



Global Strategy  
IMPROVING AG-STATISTICS



# INTEGRATED SURVEY FRAMEWORK FOR AGRICULTURAL AND RURAL STATISTICS



PIERO DEMETRIO FALORSI, **FAO**





---

# Outline

---

## 1.Goals of the Research Activity

## 2.Concept of Integration

## 3.Strategic ideas for the Research Activity

## 4.Research Lines

- ✓ Examples of the complexity
- ✓ Literature Review and identified gaps
- ✓ Interactions among sub-lines of activity

## 5.First Results of Lines A

## 6.Partners and Organization of Research Team

## 7.Conclusions





# 1. Goals of the Research Activity

**Production of integrated agricultural statistics on three different target populations:**

- ✓ **agricultural plots (environment)**
- ✓ **households (social)**
- ✓ **farms (economic).**

**This specific research program focuses on the general issues of :**

- ✓ **Building the integration of data-sets which share same common statistical units.**
- ✓ **Observing units of different populations in an integrated manner.**
- ✓ **Producing integrated estimates for the phenomena of the target populations.**

## 2. Concept of Integration

The concept of integration is not well defined in literature and it may mean different things in different contexts.

As starting point, is useful to consider the main reasons for integration.

✓ **Observational reasons:** e.g. the sampling frame is only available for one of the target populations.

✓ **Survey's goals:** in most of current statistical systems data are collected independently by sector. This division leaves no opportunity to measure the impact of an action in one sector on another.

✓ **Efficiency reasons:** phenomena related to different units may have correlation structure that could be taken into account through specific models.

### 3. Strategic idea for the Research Activity

Integration is a concept related to the whole **chain of the statistical process**(Linkage, sample design and estimate).

**Key success factor** for a good integration strategy is taking into account, through an holistic approach, the various dimensions that identify the country specific informative context.

**Software tools** for capacity building

**Dimensions to be considered:**

- ✓ **Country variables (e.g. Geographical area, statistical maturity)**
- ✓ **Features that define the types of databases and their integration (e.g. role of the Census, periodicity of the different data sources for the integration)**

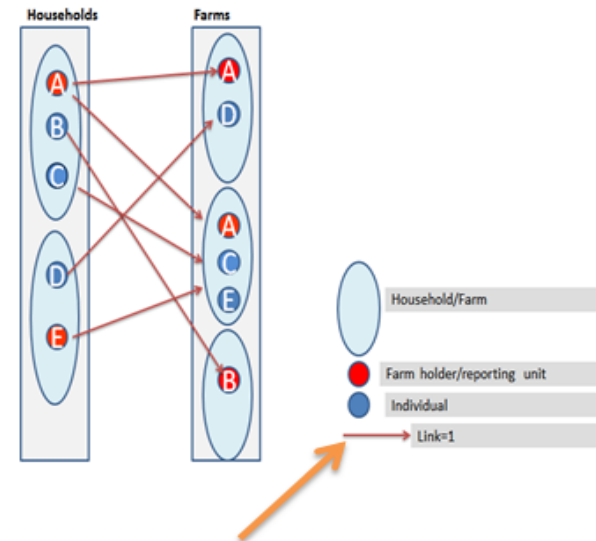
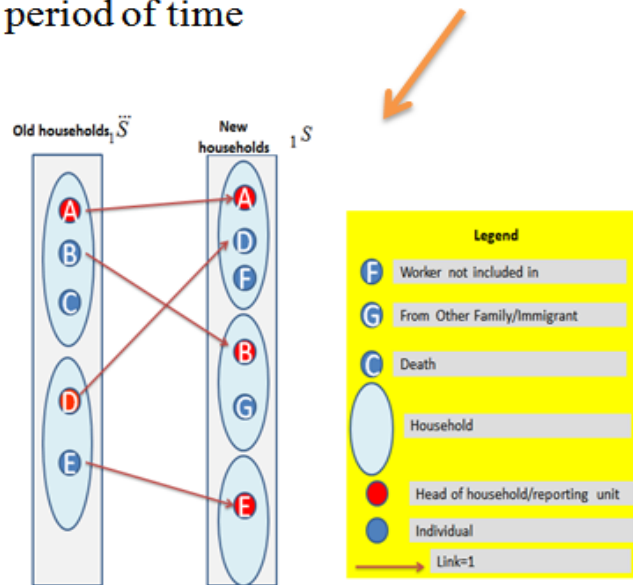
**Note that the availability of the Master Sampling Frame is a useful but not mandatory condition for achieving the Integration**



## 4. Research lines

### Examples of the complexity of Observational Contexts (1)

**A-** In some circumstances the units of the same population are observed in different period of time



**B-** When the units of two different population are jointly observed, it is necessary to consider the complex links among these.

