) 8. Total income from cropsales (int PPP\$/year) 9. Value of all crop produce (int PPP\$/year) 10. Value of all crops consumed (int PPP\$/year) 11. Total income from livestock and livestock product sales (int PPP\$/year) 12. Value of livestock and livestock produce (int PPP\$/year) 13. Value of livestock and livestock produce consumed (int PPP\$/year) 14. Livestock Orientation 15. Market Orientation 	<ol style="list-style-type: none"> 1. PPI_Likelihood 2. LivestockHoldings 3. TVA_USD_PPP_pmae_pday 4. total_income_USD_PPP_pHH_Yr 5. offfarm_income_USD_PPP_pHH_Yr 6. farm_income_USD_PPP_pHH_Yr 7. value_farm_produce_USD_PPP_pHH_Yr 8. crop_sales_USD_PPP_pHH_Yr 9. value_crop_produce_USD_PPP_pHH_Yr 10. value_crop_consumed_USD_PPP_pHH_Yr 11. livestock_prodsales_USD_PPP_pHH_Yr 12. value_livestock_production_USD_PPP_pHH_Yr 13. value_livestock_prod_consumed_USD_PPP_pHH_Yr 14. Livestock_Orientation 15. Market_Orientation
Other	<ol style="list-style-type: none"> 1. Importance of wildfoods (ordinal) 2. Number of months wildfoods consumed 3. Greenhouse gas emissions (UNIT???) 4. Male control over household produce and incomes 5. Female control over household produce and income 6. Nitrogen fertiliser use 	<ol style="list-style-type: none"> 1. RelImpWildFoods 2. NrofMonthsWildFoodCons 3. GHGEmissions 4. Gender_MaleControl 5. Gender_FemaleControl 6. NFertInput

Principals and Assumptions of RHoMIS Indicator Calculations

The RHoMIS survey is designed to quickly collect information about rural farming households. In order to calculate incomes, harvests and other measures of productivity whilst also minimising interview time it is necessary to make some assumptions. This section will use the example of crop productivity and income calculations to illustrate some of the assumptions made, further details can be found: "RHoMIS_Open_Data\Explanation_of_Calculations_and_Outputs.xlsx".

Worked Example:

When calculating crop productivity, respondents are asked to give the amount harvested over the past 12 months, they are then asked to give the unit of this amount (e.g. 'kg', 'tons', 'gorogoro' etc.). By multiplying the two it is possible to calculate the number of kilograms harvested over the past 12 months (let's label this C_H).

Respondents are also asked how much income they gain from the sale of this crop and the units of this income (e.g. 'total income per year', 'price per kg', 'price per gorogoro' etc.). This can either be used to calculate total income per year or the income per kg (let's label these P_{yr} and P_{kg} respectively).

Using these variables it is necessary to calculate: number of kg sold (C_S), number of kg consumed (C_C), price per kilo ($\$_{kg}$), income per year ($\$_{yr}$), value controlled by the man (Ctr_M), value controlled by the woman (Ctr_W) and the values of the crops which were consumed ($\$_{cons}$). These calculations mainly rely on **proportion responses**.

The respondent is asked what proportion of the harvest was consumed ($prop_{cons}$) and what proportion was sold ($prop_{sold}$), where none=0, little=0.1, under half=0.2, half=0.5, most=0.7, all=0.9. Therefore:

1. $C_C = prop_{cons} \times C_H$
2. $C_S = prop_{sold} \times C_H$

Using the amount sold (C_S) it is possible to work out price per kg and price per kilo ($\$_{kg}$) and income per year ($\$_{yr}$), depending on which sale units were given by the respondent:

$$3. \$_{kg} = \$_{yr} / C_S$$

The price per kg is then averaged across the whole project and verified against FAOstat prices, this makes it possible to calculate the value of crop consumed ($\$_{cons}$):

$$4. \$_{cons} = \$_{kg} \times C_C$$

Finally, the value controlled by the female (Ctr_W) and the male (Ctr_M) in the household can be calculated. The respondent is asked what proportion of the crop consumed is controlled by the male (Ctr_{Mcons}) or the woman (Ctr_{Wcons}), where none=0, little=0.1, under half=0.2, half=0.5, most=0.7, all=0.9. The respondent is then asked what proportion of the income from the cropsale is controlled by the male ($Ctr_{M\$}$) or the woman ($Ctr_{W\$}$). Which finally gives:

5. $Ctr_M = (Ctr_{Mcons} \times \$_{cons}) + (Ctr_{M\$} \times \$_{yr})$
6. $Ctr_W = (Ctr_{Wcons} \times \$_{cons}) + (Ctr_{W\$} \times \$_{yr})$

This style of proportion calculation is used throughout the indicator calculations, for further details please see the excel workbook:

"RHoMIS_Open_Data\Explanation_of_Calculations_and_Outputs.xlsx"

This contains detailed explanations of the indicator calculations and references to the calculations within the R-script.