Food and Agriculture Organization of the United Nations

SDG Indicator 2.4.1

Instructions Manual

on

Data Entry Operations

16/12/2019

Table of contents

Acknowledgements

This manual has been prepared in support of the methodological concept note of SDG Indicator 2.4.1, particularly the data entry operations and analysis of the indicator by Food and Agriculture Organization of the United Nations (FAO) in collaboration with Global Strategy to improve Agricultural and Rural Statistics (GSARS).

The author, Gianluigi Nico would like to thank all those who have contributed to the different aspects of the document especially; Arbab Asfandiyar Khan from Statistics division of FAO, Amy Heyman and Jean Marc Faures of Strategic Programme on Sustainable Agriculture of FAO for their continued supervision and suggestions regarding the content, structure and further improvements of the document.

The author would also like to acknowledge all peer reviewers for their valuable inputs and comments that were provided at various stages, especially Francesco Nicola Tubiello of Statistics Division of FAO.

Introduction

This instruction manual supplements the statistical tool for data entry operations and data analysis of the SDG indicator 2.4.1. It illustrates the applied procedures for data entry operations and data analysis and aims to make the data entry operators and data analysts understand the working of the statistical tool developed by the Food and Agriculture Organization of the United Nations (FAO) and the Global strategy to Improve Agricultural and Rural Statistics (GSARS).

The instruction manual is structured into two main sections as follows:

Section I describes data entry operations, i.e. all operations that must be performed in order to organize the collected data into an excel spreadsheet. Section II describes the procedures to analyse data collected and construct the 11 sub-indicators according to a dashboard approach.

This instruction manual assumes that both enumerators and data analysts are familiar with the survey questionnaire and the methodology for the SDG 2.4.1, respectively. If this is not the case, enumerators and data analysts are strongly encouraged to carefully read and get familiar with the above-mentioned documentations, before proceeding with the reading of this instruction manual.

Section I

Introduction to data entry operations

While the fieldwork is underway, questionnaires will need to be returned to the office on a regular basis so that the data entry operation stays in sync with the fieldwork and does not fall behind. Questionnaires should always be submitted to the home office on a cluster basis, and should be bundled together as such.

FAO and GSARS have prepared a statistical tool to enter data collected through the survey module for the SDG 2.4.1 into an excel spreadsheet. The statistical tool prepared by FAO and GSARS aims to organize data collected through the survey questionnaire (paper and pencil interview, PAPI) into a set of excel spreadsheet in a harmonized way. This section discusses how data entry should be conducted. This includes questionnaire receipt from the field, questionnaire storage while in process, data entry, editing, and the overall management of the data processing operation.

Questionnaire receipt

Immediately upon receipt of the questionnaires from the field, a person from the National Statistical Office, ideally the supervisor of the enumerators in charge for data collection operations, must be assigned to verify the contents of the shipment.

Normally a driver will return several clusters (i.e. enumeration areas) to the office at a time from one or more teams. Upon arrival, there should be a designated area within the storage room, or a small room adjacent to the storage room, that can accommodate these questionnaires.

For each new cluster received, the enumerators' supervisor should conduct the following checks, in the order indicated:

1) Ensure all holding questionnaires within the cluster are sorted in ascending order by holding number (i.e., lowest holding number on top), and that all modules belonging to that holding are in ascending order by module number (i.e. Section I first, and then A, B and finally C).

2) Ensure all geographic and administrative identification codes are correctly completed on the cover sheet of the questionnaire, and that no holding shares the exact combination of identifying codes with another holding.

3) Ensure the proper number of holding questionnaires exist for the cluster; i.e., how many holding should have been interviewed within the cluster? All holding questionnaires need to be returned, even if the interview was partially completed, refused, or otherwise unfinished.

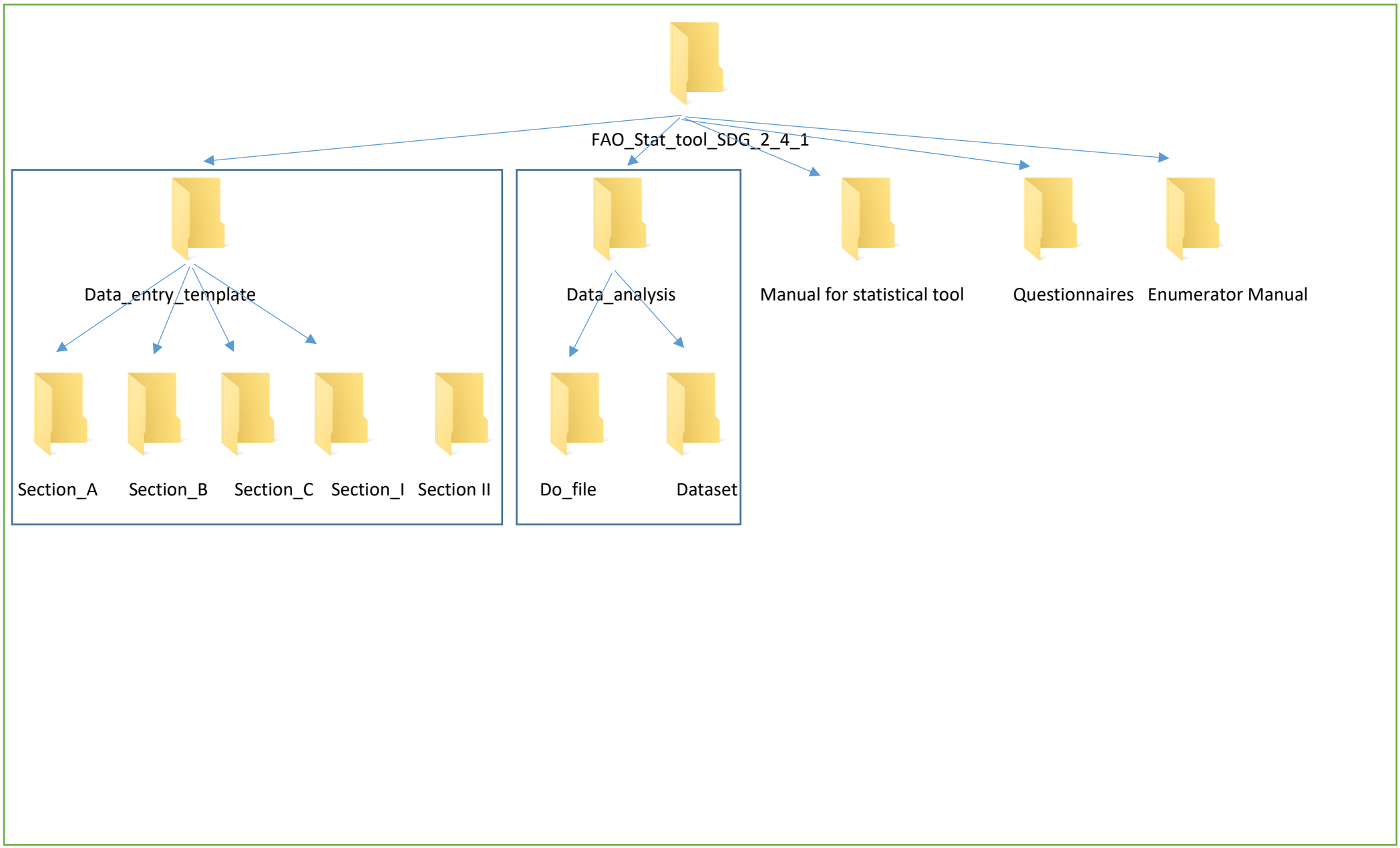
Data entry

Data entry can be defined as the process of entering data into a computerized database or spreadsheet, after data have been collected through face to face interview questions. Data entry operations that are described below are performed by an individual typing at a keyboard.

Data entry spreadsheets

FAO and GSARS have prepared a set of excel spreadsheets to be used by enumerators for entering the collected data. The developed spreadsheets will be given to enumerators in order to perform the related-data entry operations.

Each enumerator will receive a folder named “FAO_Stat_tool_SDG_2_4_1” structured as shown in the box below. The key folders for data entry and analysis are named “Data_entry_template” and “Data_analysis”.


FAO_Stat_tool_SDG_2_4_1
Data_entry_template
Data_analysis
Manual for statistical tool
Questionnaires
Enumerator Manual
Section_A
Section_B
Section_C
Section_I
Section II
Do_file
Dataset

The sub-folders “Data_entry_template” contains five additional sub-folders, namely “Section_A”, “Section_B”, “Section_C” and “Section_I” “Section_II”

The excel spreadsheets for data entry operations are organized within each of these four sub-folders and follows the typical structure of the survey questionnaire, the latter organized as follows:

1) Introductive section (Section_I);
2) Area of the holding section (Section_II);
3) a section on the economic dimension of the holding (Section_A);
4) a section collecting information on the environmental dimension of the holding (Section_B); and, finally,
5) a section collecting information for the social dimension of the holding (Section_C).