The same variable can be observed either from the target population or from another population taking properly into account the links



## 4. Research lines

### Examples of the complexity of Observational Contexts (2)

**Informative context 1:** Countries in which there is neither the census of the population nor the census of the agriculture and only some surveys (with the related frame) are available.

**Informative context 2:** Countries in which there is only the census of the population with some surveys on agriculture.

**Informative context 3:** Countries in which both the census of the population and the census of the agriculture are available.

**Informative context 4:** Countries in which the different target populations can be covered by a multiplicity of frames

In each defined informative context it is also possible to make a distinction between:

**Countries with integrated statistical systems (CISS),** i.e. countries in which the census and/or the surveys are based on a single frame;

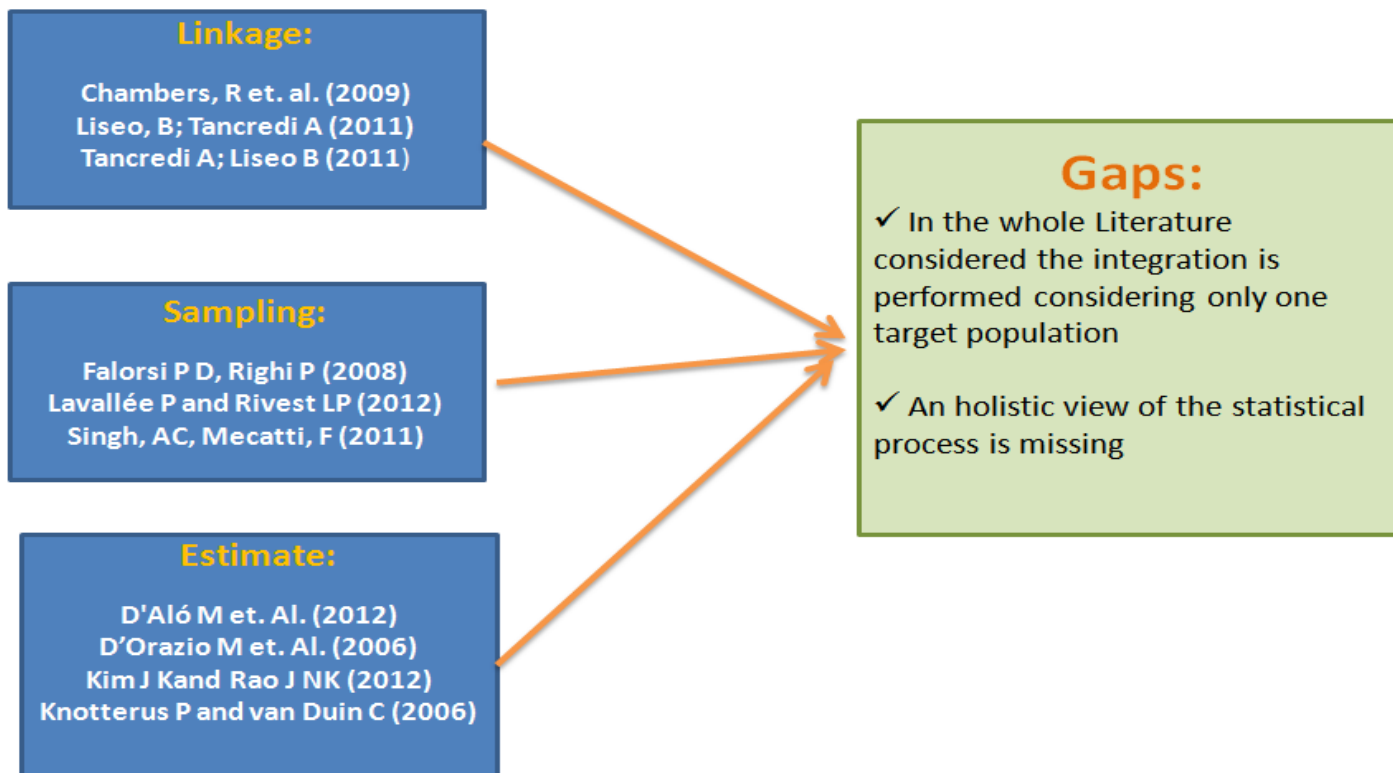
**Countries with non-integrated statistical systems (CNISS),** i.e. countries in which the census and/or the surveys are based on independent frames.

