Training and Outreach Workshop on Agricultural Surveys

FAO HQs, Rome 24-25 October 2016

MASTER SAMPLING FRAME (MSF) for Agricultural Statistics in practice

From Census to MSF

Naman Keita
Senior Consultant
Global Strategy
Outline

1. General introduction on MSF for Agricultural Statistics
   - Introduction
   - What is a MSF for agricultural Statistics?
   - Why is a MSF needed?
   - What approaches and strategies for building a MSF?

2. Using List Frame based on Population Census and Agricultural Census to build a MSF

3. Main issues in using LF to build MSF

4. Maintaining and updating MSF based on LF
Introduction

• The Global Strategy’s foundational document, indicates that the implementation of the Second Pillar (integration of agriculture into the national statistical system)
  – “begins with the development of a master sampling frame for agriculture that will be the foundation for all data collection based on sample surveys”

• Little guidance is currently available on building an MSF for agricultural surveys

=>The GS Handbook on MSF for Agricultural Statistics (GS, 2015) aims at filling this gap
What is a Master Sampling Frame (MSF) for agricultural statistics?

Sampling frame:
• a “set of source materials from which the sample is selected” (UN, 2005) and
• basis for identifying all statistical units to be enumerated in a statistical collection.

Master Sampling Frame (MSF): a frame that enables the selection of samples either for multiple surveys, each with different content (as opposed to building an ad-hoc sampling frame for each survey), or for use in different rounds of a continuing or periodic survey.
What is a Master Sampling Frame (MSF) for agricultural statistics?

For the agricultural sector and in the context of the Global Strategy, the MSF:

- a **listing of sampling units** that, when associated with reporting units, provides complete coverage of the populations of interest,
- allows **linking of the agricultural holding to the household, its socio-economic characteristics and land dimensions**.
**Why a MSF is needed and what issues it will address?**

- **Better coherence and data integration in NSS.**
  - Avoid duplication of efforts, ensure better coherence and reduce discrepancies in data from various surveys,
  - Provide a stable reference system for agricultural surveys over time,
  - Connect various aspects of the sector and allow the analysis of sampling units from different viewpoints resulting in a better understanding of the sector.

- **Cost effectiveness.** The costs of building the MSF and selecting units will be shared by all the surveys using the master sample and use of modern technologies and various sources can reduce cost (Remote Sensing, GIS, administrative sources..)

- **Better planning and coordination.** The MSF also facilitates the planning and coordination of regular surveys in an integrated survey program. *It provides an effective tool for implementation of SPARS and foundation for AGRIS.*
What approaches and Strategies for building a MSF for Ag.Stat?

The Global Strategy presents the following strategies to build a MSF, depending on country capacity and circumstances:

(i) Using list frames (based on Population Census and/or Agricultural Census and/or Business Registers of Farms, etc)

(ii) Using an area sampling frame (based on Remote Sensing, Aerial Photos, Cartographic maps etc..)

(iii) Multiple frame combining LF and AF.
Various types of area and list frames for agricultural surveys

<table>
<thead>
<tr>
<th>Frame type</th>
<th>Frame description</th>
<th>Unit component</th>
<th>Unit type example</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>List frame</td>
<td>Holding</td>
<td>Holder addresses</td>
</tr>
<tr>
<td>2</td>
<td>List frame</td>
<td>Cluster</td>
<td>Villages or Enumeration Areas</td>
</tr>
<tr>
<td>3</td>
<td>Area frame</td>
<td>Segment</td>
<td>Holding area</td>
</tr>
<tr>
<td>4</td>
<td>Area frame</td>
<td>Map grid (cluster)</td>
<td>Point</td>
</tr>
<tr>
<td>5</td>
<td>Area frame</td>
<td>Land area (cluster)</td>
<td>Physical boundaries</td>
</tr>
<tr>
<td>6</td>
<td>Area frame</td>
<td>Point</td>
<td>Area around the point</td>
</tr>
</tbody>
</table>
Types of frames used by Countries in different regions

- **AP: Asia-Pacific**: 25% List Frame, 44% Area Frame, 23% Multiple Frames (24 countries)
- **LAC: Latin America and Caribbean**: 75% List Frame, 11% Area Frame, 12% Multiple Frames (9 countries)
- **SSA: Sub-Saharan Africa**: 64% List Frame, 21% Area Frame, 15% Multiple Frames (33 countries)
- **Total**: 65% List Frame, 12% Area Frame, 23% Multiple Frames (66 countries)

Ref: AP: Asia-Pacific; LAC: Latin America and Caribbean; SSA: Sub-Saharan Africa.