Before starting data entry operations, enumerators must get familiar with the excel spreadsheet in which data collected from each of the above 5 sections will be filled in. Yet, **it is strongly advisable that FAO and GSARS statistical tool is saved on enumerator’s desktop before starting data entry operations**

General Structure of data entry spreadsheets

The template for data entry operations “Data_entry_form_Section_XXXX” is comprised of two excel sheets: the first excel sheet “Data_entry_section_XXX” contains all variables that must be filled with data collected via face-to-face interviews. The second excel sheet, “Codebook” contains a set of information to make data entry operations easier. These are:

1) **Variable**, i.e. the variable name as reported in the data entry template (sheet “Data_entry_section_I”, “Data_entry_section_A”, etc);
2) **Question number,** i.e. the question number as reported in the paper questionnaire;
3) **Question,** i.e. the verbatim question as reported in the paper questionnaire;
4) **Variable description**, the description of the variable in a concise way;
5) **Type of variable:** whether the variable must be filled in the data entry template as a string or as a numeric variable
6) **Numeric codes allowed**: each question asked in the questionnaire has its corresponding numeric code. For instance, female takes on value 2 while male takes on value 1. The numeric code is described for each variable and allows to avoid data entry mistakes. It also informs data entry operators on whether (or not) missing values are allowed for the variable under investigation
7) **Labels associated with numeric codes,** i.e. the value label that allows converting the numeric code associated with a given variable with the corresponding label

During data entry operations, enumerators are strongly encouraged to constantly consult the codebook in order to avoid data entry mistakes and ensure harmonization of data entry operations among enumerators themselves.

Box 1 below, shows how the data entry template is organized.

Box 1. Organization of the data entry template: variable to be filled in and the corresponding codebook

HHID	I_101031	I_101041	I_101051	I_102011	I_103011
0001	1	4	1	3	household sector
0002	2	4	1	1	Non-household sector
0003	1	4	1	3	Non-household sector
0004	2	3	1	1	Non-household sector
0005	2	1	1	1	Non-household sector
0006	2	2	1	2	Non-household sector
0007	1	2	1	2	Non-household sector
0008	1	2	1	3	household sector
0009	1	4	1	1	household sector
0010	2	5	1	3	household sector
0011	1	5	1	3	household sector
0012	2	5	1	1	household sector

Variable	Question number	Question	Variable description	Type of variable	Numeric codes allowed	labels associated with numeric codes
HHID	HHID	Holding Identification Number:	Identification number of the sampled holding	Numeric/general	from 00001 to 00350	Assigned by NSO
101031	I.1.3	Sex	Sex of the respondent	Numeric/general	1; 2	1= Male; 2=Female
101041	I.1.4	What is your function on the agricultural holding?	Respondent function in the holding	Numeric/general	1;2;3;4;5	1= Holder (legal and/or economically responsible for the holding) 2=Co-holder (legal and/or economically co-responsible for the holding) 3=Manager (responsible for the day-to-day decisions on the farming operations) 4= Household member working on the holding 5=Employee
101051	I.1.5	Are you able to answer questions for the agricultural holding?	Respondent awareness about socio-economic and environmental characteristics of the holding	Numeric/general	1; 2	1=Yes 2=No
102011	I.2	What is the legal status of the Holder?	Legal status of the holding	Numeric/general	1;2;3	1=Civil person/natural person 2=Group of civil persons/natural persons 3=Legal person
103011	I.3	'What type of holding is this?	Country-specific information to be pre-filled by the NSO	String/text		

General Rules for data entry operations

To ensure consistency, high quality of data entry and minimise delays some general rules need to be observed during the data entry process.

1) When to start entering data? It is advised to start data entry operation as soon as possible after the start of data collection and as soon as small batches of completed questionnaire are made available. This helps minimize delays and maximise opportunities for checking and validating data.

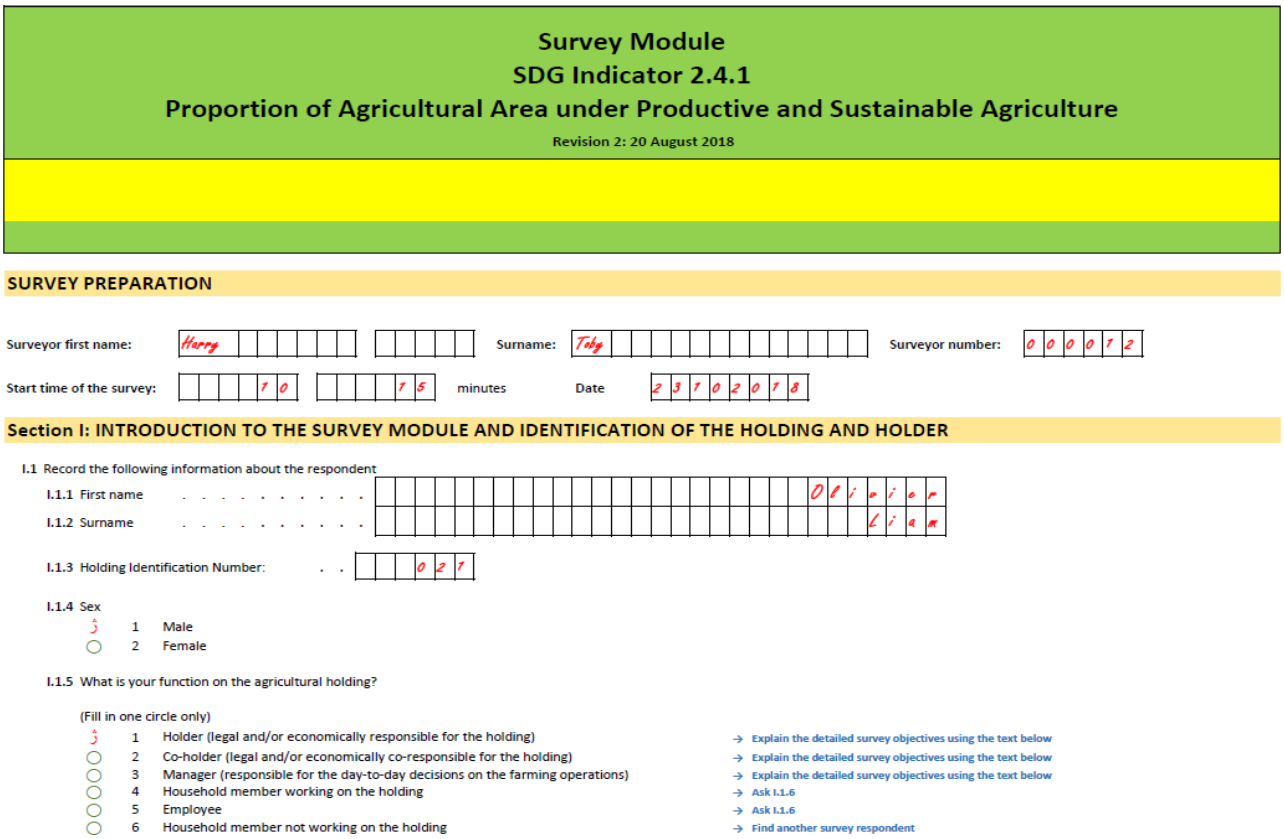
2) Confidentiality: Keep all data and information confidential. This includes not leaving paper lying where people outside of the team can see them, not discussing results or participants with anyone other than necessary to process the data, and protecting computers and files.

3) Data entry: please make sure to enter the data according to the codes associated with each question asked during face-to-face interview. Look at the example below to make sure that this step is fully understood.

Example number 1:

Let us imagine that you have just finished to collect data from the field and you now need to enter the data into the excel spreadsheet that FAO and GSARS have provided to you. You will have your paper questionnaire on your office desk and the excel spreadsheet template for data entry operations opened on your computer. You will start by entering data collected in Section I ("**introduction to the survey module and identification of the holding and holder**"), and then proceed with Section II, A, B and C, at a later stage.

The questionnaire on your desk was filled during field data collection with the information provided by the selected respondent, as per below example:



Survey Module
SDG Indicator 2.4.1
Proportion of Agricultural Area under Productive and Sustainable Agriculture
Revision 2: 20 August 2018

SURVEY PREPARATION

Surveyor first name: Harry Surname: Toby Surveyor number: 00012

Start time of the survey: 10 15 minutes Date 23/02/2018

Section I: INTRODUCTION TO THE SURVEY MODULE AND IDENTIFICATION OF THE HOLDING AND HOLDER

I.1 Record the following information about the respondent

I.1.1 First name . . . . . Olivier

I.1.2 Surname . . . . . Liam

I.1.3 Holding Identification Number: . . 021

I.1.4 Sex

- (•) 1 Male
- ( ) 2 Female

I.1.5 What is your function on the agricultural holding?

(Fill in one circle only)

- (•) 1 Holder (legal and/or economically responsible for the holding) → Explain the detailed survey objectives using the text below
- ( ) 2 Co-holder (legal and/or economically co-responsible for the holding) → Explain the detailed survey objectives using the text below
- ( ) 3 Manager (responsible for the day-to-day decisions on the farming operations) → Explain the detailed survey objectives using the text below
- ( ) 4 Household member working on the holding → Ask I.1.6
- ( ) 5 Employee → Ask I.1.6
- ( ) 6 Household member not working on the holding → Find another survey respondent

You must now open the excel spreadsheet template for data entry contained under section I **"FAO_Stat_tool_SDG_2_4_1\Data_entry_template\Section_I\Data_entry_form_section_I"** .

