Quality of Agricultural Censuses:

How to Assess and Improve Quality Leveraging New Technologies

Naman Keita

Retired FAO Principal Officer Bamako, Mali- Naman.Keita21@gmail.com

Seydou Doumbia

INSTAT, Bamako, Mali – Seydou.doumbia79@gmail.com

Abstract

Implementing a good quality Agricultural Census in a context of growing demand for more complex and inter-linked data raises particular challenges that can be addressed to a large extent by intensive use of new technologies and appropriate methodologies. Assessing census data quality is an important step towards its improvement. However, benchmarks for comparing and rating agriculture census quality are rare and not always standardized
This paper will discuss the challenges and opportunities for assessing and improving agricultural census data quality using new technologies and possible benchmarks for rating census data quality.

Keywords: Agricultural Census, Data quality, New technologies, Benchmarks.

1. Section 1: Introduction

Agricultural censuses are one of the main sources of agricultural data for evidence-based policy, program planning, and monitoring of food security and rural development. Therefore ensuring high data quality is essential to derive reliable statistics on farm structure, production capacity, labor use, and resource management. However, implementing a good quality Agricultural Census in the current context of growing demand for more complex and inter-linked data raises multiple challenges at all phases from design to implementation, processing and dissemination. The emergence of a wide range of new technologies and methodological innovation provide opportunities to enhance data quality.

The following sections will clarify the concept of data quality and its assessment in the context of an Agricultural Census.

2. Section 2 : Defining and assessing Agricultural Census data quality

The concept of quality of statistics as defined in (FAO 2018) retain the emphasis on user needs and satisfaction (fitness to use). Therefore while accuracy (mainly concerned with minimizing sampling and non-sampling errors) remain an important dimension, there are other important dimensions to take into account when considering the quality of statistical data.

The number of dimensions of statistical data quality vary across the institutions. FAO (2014) recommends considering the following five dimensions:

Relevance: the degree to which statistics (e.g. census items collected) meet users' needs, implying the need to avoid production of irrelevant data, namely, data for which no use will be found

Accuracy and reliability: closeness of the estimated value to the (unknown) true value

Timeliness and punctuality in disseminating results: time elapsed between release of data and reference period and degree to which preannounced release dates are met

Coherence and comparability of statistics: degree to which data from a single statistical programme, and data brought together across statistical programmes, are logically connected and the degree to which statistics are comparable over space (between countries and regions) and time (between different time periods).

Accessibility and clarity: Accessibility is defined as the ease, the set of conditions and the modalities by which users can obtain data, while clarity refers to the availability of adequate documentation: whether data are accompanied with appropriate metadata, illustrations such as graphs and maps, whether information on their quality are also available (including limitation in use), and the extent to which additional assistance is provided.

3. Section 3: Improving Agricultural Census data quality

a. Traditional methods

Before technological innovations, common approaches to assessing census quality included:

Pretest and Pilot Surveys: Testing questionnaires and procedures on a small sample to identify issues.

Post-Enumeration Surveys (PES): Independent re-interviews used to estimate coverage errors.

Edit and Imputation Reports: Analyses of missing values, inconsistent responses, and correction procedures.

Enumerator Monitoring: Supervision and spot-checking to evaluate data collection practices.

Consistency Checks: Cross-tabulations and coherence analysis within and across variables.

While these methods remain valuable, they are often resource-intensive and limited in scope without technological support

b. Leveraging new technologies for assessing agricultural census data quality

Emerging technologies are transforming how agricultural censuses are conducted and evaluated.

The new World Programme for Census of Agriculture 2030 released by FAO recognize that there is a agrowing use of information technology in data collection, processing and dissemination. The new FAO publication emphasizes the use of technology in census operations as this enable higher accuracy, lower costs, and real-time quality assurance which enhances the value of census data for agricultural policy, research and business, beyond the traditional statistical uses. Some of the new technologies include:

Geographic Information Systems (GIS) & Remote Sensing

- **Satellite imagery** and **drone data** can validate land use, cropping patterns, and the existence of holdings.
- High-resolution imagery supports **sample frame updating**, reducing under-coverage.
- Integration of remote sensing with GIS allows automatic detection of discrepancies between reported and observed land use.

Mobile Data Collection Platforms

- **Computer-Assisted Personal Interviewing (CAPI)** tools on tablets or smartphones reduce transcription errors and can enforce validation rules during data entry.
- Real-time GPS coordinates and timestamps help monitor enumerator coverage and detect potential fraud.
- Conditional logic minimizes respondent burden by skipping irrelevant questions automatically.

Cloud Computing and Centralized Data Management

- Cloud platforms allow real-time data upload, central dashboards, and automated quality checks.
- Supervisors can monitor progress remotely and address issues immediately, reducing delays.

Machine Learning and Automated Analytics

- **Predictive models** can flag inconsistent or outlier responses for targeted verification.
- Natural language processing (NLP) can analyze open-ended responses more efficiently than manual coding.
- Cluster analysis and pattern recognition assist in identifying systematic errors.