Source: Global Strategy Handbook on MSF (country assessments results)
USING LIST FRAME (LF) BASED ON POPULATION CENSUS DATA TO BUILD MSF

- There will be separate presentations on the use of *Area Frame* for building MSF and also on the use of *Business Registers of farms* for MSF. The rest of this presentation will discuss the use of LF based on Population or Agricultural Census.

- For LF, the ultimate sampling unit is the farm or farm household, therefore, identification of farm households is a critical step.

- In many developing countries, information from population censuses remains the primary source for building frames for household-based surveys since it provides a complete enumeration of all households and population in a country.

- In the process of the Population and Housing Census:
  - Country is divided into Enumeration Areas (EA) with complete enumeration of all households
  - GPS devices are increasingly used to create EA maps and to geo-reference the Households
USING LIST FRAME BASED ON POPULATION CENSUS DATA TO BUILD MSF

Two types of population censuses can be considered:

• *Traditional Population and housing census (PHC)*: use labour related data items to identify Agricultural Households:
  
  (i) **economic activity status** (status in employment: employer, employee, self employed/own account worker etc..)

  (ii) **main occupation** (type of work done in a job by a person) and **industry of main occupation** (production or activity of the establishment related to agriculture according to ISIC)

*There are severe limitations for the use of such limited items (FAO/UNFPA, 2011) with a high risk of serious under coverage. To be considered only when it is not possible to include dedicated agricultural items in the PHC.*
Population and Housing Census with an agricultural module

- When relevant agricultural questions are included in the PHC process (agricultural module),
  - A complete list of all farm households located in rural and urban areas can be identified.
  - Relevant auxiliary agricultural data can be collected for all farm households (with coordinates if included in the PHC questionnaire).
  - The agricultural module can be limited to the following two basic items or include more detailed questions compatible with the PHC (usually questions with responses yes or no):
    • Basic Item 1: Own account agricultural production
    • Basic Item 2: Measure of farm size.

- Many countries use multistage cluster sampling in which EAS are used as PSUs

- The MSF for the first stage is the list of EAs (PHC EAs may need to be adjusted) with the number of agricultural households and additional auxiliary agricultural variables. The EAs are now geo-referenced and digitized in many countries.
CONCEPTS OF HOUSEHOLDS AND HOLDINGS

Example of Activities of households

Growing maize  Growing cassava  Rearing goats

These are the agricultural activities of the holding(s)

Fishing (not aquaculture)  Brick making

These activities are not considered when determining the agricultural holding.
USING LIST FRAME BASED ON AGRICULTURAL CENSUS TO BUILD MSF

• The process of building an MSF based on an agricultural census applies when the census is conducted as a complete enumeration.

• This case is very similar to the PHC with an agricultural module.

• In addition:
  – More relevant auxiliary agric. information can be obtained and used
  – Data from agricultural censuses can significantly improve sample design for subsequent surveys.
  – Better account can be taken of rare items or geographically concentrated activities.
  – Master samples can be developed which can be used for the selection of subsamples for other surveys.
  – Agricultural Census data can also be used as benchmark for forthcoming surveys.
MAIN ISSUES IN USING LIST FRAMES TO BUILD MSF

Lists frames from an agricultural census (or a PHC with an adequate agricultural module) provide excellent auxiliary information for sampling design purposes:

- stratification and/or probability proportional to size sampling are facilitated.
- ratio and regression estimators can be used because there is enough information about variables at the population level to proportionate or to make regressions between sampling observations and population values.

The problem with both censuses is that the data become out of date because of the long time span between collection periods.

In some cases, it may take several years before the census data become available making them out of date rapidly.
MAIN ISSUES IN USING LIST FRAMES TO BUILD MSF

Some of the main issues (discussed in detail in GS Handbook on MSF) in using LF to build MSF include:

A) ASSOCIATION BETWEEN FRAME UNITS AND TARGET POPULATION ELEMENTS, including conceptual issues of correspondence between Households and Holdings, multiplicity etc. (clear rules needed). The effects in terms of inference need to be taken into account when initial frame units are divided or merged.