Start by exploring its content, which will look like the one below:

I_001011	I_001012	I_001013	I_001021	I_001022	I_001023	I_101011	I_101021	I_101031	I_101041	I_101051

The codebook provided in the second sheet of the same excel spreadsheet contains all the information that you need in order to proceed with data entry operations. For instance, variable I_001031 corresponds to question **I.1.3 Holding identification number, whose code is 0021, as per survey questionnaire.** It is strongly advisable that you check the codebook before entering data.

I_001011	I_001012	I_001013	I_001021	I_001022	I_001023	I_101011	I_101021	I_101031	I_101041	I_101051
HARRY	TOBY	00001	10	15	23/10/2018	OLIVER	LIAM	0021	1	1
….	….	….	….	….	….	….	….	….	….	….
….	….	….	….	….	….	….	….	….	….	….
….	….	….	….	….	….	….	….	….	….	….

Concerning question **I.1.4** "What is your function in the agricultural holding", you must transcript in the data entry template the code associated to the function of the respondent in the agricultural holding. In the example above, the function of the respondent is "**Holder (legal and/or economically responsible for the holding)**", whose corresponding code is **1. Please, make sure to transcript the code, not the associated label.** The same applies to many other variables, e.g. variable I_101051 (corresponding to question I.1.3. "Sex").

A few notes on the data entry template for section A: economic dimension of the holding

Under the folder Section_A, you will find four data entry spreadsheets, instated of one as for the other sections. While the logic applied for data entry operations remains the same (see example 1), this section of the survey questionnaire is much more complex. This is because Section A "**economic dimension of the holding**" contains a considerable number of variable to capture:

1) information on holding profitability and resilience;
2) Value of production of the holding from crop-related activities;
3) Value of production of the holding from livestock-related activities; and
4) Value of production of the holding from on-farm activities.

In order to avoid an excessive burden in data entry operations, 5 excel spreadsheet have been developed by FAO and GSARS, once for each of the above-mentioned topic covered by the survey questionnaire.

Under this folder (Section_A), you will find the following 5 excel spreadsheet for data entry:

- **Data_entry_form_Section_A**: used for data entry of questions **from A.1 and A.7-A.8; Data_entry_form_Section_A_Area_value_added_crop**: used for data entry of questions **A.02;**
- **Data_entry_form_Section_A_Area_value_added_livestock**: used for data entry of questions **A.03;**
- **Data_entry_form_Section_A_Area_value_added_on_farm**: used for data entry of questions **A.06;**

You will notice that, given the complexity through which most of these questions are collected, a considerable number of variables correspond to one single question. This is the case, for instance, of questions A.02, "What was the total value of crops and its by-products produced by the holding?", whose total number of variables needed for data entry operations is equal to 65.

Section II

Introduction to data analysis

Once data collection through paper questionnaires have been properly transcript in the template for data entry, as per Section I above, you will need to start data analysis for the 11 sub-indicator.

The Food and Agriculture Organization of the United Nations (FAO) and the Global strategy to improve agricultural and rural statistics (GSIARS) have prepared a statistical tool to analyse the collected data into a harmonized way across surveys and countries. In this section we will discuss:

1. Anonymization of sensitive information collected.
2. Labelling of variables and values to facilitate data analysis.
3. Data analysis procedure.

Before starting the exploration of this section, please make sure that you have installed in your computer the statistical software "Stata", which is functional to data analysis operations.

Data anonymization

Data anonymization is a type of procedure whose intent is privacy protection of sensitive information (e.g. name and family name of the respondent, address of the respondent, etc.). It is the process of encrypting personally identifiable information from the datasets that have been constructed, so that the people whom the data describe remain anonymous.

Sensitive information collected for the calculation of the SDG 2.4.1 will be hidden automatically from the datasets by using a dedicated script that you will find under the folder "Data_analysis", sub-folder "**Do.file**".

Before explaining how the anonymization process works, it is worth to remind what a **do.file** is, since the latter constitutes the most important instrument for the data analysis process as a whole. Rather than typing commands at the keyboard, FAO and GSAR have created a text file containing commands and instruct Stata to execute the anonymization process which is stored in that file. Such files are called do-files because the command that causes them to be executed is do (see figure below=. A do-file is a standard text file that is executed by Stata when you type do-filename. By writing a standard do.file you can easily manage to anonymize your data and you can reproduce your work later. Also, writing a do-file makes the inevitable debugging process much easier. If you decide to change one part of your analysis, changing the relevant commands in your do-file is much easier than having to start back at square one, as is often necessary when working interactively.

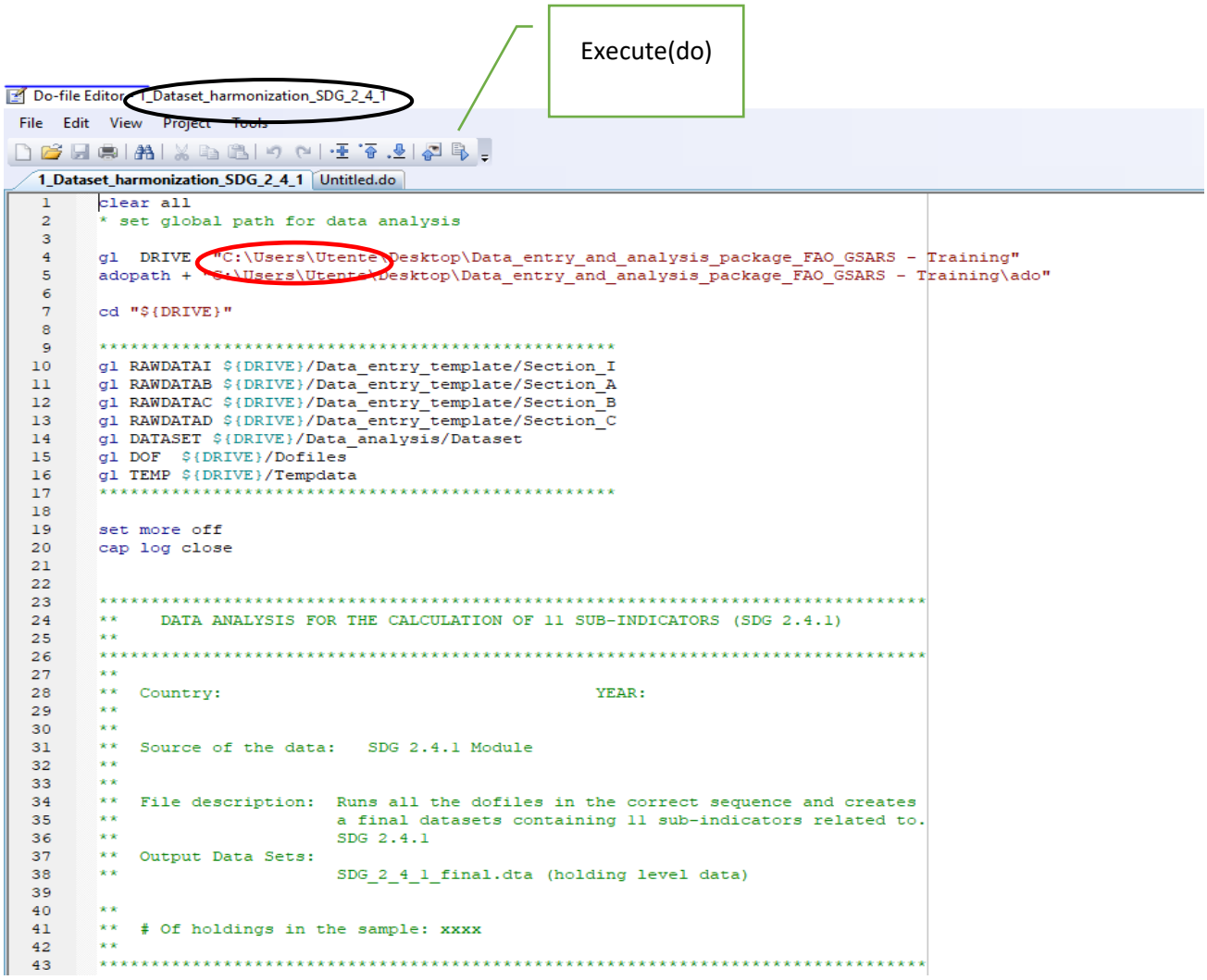
In order to anonymize sensitive information you need to open the folder "data_analysis" and the corresponding sub-folder named "do.file". The latter contains a set of standardized do.files that have been created for data analysis and data anonymization/harmonization, including, therefore, the do file "1_Dataset_harmonization_SDG_2_4_1".

The set of commands contained in this do.file (1_Dataset_harmonization_SDG_2_4_1) allow to:

1) anonymize sensitive information collected

2) attaching label to variables and coded responses given by the sample holdings in ordet to make the subsequent data analysis easier.

The do.file “1_Dataset_harmonization_SDG_2_4_1” is represented in the below figure



Before running the do file for anonymizing data, you need to change the working directory according to your computer path. This is to say that the path “C:\Users\Utente” need to be replaced by your own path (e.g. **C:\userename\username**)

The global path is the main folder from which all dataset that you need for data analysis and data entry are stored. All datasets that you need for data analysis and anonymization are found under different sub-folder, which are

1. Data_entry_template/Section_I
2. Data_entry_template/Section_II
3. Data_entry_template/Section_A
4. Data_entry_template/Section_B
5. Data_entry_template/Section_C

The lines in the do.file (figure above) from 9 to 16 allow setting the working directory in Stata, which is the folder from which each dataset will be imported into Stata.

