Statistical confidentiality of agricultural surveys in the context of the AGRISurvey program
Thijs Benschop and Clara Aida Khalil (FAO)
Thijs.Benschop@fao.org, ClaraAida.Khalil@fao.org

Abstract and Paper

The Agricultural Integrated Survey (AGRIS) developed by the Food and Agricultural Organisation of the United Nations (FAO) is a farm-based modular 10-year survey model designed as a cost-effective tool to improve the production of quality data on the technical, economic, environmental and social dimensions of agriculture. In 2016 the FAO Statistics Division began supporting countries to implement the AGRIS model at national level through the AGRISurvey program. Besides enhancing data collection, one of the core objectives of the AGRISurvey program is to support partner countries in the dissemination of the collected agricultural microdata. The practice of and, hence, the experience with dissemination of microdata from agricultural surveys is limited. Reasons are the specific characteristics of farm survey data, such as the inclusion of (large) commercial farms and multi-level nested modules (household, parcels, crops). This contribution presents the particular challenges with respect to disseminating microdata in the AGRISurvey project as well as the developed solutions to anonymizing and disseminating microdata as public use files or scientific use files. These are illustrated by the practice in case of the first program countries Senegal and Uganda.
Statistical confidentiality of agricultural surveys in the context of the AGRISurvey program

Thijs Benschop*, Clara Aida Khalil**

* Food and Agricultural Organization of the United Nations, Viale delle Terme di Caracalla, 00153 Rome, Italy, thijs.benschop@fao.org

** Food and Agricultural Organization of the United Nations, Viale delle Terme di Caracalla, 00153 Rome, Italy, claraaida.khalil@fao.org

Abstract. The Agricultural Integrated Survey model (AGRIS) developed by the Food and Agricultural Organisation of the United Nations (FAO) is a farm-based modular 10-year survey model designed as a cost-effective tool to improve the production of quality data on the technical, economic, environmental and social dimensions of agriculture. In 2016, the FAO Statistics Division began supporting countries to implement the AGRIS model at national level through the AGRISurvey program. Besides enhancing data collection, one of the core objectives of the AGRISurvey program is to support partner countries in the dissemination of the collected agricultural microdata. The practice of and, hence, the experience with dissemination of microdata from agricultural surveys is limited. Reasons are the specific characteristics of farm survey data, such as the presence of (large) commercial farms and multi-level nested modules (household, parcels, crops). This contribution presents AGRISurvey and the particular challenges with respect to disseminating microdata in the AGRISurvey project as well as the approaches used to anonymizing and disseminating microdata as public use files or scientific use files. These are illustrated by the practice in the first program countries Senegal and Uganda.

1 Introduction

The Agricultural Integrated Survey (AGRIS) developed by the Food and Agricultural Organisation of the United Nations (FAO) is a farm-based mod-
ular 10-year survey model designed as a cost-effective tool to improve the production of quality data on the technical, economic, environmental and social dimensions of agriculture (see [2]). In 2016, the FAO Statistics Division began supporting countries to implement the AGRIS model at national level through the AGRISurvey program. This program, financed by USAID and the Bill and Melinda Gates Foundation, aims to support agricultural data collection, processing and dissemination efforts of National Statistical Offices (NSOs), Ministries of Agriculture (MoA) and line Ministries. The initiative responds to the demand for more and higher quality agricultural data, in view of fostering evidence-based decision-making at national level, and countries’ ability to monitor the Sustainable Development Goal indicators, particularly those related to food and agriculture. An integral part of the AGRISurvey program is the Open Agricultural Statistics (OAS) component.

The OAS component of the AGRISurvey program aims at opening agricultural data in general, with special focus on the dissemination of agricultural census and survey non-confidential microdata. Under the OAS component, an assessment of the data dissemination policy and program (i.e., protocols and practices) in the context of agriculture statistics is performed and customized support to improve these are provided in view of making agriculture statistics, including agricultural survey microdata, accessible to the public in formats that ensure greater readability, usability and findability of the data.