Informative context	Integration of the statistical system	
	A : CISS	B : CNISS
1 : Only some surveys	Ex. Guyana	Ex. Angola, Sierra Leone, Yemen
2: Only one out of the two censuses	Ex. Papua New Guinea, Nepal, Tanzania	Ex. Israel, Liberia, Mauritius
3: Agricultural and population census	Ex. Canada, Mozambique,	Ex. Italy
4: Multiple frames		

## 4. Research lines

### Literature review and identified gaps

The entire literature review and gaps may be found in the Research Plan





## 4. Research lines

### Interactions among sub-lines of activity (1)

#### Two Research lines

**Line A: Guidelines** for consolidation of current survey practices (**Short time result**).

This line starts from current advancements in survey sampling theory and is aiming at developing guidelines which, using indirect sampling techniques as unified theoretical framework, proposes feasible survey strategies which assures at the same times:

- ✓ **Integrated observation**
- ✓ **Coherent estimates**

**Line B: Advanced topics** for quality, design and estimation

This line of research is supposed to find innovative solutions for integration considering the whole statistical process:

- ✓ **Linkage**
- ✓ **Sampling**
- ✓ **Estimation**

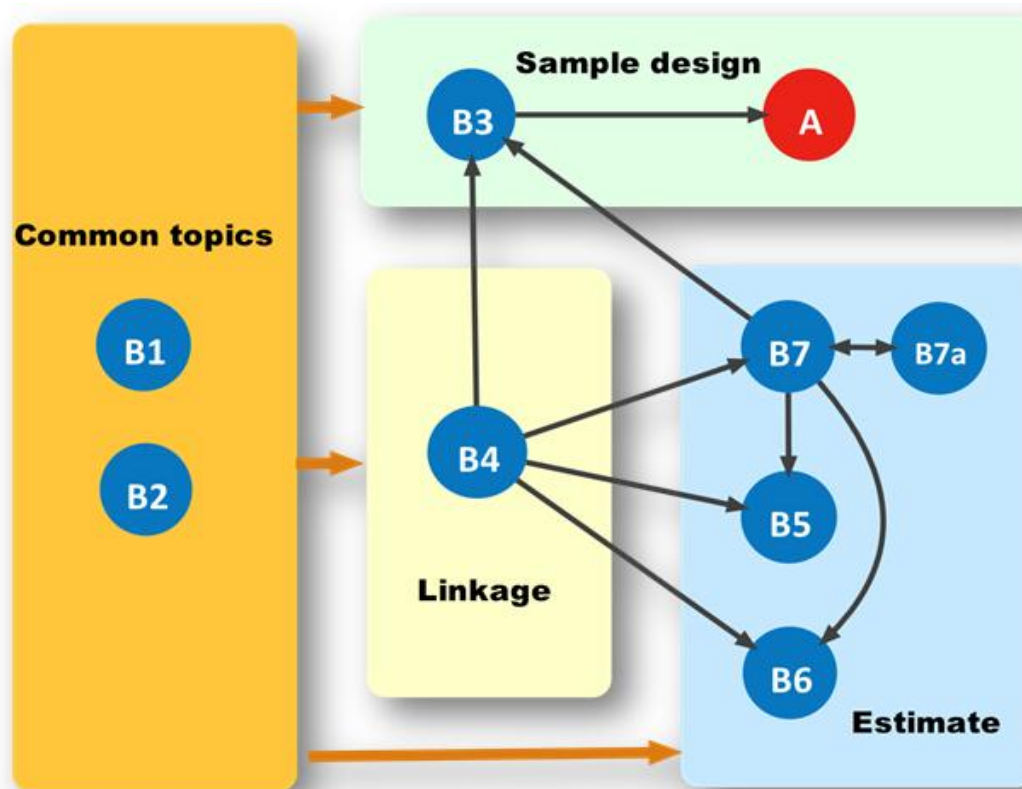
The output of this line are especially useful for achieving reasonable solutions for the less structured informative contexts.