In order to anonymize sensitive data, you will have to hide some personal information that are contained in Section I of the survey questionnaire. To this aim you need to, first import the excel file "Data_entry_form_Section_I.xlsx" under the sub-folder "Section_I". Data will do it automatically as per figure below. What stata does, is to import the excel file used for data entry operation and run a set of command that allows to hide sensitive information collected.

```
* Input dataset: Section I:                                    /*INTRODUCTION TO THE SURVEY MODULE AND IDENTIFICATION OF THE HOLDING AND HOLDER */

import excel "$RAWDATAI\Data_entry_form_Section_I.xlsx", sheet("Data_entry_section_I") firstrow

*********************************************
*        Anonymization of sensitive data     *
*********************************************

* Anonymizing string variables

foreach var in I_001011 I_001012 I_101011 I_101021 I_104011 I_105011 I_105021 {
replace `var' = " "
destring `var', replace
}

* Anonymizing both string variables and numeric variable
foreach var in I_001011 I_001012 I_101011 I_101021 I_104011 I_105011 I_105021 ///
I_106011 I_106021 {
replace `var'=1

    lab values `var' `var'
        lab def `var' 1 "xxxxxxxxxx"
}
```

The anonymization procedure works in a simple way: it creates a loop (**for each variable to be anonymized**) according to which a set of variables containing sensitive information are first replaced with missing values and then replaced with a set of "xxxxx" values that allows hiding the original information.

In order to do so, you will only need to change the Stata path (line 3 and 5 of the Stata do.file "1_Dataset_harmonization_SDG_2_4_1" and push the button Excute (do) in the upper right of the do.file. The anonymized dataset will look like the one below:

Before anonymization

	I_001011	I_001012	I_001013	I_001021	I_001022	I_001023	I_101011	I_101021
1	HARRY	TOBY	00001	6	26	03aug2018	OLIVER	LIAM
2	HARRY	TOBY	00001	17	23	04aug2018	JACK	ALEXANDER
3	HARRY	TOBY	00001	24	19	05aug2018	JACOB	ASHTON
4	HARRY	TOBY	00001	17	50	06aug2018	RILEY	CONNOR
5	HARRY	TOBY	00001	4	41	07aug2018	ETHAN	HARVEY
6	HARRY	TOBY	00001	8	59	08aug2018	ALFIE	LUKE
7	HARRY	TOBY	00001	4	57	09aug2018	HARRY	KIAN
8	HARRY	TOBY	00001	17	22	10aug2018	DYLAN	CAMERON
9	HARRY	TOBY	00001	13	30	11aug2018	THOMAS	IESTYN
10	HARRY	TOBY	00001	16	41	12aug2018	CHARLIE	JENSON
11	HARRY	TOBY	00001	11	43	13aug2018	LOGAN	IOAN
12	HARRY	TOBY	00001	17	15	14aug2018	JOSHUA	COREY
13	HARRY	TOBY	00001	23	21	15aug2018	WILLIAM	MICHAEL
14	HARRY	TOBY	00001	3	51	16aug2018	RHYS	ADAM
15	HARRY	TOBY	00001	0	32	17aug2018	JAMES	ALEX
16	HARRY	TOBY	00001	22	27	18aug2018	SAMUEL	ZACHARY
17	HARRY	TOBY	00001	7	7	19aug2018	NOAH	AARON
18	HARRY	TOBY	00001	11	3	20aug2018	JAYDEN	BLAKE
19	HARRY	TOBY	00001	7	29	21aug2018	MASON	CALEB
20	MAX	RHYS	00002	0	24	22aug2018	TYLER	ELLIS
21	MAX	RHYS	00002	15	9	23aug2018	FINLEY	MATTHEW
22	MAX	RHYS	00002	23	36	24aug2018	DANIEL	AIDEN
23	MAX	RHYS	00002	14	56	25aug2018	MAX	HENRY
24	MAX	RHYS	00002	6	5	26aug2018	LEO	OWAIN
25	MAX	RHYS	00002	14	13	27aug2018	JAKE	KAYDEN

After anonymization

	I_001011	I_001012	I_001013	I_001021	I_001022	I_001023	I_101011	I_101021
1	xxxxxxxxxx	xxxxxxxxxx	00001	6	26	03aug2018	xxxxxxxxxx	xxxxxxxxxx
2	xxxxxxxxxx	xxxxxxxxxx	00001	17	23	04aug2018	xxxxxxxxxx	xxxxxxxxxx
3	xxxxxxxxxx	xxxxxxxxxx	00001	24	19	05aug2018	xxxxxxxxxx	xxxxxxxxxx
4	xxxxxxxxxx	xxxxxxxxxx	00001	17	50	06aug2018	xxxxxxxxxx	xxxxxxxxxx
5	xxxxxxxxxx	xxxxxxxxxx	00001	4	41	07aug2018	xxxxxxxxxx	xxxxxxxxxx
6	xxxxxxxxxx	xxxxxxxxxx	00001	8	59	08aug2018	xxxxxxxxxx	xxxxxxxxxx
7	xxxxxxxxxx	xxxxxxxxxx	00001	4	57	09aug2018	xxxxxxxxxx	xxxxxxxxxx
8	xxxxxxxxxx	xxxxxxxxxx	00001	17	22	10aug2018	xxxxxxxxxx	xxxxxxxxxx
9	xxxxxxxxxx	xxxxxxxxxx	00001	13	30	11aug2018	xxxxxxxxxx	xxxxxxxxxx
10	xxxxxxxxxx	xxxxxxxxxx	00001	16	41	12aug2018	xxxxxxxxxx	xxxxxxxxxx
11	xxxxxxxxxx	xxxxxxxxxx	00001	11	43	13aug2018	xxxxxxxxxx	xxxxxxxxxx
12	xxxxxxxxxx	xxxxxxxxxx	00001	17	15	14aug2018	xxxxxxxxxx	xxxxxxxxxx
13	xxxxxxxxxx	xxxxxxxxxx	00001	23	21	15aug2018	xxxxxxxxxx	xxxxxxxxxx
14	xxxxxxxxxx	xxxxxxxxxx	00001	3	51	16aug2018	xxxxxxxxxx	xxxxxxxxxx
15	xxxxxxxxxx	xxxxxxxxxx	00001	0	32	17aug2018	xxxxxxxxxx	xxxxxxxxxx
16	xxxxxxxxxx	xxxxxxxxxx	00001	22	27	18aug2018	xxxxxxxxxx	xxxxxxxxxx
17	xxxxxxxxxx	xxxxxxxxxx	00001	7	7	19aug2018	xxxxxxxxxx	xxxxxxxxxx
18	xxxxxxxxxx	xxxxxxxxxx	00001	11	3	20aug2018	xxxxxxxxxx	xxxxxxxxxx
19	xxxxxxxxxx	xxxxxxxxxx	00001	7	29	21aug2018	xxxxxxxxxx	xxxxxxxxxx
20	xxxxxxxxxx	xxxxxxxxxx	00002	0	24	22aug2018	xxxxxxxxxx	xxxxxxxxxx
21	xxxxxxxxxx	xxxxxxxxxx	00002	15	9	23aug2018	xxxxxxxxxx	xxxxxxxxxx
22	xxxxxxxxxx	xxxxxxxxxx	00002	23	36	24aug2018	xxxxxxxxxx	xxxxxxxxxx
23	xxxxxxxxxx	xxxxxxxxxx	00002	14	56	25aug2018	xxxxxxxxxx	xxxxxxxxxx
24	xxxxxxxxxx	xxxxxxxxxx	00002	6	5	26aug2018	xxxxxxxxxx	xxxxxxxxxx
25	xxxxxxxxxx	xxxxxxxxxx	00002	14	13	27aug2018	xxxxxxxxxx	xxxxxxxxxx
26	xxxxxxxxxx	xxxxxxxxxx	00002	12	56	28aug2018	xxxxxxxxxx	xxxxxxxxxx

Data harmonization

Data harmonization is the process of replacing coded associated with the variables collected with a set of labels that allows to easily identify the variable content (e.g. male/female, yes/no, etc), as well as adding a short description to the variable (variable label).

As for the case of data anonymization, this procedure is done automatically assuming that data entry operations were done properly.

The data harmonization process is performed immediately after the anonymization of sensitive information. The set of commands to harmonize datasets has been developed by FAO and GSARS and contained in the do. file "1_Dataset_harmonization_SDG_2_4_1".

There is very little to be done, apart from ascertaining that all variables have been properly stored during data entry operations.