When disseminating microdata from agricultural surveys or censuses, new challenges are encountered due to the specific nature of agricultural data. These are different from those encountered when disseminating microdata from standard household surveys. The main issues are the inclusion of large (non-household or commercial) agricultural producers, the hierarchy between the different modules as well as the potential sensitivity of the data. The AGRIS model consists of a core module and several rotating modules. This modular design poses further challenges to the anonymization.
2 The AGRIS model

The Agricultural Integrated Survey (AGRIS) model is a farm-based modular 10-year survey model designed as a cost-effective tool to improve the production of quality microdata on multiple dimensions of agriculture. Considering challenges linked with survey nonresponse, missing data and poor data quality connected with long survey questionnaires, the AGRIS model advises partner countries to adopt a modular approach for agricultural surveys. A core module on agricultural production administered on an annual basis, is complemented by a set of rotating modules, implemented at different frequencies, covering all the major dimensions of agriculture.

These rotating modules intend to bring additional knowledge to the following thematic domains:

- The economic dimension of agricultural holdings, including the income of family holdings;
- The labor input used in agricultural holdings, allowing to measure indicators on agricultural labor productivity;
- The production methods adopted by agricultural holdings, allowing to assess, among other factors, the environmental impact of agriculture and the efficiency of the adopted production techniques;
- The machinery and equipment used and assets available to agricultural holdings, allowing to assess the stage of modernization of the agricultural production process.

The AGRIS model also provides recommendations on sampling strategies, which are issued at two levels. First, commercial entities and the household sector are considered separately, which allows designing a two-stage stratified plan for sampling households and a one-stage stratified plan for commercial holdings. Second, a partial sample rotation over time is advised, in order to limit survey burden and allow for the analysis of longitudinal panel data.

Data generated by AGRIS is meant to inform policy design and implementation, strengthen the efficiency of agricultural markets and support the research community. Finally, the AGRIS model acknowledges the link with
the Sustainable Development Goals (SDG) monitoring agenda, as the proposed set of AGRIS Generic Questionnaires will generate the data necessary to calculate several SDG indicators.

3 Anonymization of agricultural surveys

3.1 Literature

In the literature on confidentiality and anonymization of microdata, agricultural surveys or censuses are seldom dealt with separately. Depending on the population of interest, there are two options where to look in the literature:

1. if the population of interest consists only of holdings in the household sector, the characteristics resemble those of other household surveys, such as household budget surveys or multi-purpose household surveys;
2. if the population of interest includes commercial holdings, which are often larger in size, the characteristics resemble those of enterprise surveys.

In the literature, it is often pointed out that the release of microdata from enterprise surveys should be more restricted, for example by choosing a more restrictive release method such as a controlled research data center. Two reasons why the release of enterprise data needs to be more restricted are (see e.g. [3]):

1. the difficulty to anonymise large businesses, such as the only utility company;
2. the (commercial) interest to try to reidentify records, e.g., by competitors in the same economic sector, is much larger than in the case of household surveys.

3.2 Practice

The release of microdata from agricultural surveys and censuses as PUF or SUF is limited, partially because of the aforementioned issues. There are, however, some noteworthy exceptions from which we can learn for the AGRISurvey microdata:
• At the Italian NSO Istat, the 2005 Italian Farm Structure Survey was anonymized and released as SUF. The report mentions that records were protected by using variable suppression, variable aggregation and perturbation. [4]

• The microdata of the Living Standards Measurement Study (LSMS) run by the World Bank are released as anonymized PUF. The LSMS data includes data on agricultural holdings. However, the statistical units in the LSMS are mainly households. There is no public documentation on the anonymization of the LSMS data.

3.3 Anonymization of AGRISurvey

In the context of the implementation of AGRISurvey in low income and low-middle income countries, the only realistic release types are Scientific Use Files (SUF) and Public Use Files (PUF), as the technical and financial capacity of the data producer (mostly the NSO or Ministry of Agriculture) does not allow for the use of other release types, such as remote access. One alternative access type, on-site access, is considered to be too restrictive in terms of data access to be a viable option for the donors.

As the countries where AGRISurvey is implemented are mainly low income and low-middle income countries, the disclosure scenarios are typically different from these in higher income countries. The availability of external data and the capacity and interest of intruders is much lower than in other countries. Therefore, the need for anonymization is generally lower.