## 4. Research lines

### Interactions among sub-lines of activity (2)

Most of the research sub-lines focus on specific statistical methods for achieving the integration. Conversely there are other sub-lines that cover cross-cutting issues.



Relationships between research lines



## 5. First results of line A: Summary

The first results of line A are summarized in the “*Concept note on the use of Indirect Sampling as unified framework for Integrated Agricultural Statistics*” .

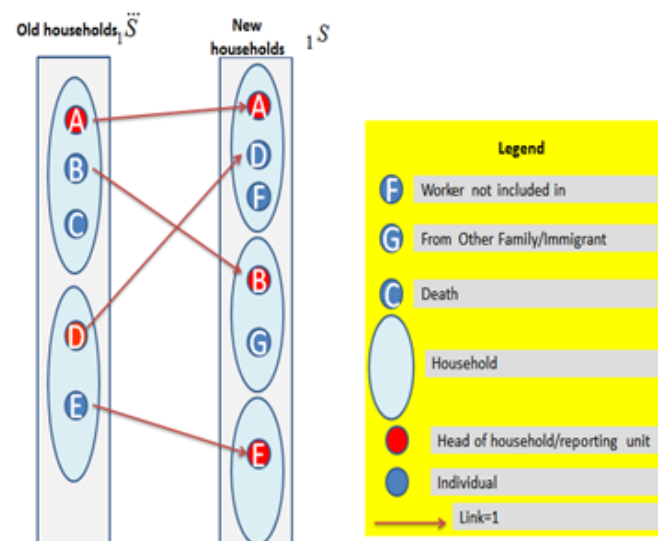
- ✓ We have shown as **indirect sampling** may represent a unified approach for assuring the consistency of integrated agricultural statistics
- ✓ and for dealing with **frame imperfection**.
- ✓ The methodological solution here described covers as particular cases different methods proposed in literature for dealing with:
  - **imperfect frames**
  - **rare populations** (as the snowball sampling or adaptive sampling (Chaudhuri, 2010)).
  - **Different estimators for area frame**
- ✓ The approach here proposed is **quite flexible** and may be adopted for the particular country case.

## 5. First results of line A

### Example 1: starting from the households

- 1) Selecting a sample of Enumeration Areas (EA).
- 2) Making a census of all existing households in the sample EAs.
- 3) All the farms having people of  ${}_1S$  as workers (either as employees or holders) are surveyed. In this way an indirect sample of farm  ${}_2S$  is observed
- 4) All the land of  ${}_2S$  is surveyed

$${}_1A \xrightarrow[\text{dir}]{EA} {}_1S \xrightarrow[\text{indir}]{\text{person}} {}_2S \rightarrow {}_3S$$



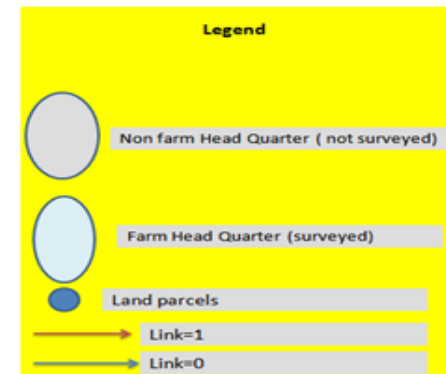
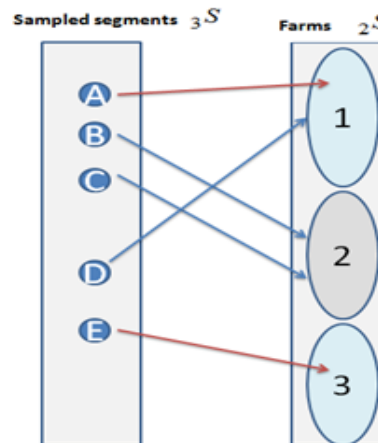
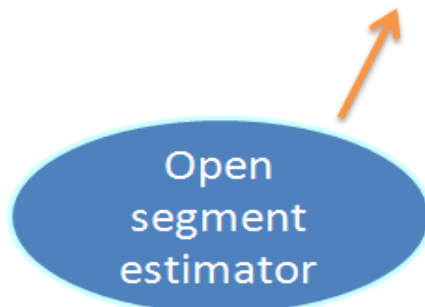