C) DEALING WITH IMPERFECTIONS OF LIST FRAMES (over-coverage, under-coverage)

D) NON-SAMPLING ERRORS IN LIST FRAMES (duplicate names, inclusion of elements from other populations, missing population elements)
MAINTAINING AND UPDATING LIST FRAMES

- Updating the sampling frame from PHC or Ag Census could be considered at different sampling stages when multistage sampling is used.

- In such cases there are two frames: a frame of land areas constituted by the enumeration areas (area frame) and a list frame of holdings or households inside the selected PSU.

- If the sample is going to be used in several surveys, the updating of the list frame must only refer to the elements in the selected PSU.
As far as the updating of the area frame of PSU (EAs) is concerned, it usually presents no major problem because of their stability.

Its updating could be facilitated by the growing use of geo-referenced and digitised EA maps as part of the PHC process:
- A database of all EAs in the country will be available with agricultural related data collected during the PHC.
- This information can be combined with satellite images (with land cover and use information) to build an area frame that is much easier to update.

Updating of sampling frames for the second stage can be done by using rotating sample selection of PSUs and performing a complete enumeration of selected PSUs.
MAINTAINING AND UPDATING LIST FRAMES

• An ideal process for the Master Sample Frame development as described in United Nations (1986) would be:

  (1) develop the frame needed for the agricultural census (from PHC for example),

  (2) enhance the frame with census outputs (auxiliary information), and

  (3) structure these materials in a form suitable for the anticipated sample selection operation.

• As explained in United Nations (1986): “To follow this ideal sequence, one must start with a reasonably clear picture of the purposes for which MSF will be used. Once these requirements are established, the steps necessary to meet them can be incorporated in the census plan. This process can also work in the opposite direction. If a Master Sampling Frame is developed from a census and is updated appropriately during the subsequent 5- or 10-year period, the job of developing the frame for the next census is likely to be considerably easier.”
Mozambique conducted its Population and Housing Census in 2007. It included an agricultural module (Section G) in the population census household questionnaire. The questions were as follows:

- **G 1**: Does any member of the household practice agriculture for himself? Yes/No
- **G 2**: Does the household have any tank for aquaculture? Yes/No If yes, how many? -----
- **G 3**: Does any member of the household practice fishing through traditional methods? Yes/No
- **G 4**: Does this household have cashew trees? Yes/No If yes, how many? -----
- **G 5**: Does this household have coconut trees? Yes/No If yes, how many? -----
- **G 6**: How many animals does this household have?
  - **G 6.1** Cows/Bullocks -----
  - **G 6.2** Goats ------
  - **G 6.3** Sheep ------
  - **G 6.4** Pigs ------
  - **G 6.5** Chicken ------
  - **G 6.6** Ducks ------

These questions allowed identification of Agricultural Holders and provided additional auxiliary data on:

- Types of permanent crops on the holding.
- Number of animals on the holding for each livestock type.
- Presence of aquaculture on the holding.

The data collected was used to build an effective sampling frame for the (sample based) agricultural census conducted in 2010 and subsequent annual agricultural surveys.
ETHIOPIA: Use of list frame based on population census to build a MSF.

- Ethiopia has extensive experience collecting data through an annual agricultural survey based on a sample of PSU’s (PHC/EAs) selected with PPS from which households are listed and selected (SSU).

- A listing questionnaire was used at the beginning of the population census to collect data to be used for developing the master frame.

- The sampled PSU’s are also used for household income and consumption surveys.
ETHIOPIA: Use of list frame based on population census to build a MSF.

Some lessons learned:

• The list frame approach facilitates integration between different household surveys.
• The list frame reduces the cost and allows data to be linked for in-depth analysis.
• The list frame is good for collecting socio economic data, and provides a well distributed sample.
• Data collection can be time consuming as households are distributed across the PSU.
• Some holdings can be missed and not all parcels associated with a holding can be identified.
• It is difficult to update the master frame based on LF because of boundary changes in administrative areas like district.
For More information

• Global Strategy to improve Agricultural and Rural Statistics. 2010. UNSD, World Bank, FAO.

• **Guidelines for Linking Population and Housing Censuses with Agricultural Censuses**—with selected country practices, *FAO/UNFPA, 2012*


See: [http://www.gsars.org](http://www.gsars.org)
Thank You