The do file "1_Dataset_harmonization_SDG_2_4_1" allows to:

1) associate a label to each code in a given variable;

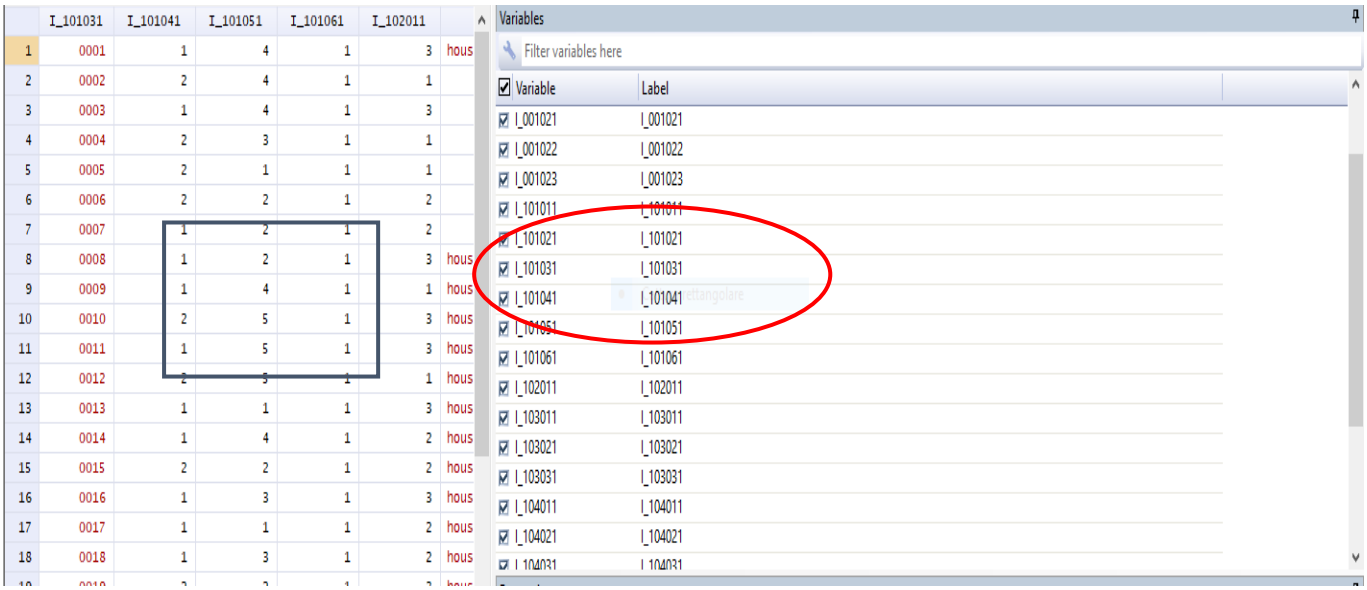
2) associate a short description to the variable name.

```
**************************************
defining labels for numeric variable*
**************************************
       lab values I_101031 I_101031                                    /*(sex of the respondent) */

       lab def I_101031 1 "Male"
       lab def I_101031 2 "Female", add



       lab values I_101041 I_101041                                    /* respondent function in the holding*/

       lab def I_101041 1 "Holder (legal and/or economically responsible for the holding)"
       lab def I_101041 2 "Co-holder (legal and/or economically co-responsible for the holding)", add
       lab def I_101041 3 "Manager (responsible for the day-to-day decisions on the farming operations)", add
       lab def I_101041 4 "Household member working on the holding", add
       lab def I_101041 5 "Employee", add
```

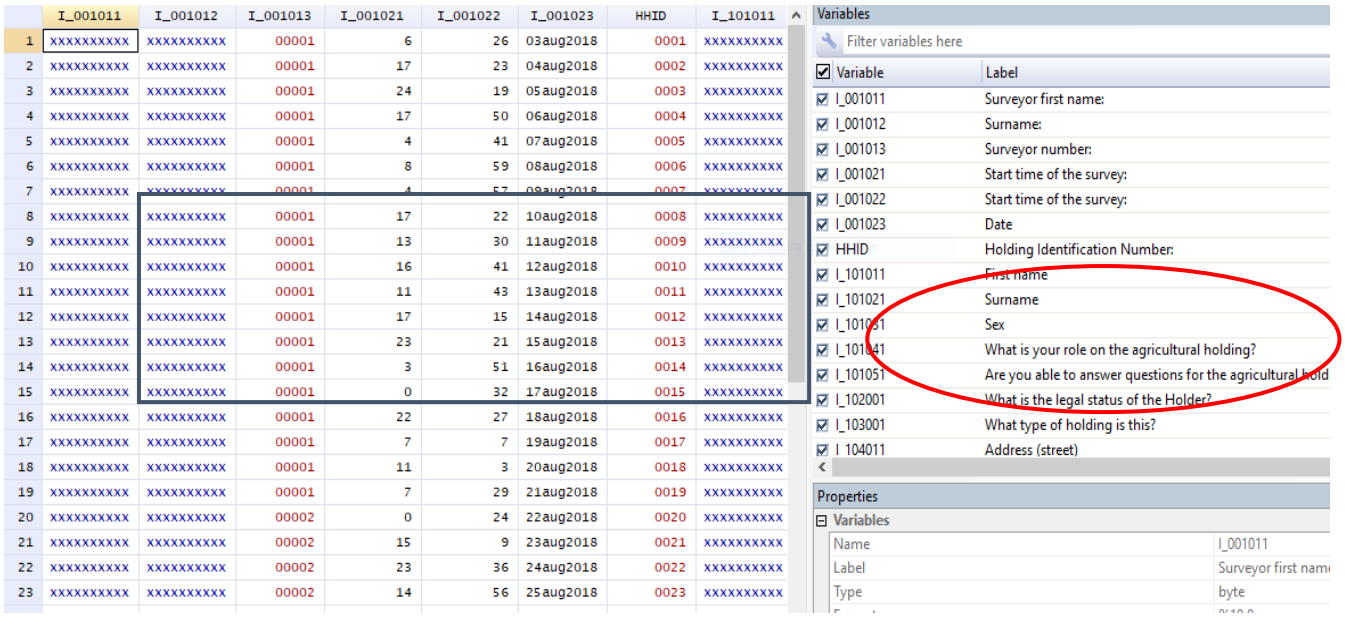
For instance, variable **I_101031** gives information about the sex of the respondent, which is **1** for male and **2** for female. During data entry operations, enumerator only filled the template with code 1 or 2. Data harmonization procedure allows assigning a label to the two codes. By doing so, 1 will be replaced with the label “Male” and 2 with the label “Female”.

In the same vein, this procedure also attaches a short description to the variable, which is: variable I**_101031 will be labelled as “Sex of the respondent”. The two figures bellows shows the before-after output of the data harmonization process.**

Before data harmonization



After data harmonization



	I_001011	I_001012	I_001013	I_001021	I_001022	I_001023	HHID	I_101011
1	xxxxxxxxxx	xxxxxxxxxx	00001	6	26	03aug2018	0001	xxxxxxxxxx
2	xxxxxxxxxx	xxxxxxxxxx	00001	17	23	04aug2018	0002	xxxxxxxxxx
3	xxxxxxxxxx	xxxxxxxxxx	00001	24	19	05aug2018	0003	xxxxxxxxxx
4	xxxxxxxxxx	xxxxxxxxxx	00001	17	50	06aug2018	0004	xxxxxxxxxx
5	xxxxxxxxxx	xxxxxxxxxx	00001	4	41	07aug2018	0005	xxxxxxxxxx
6	xxxxxxxxxx	xxxxxxxxxx	00001	8	59	08aug2018	0006	xxxxxxxxxx
7	xxxxxxxxxx	xxxxxxxxxx	00001	4	57	09aug2018	0007	xxxxxxxxxx
8	xxxxxxxxxx	xxxxxxxxxx	00001	17	22	10aug2018	0008	xxxxxxxxxx
9	xxxxxxxxxx	xxxxxxxxxx	00001	13	30	11aug2018	0009	xxxxxxxxxx
10	xxxxxxxxxx	xxxxxxxxxx	00001	16	41	12aug2018	0010	xxxxxxxxxx
11	xxxxxxxxxx	xxxxxxxxxx	00001	11	43	13aug2018	0011	xxxxxxxxxx
12	xxxxxxxxxx	xxxxxxxxxx	00001	17	15	14aug2018	0012	xxxxxxxxxx
13	xxxxxxxxxx	xxxxxxxxxx	00001	23	21	15aug2018	0013	xxxxxxxxxx
14	xxxxxxxxxx	xxxxxxxxxx	00001	3	51	16aug2018	0014	xxxxxxxxxx
15	xxxxxxxxxx	xxxxxxxxxx	00001	0	32	17aug2018	0015	xxxxxxxxxx
16	xxxxxxxxxx	xxxxxxxxxx	00001	22	27	18aug2018	0016	xxxxxxxxxx
17	xxxxxxxxxx	xxxxxxxxxx	00001	7	7	19aug2018	0017	xxxxxxxxxx
18	xxxxxxxxxx	xxxxxxxxxx	00001	11	3	20aug2018	0018	xxxxxxxxxx
19	xxxxxxxxxx	xxxxxxxxxx	00001	7	29	21aug2018	0019	xxxxxxxxxx
20	xxxxxxxxxx	xxxxxxxxxx	00002	0	24	22aug2018	0020	xxxxxxxxxx
21	xxxxxxxxxx	xxxxxxxxxx	00002	15	9	23aug2018	0021	xxxxxxxxxx
22	xxxxxxxxxx	xxxxxxxxxx	00002	23	36	24aug2018	0022	xxxxxxxxxx
23	xxxxxxxxxx	xxxxxxxxxx	00002	14	56	25aug2018	0023	xxxxxxxxxx

Data analysis: methodology to construct the 11 sub-indicator

The methodology note endorsed by the IAEG-SDG defines the Indicator 2.4.1 as “Proportion of agricultural area under productive and sustainable agriculture”, which is expressed by the following formula:

$$SDG2.4.1 = \frac{Area\ under\ productive\ and\ sustainable\ agriculture}{Agricultural\ land\ area*}$$

This implies the need to measure both the extent of land under productive and sustainable agriculture (the numerator), as well as the extent of land area under agriculture (the denominator). The nominator is the subject of this note and its computation is described in the sections “Assessing sustainability performance for each sub-indicator” and “Reporting the indicator at national level”. The denominator, in turn, is a function of the scope of the indicator. It is the agricultural land area managed by agricultural holdings, defined as the sum of agricultural area utilized by agricultural holdings that are owned (excluding rented-out), rented-in, leased, sharecropped or borrowed.