## 5. First results of line A

### Example 2: starting from the land

- 1) Selecting a direct sampling of land segments.
- 2) Making a census of all existing tracts in the sample segments, thus obtaining the sample  ${}_3S$ .
- 3) Creating an indirect sample of a farms,  ${}_2S$ , by considering only the farms which have their headquarter located in the area of the selected sample of land segments. Creating an indirect sample of a farms,  ${}_2S$ , by considering only the farms which have their headquarter located in the area of the selected sample of land segments.
- 4) Forming a sample of households by considering all the households of the farm workers in  ${}_2S$ .

${}_3A \xrightarrow[\text{dir}]{\text{Segments}} {}_3S \xrightarrow[\text{indir}]{\text{farmheadquarte}} {}_2S \xrightarrow[\text{indir}]{\text{woke}} {}_1S$





## 5. First results of line A

### Estimation: starting from the households

$${}_j\hat{Y} = \sum_{j,k \in_j S} w_{jk} {}_jY_{jk}$$

$${}_j\hat{Y} = \sum_{1,k \in_1 S} \sum_{i \in_1 k} d_{1ki} z_{1ki}$$



- ✓ The sampling weights are those defined in the sampling on the households.
- ✓ **UNIQUE FILE FOR ALL THE VARIABLES**
  - The variables values are those original if the variable is related to the population of households.
  - The variables values are the transformed ones, if the variable of interest is related to the populations of farms and land parcels. These variables take into account the links existing with the households.
- ✓ The calibrated form assures the **coherence with the known total of auxiliary variables for all population values.**
- ✓ The estimator assures the coherence even a variable is collected from the units of a population different from the one the population is related

## 6. Partners and Organization of the research activity

### Researchers from:

- ✓ University of Rome, La Sapienza
- ✓ Italian Institute of Statistics
- ✓ INE Mozambique
- ✓ NBS Tanzania

**Strong management of research activity (Monthly meetings and three external meeting with discussion of research outputs)**

*Timetable of the research activities and expected outputs*

RESEARCH LINE	TIMETABLE OF EXPECTED OUTPUTS BY MONTHS								
	2013						2014		
	6	7	8	9	10	11	12	1	2
<b>Line A</b> Guidelines on existing practices for observing units of different populations in integrated framework							First version	Revised version	End
<b>Line B</b> Advanced topics for quality, design and estimation				1		2	3	4	End

1: Report on ongoing or already completed research activities, on the review of the relevant literature, on gaps analysis and remaining methodological issue

2: Output Empirical studies designed and tested on FAO data sets and Open software identification/development.

3: Technical report on the findings and recommendations for possible solutions to methodological issues prepared, peer reviewed, and validated by experts

4: Report on Methodological basis for the guidelines for a advanced technical assistance and training.

## 7. Conclusions

### Strengths

- ✓ **Holistic approach** aiming at covering the whole chain of the statistical production process
- ✓ **Flexible strategies** that enable to achieve the integration in different informative contexts (also in those less structured)
- ✓ **Quick wins.** First results of line A available in a relative short time for guidelines
- ✓ **Results yet available.** Accepted papers in several scientific meetings (ISI, ICAS, International Symposium on imperfect frames –Statistics Canada)
- ✓ The entire set of methods and software tools produced will be tested on **real data**

### Risks

- ✓ Complex governance/cooperation
- ✓ Time Limited

} **Need of a strong management**







---

**Thank you**

**[Piero.Falorsi@fao.org](mailto:Piero.Falorsi@fao.org)**

---