Three characteristics of the AGRIS model pose an additional challenge to the anonymization of the microdata:

• the AGRIS model proposes the use of a partially rotating sampling, which mean that the data has the structure of panel data. This increases the risk of reidentification, as the information can be obtained from datasets from previous survey rounds.

• in many agricultural surveys, the location of plots and parcels is registered by using GPS coordinates. GPS coordinates can be interpreted as a direct identifier. The easiest way to deal with geo-referenced data
is to remove these variables and instead leave in or create other administrative geographical variables, such as province or region. However, this approach also removes the benefits of geospatial data. Another option is the geographical displacement of areas and/or records, which needs to be further investigated.

- the AGRIS model proposes rotating modules. Rotating modules means that for the first five years the anonymization approach would have to be different because we are collecting always new data. Furthermore, after five years (when implementing a module for the second time) it cannot be expected that the same exact questionnaire is used as the country context will have evolved and, as a consequence, the data that need to be collected.

4 Country practice

AGRISurvey is gradually rolled out in low income and low-middle income countries. In two countries, Senegal and Uganda, the OAS component is currently being implemented. This section describes the experiences and issues encountered in both countries in respect of microdata dissemination and confidentiality.

4.1 Senegal

Since 2017, FAO has been formally engaged with the "Direction de l’Analyse, de la Prévision et des Statistiques" (DAPSA) of the Ministry of Agriculture and Rural Equipment (MAER) toward implementation of the AGRISurvey program in Senegal. This includes the OAS component with associated activities. The *Enquête Agricole Annuelle* is the annual agricultural survey in Senegal.

The EAA estimates the production of the most important rain-fed crops produced by agricultural households, using the crop-cutting method. It also provides information on the physical characteristics of cultivated plots (geo location, area) and major investments made (agricultural inputs, cultural operations, soil management and restoration). Structural data, such as agricultural equipment, agricultural income, agricultural risks and adaptation
strategies, are also collected once every three years. The EAA covers the entire country and generates representative production estimates for 14 regions and 42 agricultural departments of Senegal. [1]

The first round of the EAA within the AGRISurvey program was the EAA 2017/18. The EAA questionnaire is relatively small and the data is collected during two field visits: post-planting and post-harvest. As it was the first round of the EAA under the AGRISurvey program, the quality of the data collected can still be improved by reformulating questions and responses in the questionnaire and in the data collecting process. After careful consideration, none of the collected variables was identified as sensitive or confidential, which reduced the need for anonymization. It was decided to release the dataset only as SUF at this stage, as the main users are researchers. In terms of anonymization, direct identifiers were removed (including the GPS coordinates) and any geographical variable below the département was removed. The sample is representative at the département level. Furthermore, recoding and local suppression were used to achieve 2-anonymity. By using PRAM, uncertainty was added to variables that are easily observable, such as the ownership of a tractor.

4.2 Uganda

Since 2018, FAO has been formally engaged with the Uganda Bureau of Statistics (UBOS) toward implementation of the AGRISurvey program in Uganda including the Open Agriculture Statistics (OAS) component with associated activities. The Annual Agricultural Survey (AAS) was first rolled out in 2017 and in 2018 it has been revised and improved according to the AGRIS methodology. Data on agricultural production, labour input, cost of production and women participation to agriculture, food insecurity of agricultural households and transfers received is collected. Households are visited four times per year. In Uganda there are two agricultural seasons and in each season there is a post-planting and post-harvesting visit. The sample is representative at the zardi level.

Due to delays in the data collection and processing, the microdata has not been released as to date. Only an initial assessment of possible disclosure scenario was undertaken. The NSO is currently revising the data dissemination policy which includes also a policy for microdata release. UBOS, the
Ugandan NSO, has no previous experience and practice in anonymizing and disseminating microdata.

5 Conclusion

More countries will be included in the near future, with the ambitious objective of reaching 50 countries by 2030 in the framework of a greater umbrella project, the Global Rural and Agricultural Integrated Surveys (GRAInS), joining the work of the AGRISurvey and the Living Standard Measurement Study (LSMS). While working with these countries, new challenges, general or country-specific, will appear in the context of disseminating microdata. The most pressing question is likely how to deal with large-scale commercial farms.

References