The proposed list of themes and sub-indicators was obtained through a series of consultations. In total 11 themes are included. The questionnaire modules contain a minimum set of questions needed to measure each sub-indicator at holding level have.

No.	Themes	Sub-indicators
1	Land productivity	Farm output value per hectare
2	Profitability	Net farm income
3	Resilience	Risk mitigation mechanisms
4	Soil health	Prevalence of soil degradation
5	Water use	Variation in water availability
6	Fertilizer pollution risk	Management of fertilizers
7	Pesticide risk	Management of pesticides
8	Biodiversity	Use of agro-biodiversity-supportive practices
9	Decent employment	Wage rate in agriculture
10	Food security	Food insecurity experience scale (FIES)
11	Land tenure	Secure tenure rights to land

For each sub-indicator, criteria to assess sustainability levels are developed. The concept of sustainability implies an idea of continuous progress and improvement towards better performances across all themes, and such performances can therefore be more or less sustainable. In order to capture the concept of continuous progress towards sustainability, a 'traffic light' approach is proposed, in which three sustainability levels are considered for each sub-indicator:

- Green: desirable
- Yellow: acceptable
- Red: unsustainable.

While a certain level of subjectivity is unavoidable, this approach allows identification, for each theme, of conditions of critical unsustainability (red), conditions that can be considered 'ideal' (green) and, in between, intermediate conditions that are considered 'acceptable' but would need to be scrutinized in terms of possible improvements (yellow). This approach also acknowledges the trade-offs existing between sustainability dimensions and themes, and the need to find an acceptable balance between them. Each sub-indicator is assessed at the level of the agricultural holding. The sustainability level is then associated with the agricultural land area of the agricultural holding. All sub-indicators for a given agricultural holding therefore refer to the same agriculture land area.

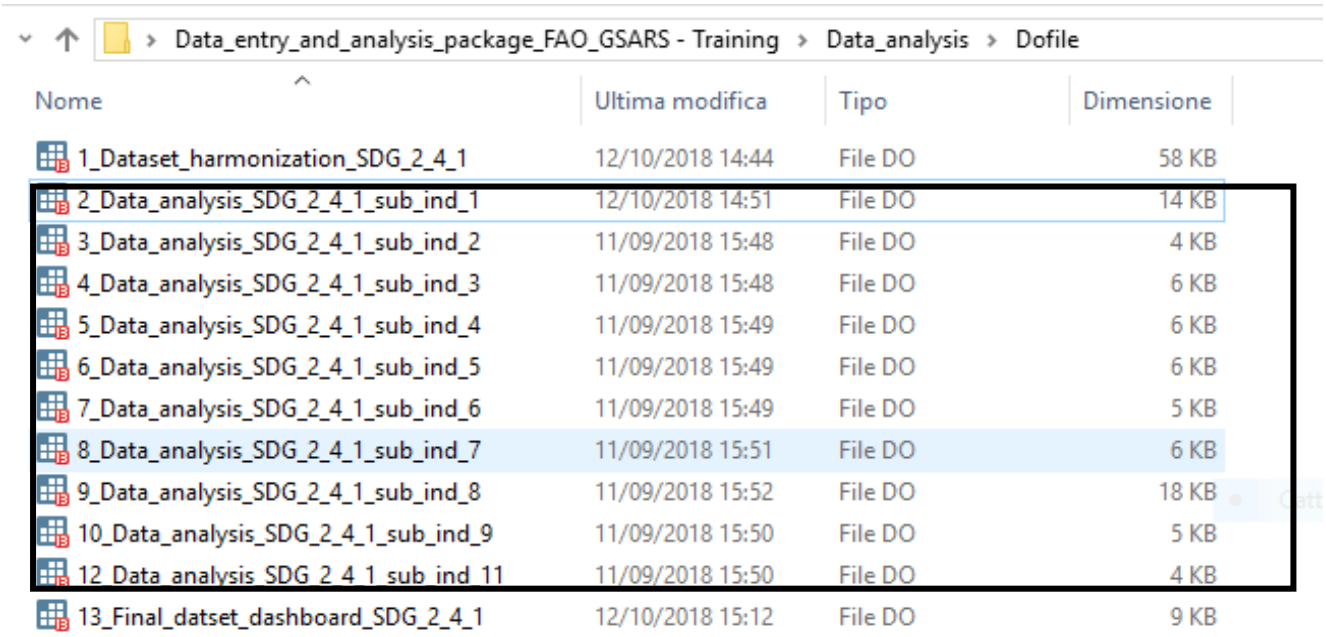
The revised methodology proposes to focus on a dashboard presenting the different sub-indicators separately. The dashboard is chosen for reporting the indicator, as sustainability is about finding an acceptable balance between its three dimensions. It offers several advantages, including the possibility of combining data from different sources and clarity about the main unsustainability issues: countries can easily visualize their performance in terms of the different sustainability dimensions and themes, and understand where policy efforts can be focused (see below).

Computation of results and construction of the dashboard is performed for each sub-indicator separately: for each sub-indicator, aggregation at national level is done by summing the agricultural land area of all agricultural holdings by sustainability category (red, yellow or green), and reported as percentage of the total agricultural land area of the country (minus the common land, as discussed earlier).

Data analysis: construction of the 11 sub-indicators through the statistical tool

The statistical tool developed by FAO and GSARS allows to easily construct the 11 sub-indicators according to the proposed dashboard approach. 11 do.files have been developed accordingly, each of them containing a set of standardized scripts to derive the corresponding 11 sub-indicators.

The do files are found under the folder “Data_analysis”, sub-folder “Dofile” (see figure below). An additional do file “Final_dataset_dashboard_SDG_2_4_1” was also developed to aggregate the 11 sub-indicators at the national level according to the above-mentioned dashboard approach.

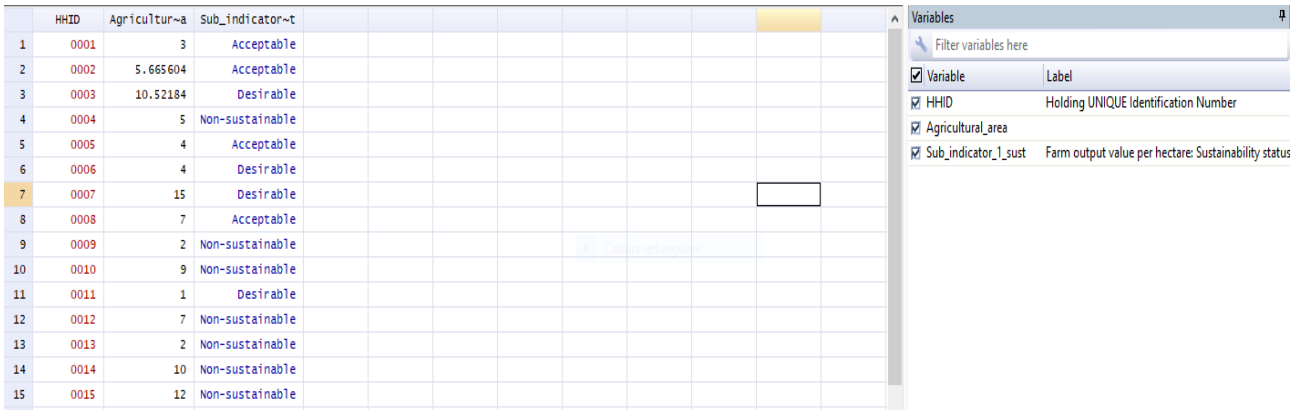


Data analysists must execute the 11 do.files consequently which is, in ascending order from sub-indicator number 1 to sub-indicator number 11. The do file does not need any relevant modification, apart from changing the path.