Digital Administrative Records and Integrated Data Sources

- Linking census data with **tax records**, **land registries**, and **market data** enhances completeness and reduces respondent burden.
- Integrated systems support longitudinal analysis and cross-validation.

c. Challenges and limitations

Despite clear advantages, several obstacles exist:

- High initial costs for technology acquisition and training.
- Dependence on internet and electricity infrastructure.
- Risk of algorithmic bias if ML models are trained on incomplete data.
- Technological solutions cannot fully replace sound statistical design.

4. Section 4 : Possible benchmarks for rating census data quality

Quality benchmarks for agricultural censuses may cover several core dimensions:

Coverage Benchmarks

- *Undercoverage and overcoverage rates* (levels of missing or duplicate holdings).
- Post-enumeration survey (PES) comparisons to estimate how complete the census is relative to a frame or independent benchmark.

Accuracy Benchmarks

- *Error rates* from editing and validation checks (e.g., inconsistencies, non-response proportions).
- *Sampling error measures* (standard errors or coefficients of variation) when sampling is used.

Timeliness and Punctuality

- Whether census data are released according to planned schedules.

- Time lags between collection, processing, and dissemination.

Comparability Benchmarks

- Alignment with **international definitions and classifications** (e.g., FAOSTAT concepts, Indicative Crop Classification) for comparability across countries and time periods.

Documentation & Metadata Standards

- Presence of **standardized metadata** explaining content, methods, quality indicators, and limitations (e.g., documentation of survey instruments, coding schemes).
- Use of international metadata standards to facilitate reuse and interoperability.

These benchmarks are aligned with widely accepted data quality frameworks used in official statistics (e.g., by UNECE, Eurostat, Statistics Canada, and other national offices).

A possible Agricultural Census Quality Benchmarking Framework (ACQBF) may be structures around the following four dimensions:

1. Institutional Readiness
    - Legal Framework
    - Budget Adequacy
    - Staff Training Coverage
    - Digital Infrastructure
2. Methodological Soundness
    - Frame Coverage Error
    - Overcoverage Rate
    - Adoption of International Definitions
    - Pilot Testing Conducted
3. Operational Implementation
    - Response Rate
    - Item Nonresponse
    - Real-Time Validation Rules

- GPS Verification Coverage
- Enumerator Error Rate

4. Data Output & Dissemination
    - Timeliness
    - Revision Rate
    - Metadata Completeness
    - Microdata Accessibility

For each dimension and sub-element, scores and weight could be defined in order to compute an overall composite index for the Census.

For example:

- **Statistics Canada's Census of Agriculture quality indicators** assign letter grades (A–F) based on **coefficients of variation** (CV) to reflect reliability of estimates. "A" indicates excellent quality while "F" is too unreliable to publish.
- Response rates and coverage measures (percentage of complete returns and under coverage estimates) are widely used as simple operational benchmarks of census execution quality.

Such schemes — where quantitative measures (CVs, response rates, coverage percentages) are used to classify results — serve as **benchmarks for internal quality reporting and cross-topic comparison** within a given census.

5. Conclusions

There are established quality frameworks for agricultural censuses — particularly in the **World Programme for the Census of Agriculture** .However, to truly compare and *rate* census quality internationally, these benchmarks must be standardized further by:

- agreeing on **common performance indicators and thresholds**,
- adopting **interoperable metadata standards**, and
- building **international reporting platforms**.

This harmonization would make it easier to assess quality consistently across countries and census cycles, to learn from best practices, and to support data-driven agricultural policies.

References

- FAO. *World Programme for the Census of Agriculture* 2030
- FAO (2018) WORLD PROGRAMME FOR THE CENSUS OF AGRICULTURE 2020 VOLUME 2 **Operational guidelines**
- **Eurostat**. 2015. *Quality Assurance Framework of the European Statistical System Version 1.2.*
- Eurostat. 2013. Quality Assurance Framework. Version 1.1.
- **FAO**. 2015. *World Programme for the Census of Agriculture 2020* Volume 1*: Programme, concepts and definitions*. Rome.
- **FAO**. 2014. *Statistics Quality Assurance Framework*. FAO. Rome.
- **FAO**. 2012. *World Census of Agriculture 2000 Methodological Review*. FAO. Rome.
- **Statistics Canada**. 2017. *Quality Assurance Framework Third Edition*.
- **United Nations (UN)**. 2018. In: *2020 World Population and Housing Census Programme* [online]. New York. [Cited 30 January 2018].
- **United Nations (UN)**. 2017. Principle and Recommendations for Population and Housing Censuses. Rev.3. New York.
- **United Nations (UN)**. 2012. *Guidelines for the template for a generic national quality assurance framework (NQAF)*.
- **United Nations Economic Commission for Europe (UNECE)**. 2006. *Recommendations for the 2010 Censuses of Population and Housing*. Conference of European Statisticians New York and Geneva.
- **Zarkovich, S.S.** 1966. *Quality of statistical data*. FAO. 395 pp.