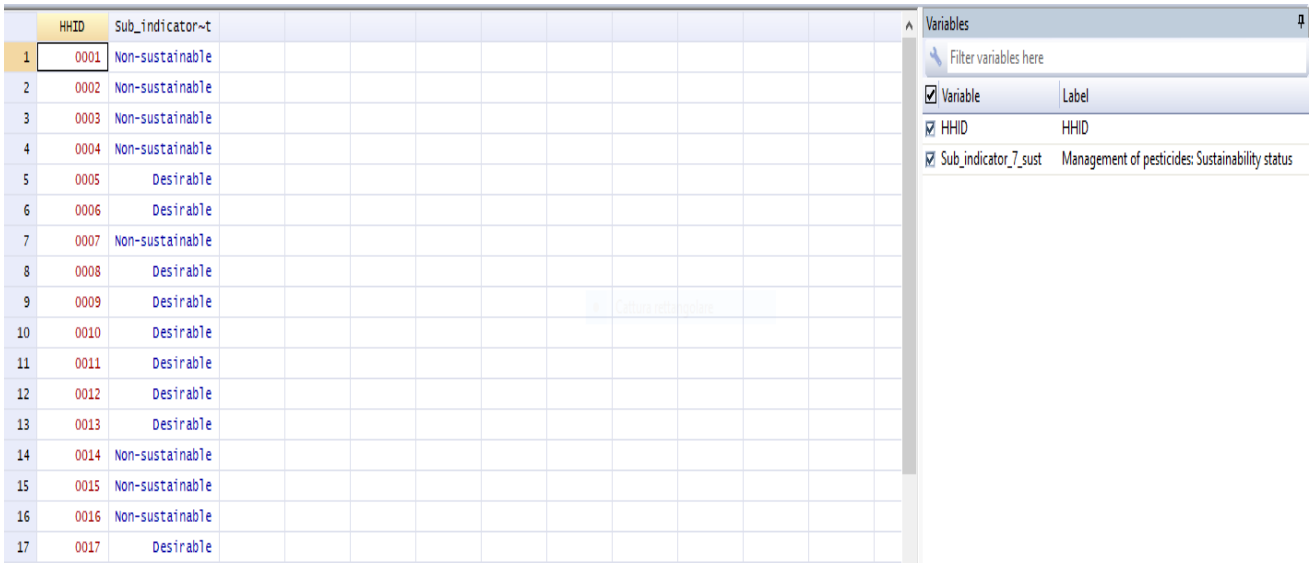
Each sub-indicator do.file was developed in order to account for the sustainability criteria. They finally generate a holding level dataset containing the following information

1. holding unique identification number
2. agricultural area of the holding
3. sustainability status associated to a given sub-indicator

Figure below provides an example of the generated dataset for sub-indicator 1 and 8.



	HHID	Agricultur~a	Sub_indicator~t
1	0001	3	Acceptable
2	0002	5.665604	Acceptable
3	0003	10.52184	Desirable
4	0004	5	Non-sustainable
5	0005	4	Acceptable
6	0006	4	Desirable
7	0007	15	Desirable
8	0008	7	Acceptable
9	0009	2	Non-sustainable
10	0010	9	Non-sustainable
11	0011	1	Desirable
12	0012	7	Non-sustainable
13	0013	2	Non-sustainable
14	0014	10	Non-sustainable
15	0015	12	Non-sustainable



	HHID	Sub_indicator~t
1	0001	Non-sustainable
2	0002	Non-sustainable
3	0003	Non-sustainable
4	0004	Non-sustainable
5	0005	Desirable
6	0006	Desirable
7	0007	Non-sustainable
8	0008	Desirable
9	0009	Desirable
10	0010	Desirable
11	0011	Desirable
12	0012	Desirable
13	0013	Desirable
14	0014	Non-sustainable
15	0015	Non-sustainable
16	0016	Non-sustainable
17	0017	Desirable

The two datasets generated from the do files "2_Data_analysis_SDG_2_4_1_sub_ind_1" and "9_Data_analysis_SDG_2_4_1_sub_ind_8" inform about the sustainability status of the holding related to sub-indicator 1 (Farm output value per hectare) and sub-indicator 7 (Management of pesticides). Depending on whether or not the sustainability criteria, as endorsed in the methodological note of SDG 2.4.1., are met by the holdings, the latter are classified as 1) desirable, 2) acceptable and 3) non-sustainable.

Data analysis: important note sub-indicators number 9

Sub-indicator number 9 "Wage rate in agriculture" defines the sustainability status of the holding taking as a reference the wage rate paid to unskilled agricultural worker rate in relation to national or agriculture sector minimum wage rate. In case there is no national or agriculture sector minimum wage rate, the national poverty line is used instead:

- Green (desirable): If the wage rate paid to unskilled labour is above the minimum national wage rate or minimum agricultural sector wage rate (if available). Default result for farms not hiring labour.
- Yellow (acceptable): if the wage rate paid to unskilled labour is equals to the minimum national wage rate or minimum agricultural sector wage rate (if available).
- Red (unsustainable): if the wage rate paid to unskilled labour is below the minimum national wage rate or minimum agricultural sector wage rate (if available).

This implies that the threshold for sustainability is likely to be different from country to country and even from year to year within the same country of reference.

The do file “10_Data_analysis_SDG_2_4_1_sub_ind_9” was developed in order to derive the sustainability status of the holding as per criteria listed above. Since the national or agricultural sector minimum wage rate differs from country to country, data analysis need to slightly amend the do file in order to account for the country-specific national or agricultural sector minimum wage rate. This modification can be done easily by simply changing the threshold for sustainability, as represented in figure below.

```
use "$RAWDATAD/Data_entry_form_Section_C_final.dta", clear

* Step 1: calculate the daily wage paid to unskilled agricultural workers
**********************************
*Daily wage
**********************************


egen  Daily_wage= rowtotal(C_C04011 C_C04021), missing                    /// sum up daily wage in-cash and in-kind

sum Daily_wage, det
**********************
*Final indicator #  9*
**********************
 The national minum wage (per day) is set at 15 LCU

    g Sub_indicator_9_sust=.
    replace Sub_indicator_9_sust=1 if C_C03000==2                          // No workers employed
    replace Sub_indicator_9_sust=1 if Sub_indicator_9_sust==. & (Daily_wage>15)                          // employed workers have a wage above the
    replace Sub_indicator_9_sust=1 if Sub_indicator_9_sust==. & ( C_C01000==1) & (C_C020A1!=. & C_C010B1!=" ")  // holding has a fair labour certificati
    replace Sub_indicator_9_sust=2 if Sub_indicator_9_sust==. & (Daily_wage==15) // employed workers have a wage above the national miunimum wage
    replace Sub_indicator_9_sust=3 if Sub_indicator_9_sust==. & (Daily_wage<15)  // employed workers have a wage above the national miunimum wage
```

Data analysts must change the threshold for sustainability by replacing the one set in the do. File (15 local currency unit) with the country-specific national or agricultural sector minimum wage rate.

Data analysis: obtaining the final dashboard

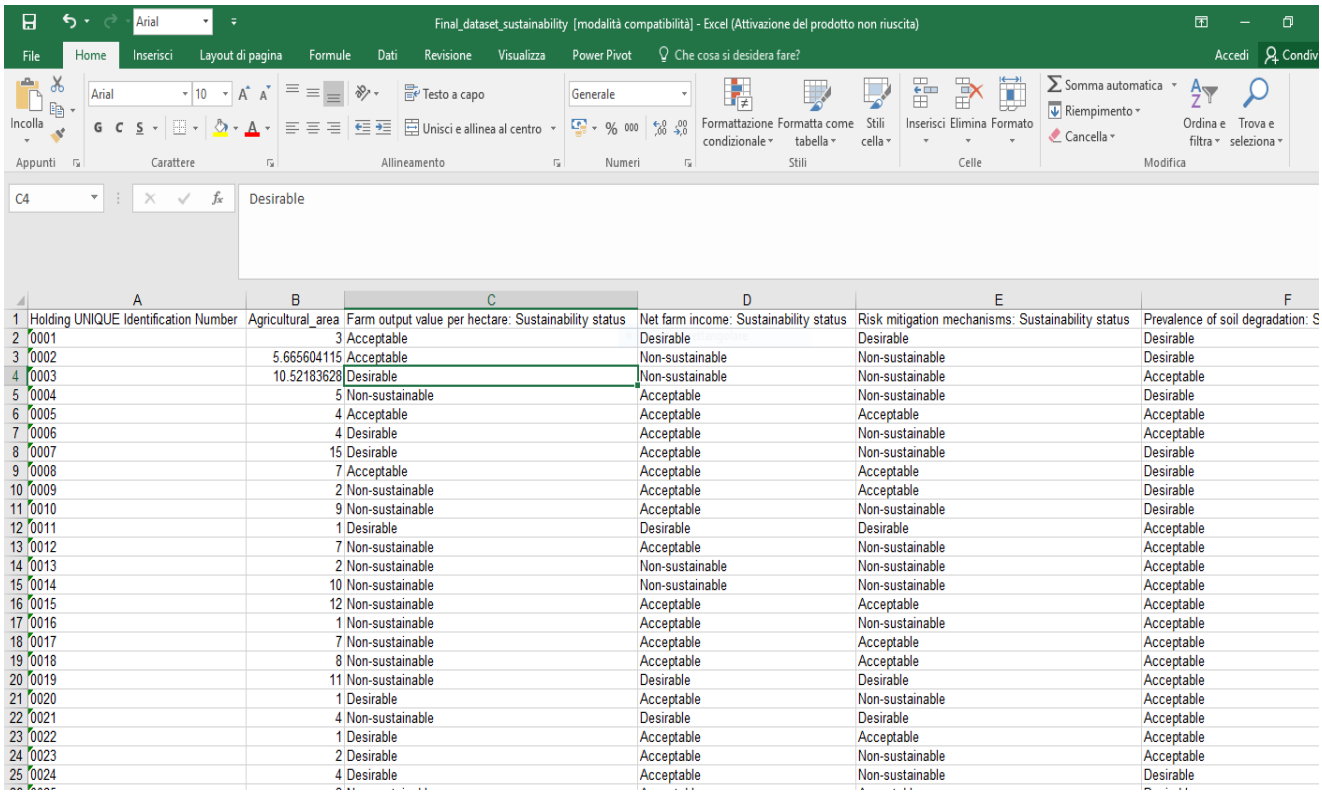
To obtain the proportion of agriculture area that is sustainable, the assessment of sustainability should be made across all sub-indicators for each holding that is part of the sample. The holding would then be assigned a sustainability level that is the most constraining across all sub-indicators, and the results would then be aggregated at the national level. This approach is automatically implemented by using the FAO and GSARS statistically approach.

Once all sub-indicators have been collected at the holding level, implying that for holding the sustainability status has been assigned for each sub-indicator, data analysts must proceed with the construction of the final dataset. The procedure to obtain a final dataset is easily implementable and can be found within the do file “13_Final_datset_dashboard_SDG_2_4_1”. The just mentioned do file produces two separates datasets: the first generated dataset contains all information about the sustainability status of each sampled holding and related to each sub-indicators, as well as the agricultural area of the holding in the second column.

The dataset will contain a total of 13 variables, as indicated below

1. Holding Unique identification number
2. Agricultural area of the holding
3. Sustainability status of the holding: sub-indicator 1 Farm output value per hectare
4. Sustainability status of the holding: sub-indicator 2 Net farm income
5. Sustainability status of the holding: sub-indicator 3 Risk mitigation mechanisms
6. Sustainability status of the holding: sub-indicator 4 Prevalence of soil degradation
7. Sustainability status of the holding: sub-indicator 5 Variation in water availability
8. Sustainability status of the holding: sub-indicator 6 Management of fertilizers
9. Sustainability status of the holding: sub-indicator 7 Management of pesticides
10. Sustainability status of the holding: sub-indicator 8 Use of biodiversity-supportive practices
11. Sustainability status of the holding: sub-indicator 9 Wage rate in agriculture
12. Sustainability status of the holding: sub-indicator 10 Food insecurity experience scale (FIES)
13. Sustainability status of the holding: sub-indicator 11 Secure tenure rights to land

The dataset will be automatically exported in excel format and will look like the one represented in figure below.



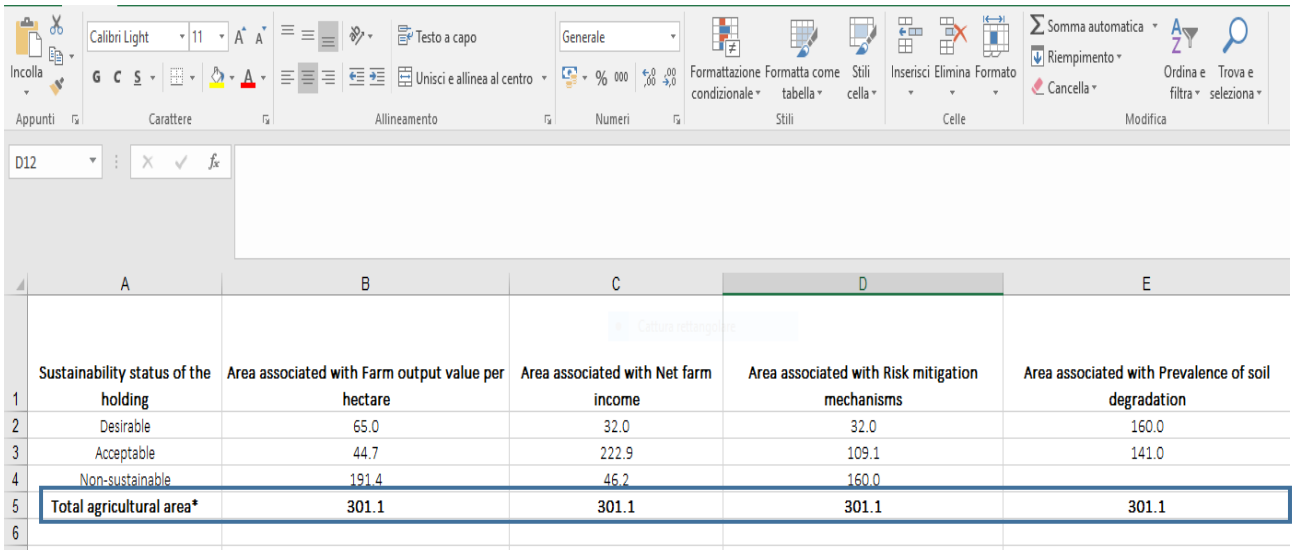
Final_dataset_sustainability [modalità compatibilità] - Excel (Attivazione del prodotto non riuscita)

C4: Desirable

	A	B	C	D	E	F
1	Holding UNIQUE Identification Number	Agricultural_area	Farm output value per hectare: Sustainability status	Net farm income: Sustainability status	Risk mitigation mechanisms: Sustainability status	Prevalence of soil degradation: S
2	0001	3	Acceptable	Desirable	Desirable	Desirable
3	0002	5.665604115	Acceptable	Non-sustainable	Non-sustainable	Desirable
4	0003	10.52183628	Desirable	Non-sustainable	Non-sustainable	Acceptable
5	0004	5	Non-sustainable	Acceptable	Non-sustainable	Desirable
6	0005	4	Acceptable	Acceptable	Acceptable	Acceptable
7	0006	4	Desirable	Acceptable	Non-sustainable	Acceptable
8	0007	15	Desirable	Acceptable	Non-sustainable	Desirable
9	0008	7	Acceptable	Acceptable	Acceptable	Desirable
10	0009	2	Non-sustainable	Acceptable	Acceptable	Desirable
11	0010	9	Non-sustainable	Acceptable	Non-sustainable	Desirable
12	0011	1	Desirable	Desirable	Desirable	Acceptable
13	0012	7	Non-sustainable	Acceptable	Non-sustainable	Acceptable
14	0013	2	Non-sustainable	Non-sustainable	Non-sustainable	Acceptable
15	0014	10	Non-sustainable	Non-sustainable	Non-sustainable	Acceptable
16	0015	12	Non-sustainable	Acceptable	Acceptable	Acceptable
17	0016	1	Non-sustainable	Acceptable	Non-sustainable	Acceptable
18	0017	7	Non-sustainable	Acceptable	Acceptable	Acceptable
19	0018	8	Non-sustainable	Acceptable	Acceptable	Acceptable
20	0019	11	Non-sustainable	Desirable	Desirable	Acceptable
21	0020	1	Desirable	Acceptable	Non-sustainable	Acceptable
22	0021	4	Non-sustainable	Desirable	Desirable	Acceptable
23	0022	1	Desirable	Acceptable	Acceptable	Acceptable
24	0023	2	Desirable	Acceptable	Non-sustainable	Acceptable
25	0024	4	Desirable	Acceptable	Non-sustainable	Desirable

The second dataset, instead will report, for each sub-indicator, the corresponding area aggregated at the national level which is: 1) desirable; 2) acceptable; and 3) non-sustainable.

The aggregation of the agricultural area under a given sustainability status is reported into a final dataset that will look like the one below:



	A	B	C	D	E
1	Sustainability status of the holding	Area associated with Farm output value per hectare	Area associated with Net farm income	Area associated with Risk mitigation mechanisms	Area associated with Prevalence of soil degradation
2	Desirable	65.0	32.0	32.0	160.0
3	Acceptable	44.7	222.9	109.1	141.0
4	Non-sustainable	191.4	46.2	160.0	
5	**Total agricultural area***	**301.1**	**301.1**	**301.1**	**301.1**
6					

It is important to double check that the sum of the agricultural areas, as expressed in hectares, is identical across all 11 sub-indicators, while it changes depending on the sustainability status by sub-indicator type.

The final dashboard can be easily obtained by calculating, for each sub-indicator, the proportion of total agricultural area which is desirable, acceptable and non-sustainable, over total national agricultural area.

Computation of results and construction of the dashboard is performed for each sub-indicator separately: for each sub-indicator, aggregation at national level is done by summing the agricultural land area of all agricultural holdings by sustainability category (desirable, acceptable and non-sustainable), and reported as percentage of the total agricultural land area of the country.

	A	B	C	D	E	F	G
1	Sustainability status of the holding	Area associated with Farm output value per hectare	Area associated with Net farm income	Area associated with Risk mitigation mechanisms	Area associated with Prevalence of soil	Area associated with Variation in water availability	Area associated with Management of fertilizers
2	Desirable	22%	11%	11%	53%	76%	51%
3	Acceptable	15%	74%	36%	47%	15%	3%
4	Non-sustainable	64%	15%	53%	0%	9%	46%
5							
6							