Harnessing data revolution for cost-effective agricultural censuses in the 21st century with widely accessible and used results

Naman Keita

Senior Agricultural Statistics Consultant FAO/World Bank

Viale Terme di Caracalla, CP 64259

Rome, Italy

Naman.Keita21@gmail.com

DOI: 10.1481/icasVII.2016.f37d

ABSTRACT

Data revolution is defined as *'an explosion in the volume of data, the speed with which data are produced, the number of producers of data, the dissemination of data, and the range of things on which there is data, coming from new technologies such as mobile phones and the 'internet of things' and from other sources such as qualitative data, citizen-generated data and perceptions data'*. The other associated feature is '*a growing demand for data from all parts of society*'[1].

This emerging development will have a growing impact on all major statistical data production activities in coming years, including the Census of Agriculture. The Data Revolution raises challenges but also offers unprecedented opportunities to modernize the process of conducting a Census of Agriculture. In this 21st Century, statisticians have access to a variety of methods and tools that was not available before for conducting a census of agriculture. If used appropriately, they can generate significant efficiency savings and provide timelier and better quality data which can be widely accessed and used by a variety of users.

In a context of growing demand for more diversified and frequent data and at the same time the scarcity of resources for census taking, the new tools and methods can be used to address the new challenges for ensuring that the census is conducted in the most relevant and cost-effective way. This requires taking a fresh look at the concept of census of agriculture, the overall strategy and

[1] The UN Secretary-General's Independent Expert Advisory Group on a Data Revolution for Sustainable Development (IEAG) report *A World That Counts: Mobilising The Data Revolution for Sustainable Development*.

modalities of designing and conducting a census of agriculture in order to use innovative options for more relevant and efficient census operation.

The availability of digital and mobile computing tools for data capture using smartphones or tablets, geo positioning tools such as handheld GPS devices and more precise and cheaper remote sensing images can provide cost-effective alternatives to traditional way of collecting, centralizing and processing data depending on country situation. Also the increasing use of other sources of data such as administrative sources in a growing number of countries can complement (or even partially replace field data collection for a number of items) and reduce census burden and cost.

The FAO World Programme for Census of Agriculture 2020 in its forthcoming Volume 2 on Operational Guidelines on Implementing the Census of Agriculture will provide specific guidance on various modalities for conducting the Census in the next decade as well as the use of new tools and methods.

This paper will discuss the challenges and opportunities of the Data Revolution and its potential impact for designing and implementing a cost effective Census in the 21st Century in a context of growing demand for more complex and inter-linked data particularly for monitoring the newly adopted Sustainable Development Goals by the United Nations. The paper will refer to selected country experiences of agricultural censuses using new and innovative methods that take advantage of the on-going data revolution.

Keywords:

Data revolution, Census of Agriculture, Cost effectiveness, open data

Introduction

Data revolution is defined as 'an explosion in the volume of data, the speed with which data are produced, the number of producers of data, the dissemination of data, and the range of things on which there is data, coming from new technologies such as mobile phones and the 'internet of things' and from other sources such as qualitative data, citizen-generated data and perceptions data'. The other associated feature is 'a growing demand for data from all parts of society'[2].

However, the (UN report, 2014) indicates that 'People, economies and societies are adjusting to a world of faster, more networked and more comprehensive data – and all the fears and dangers, as well as opportunities, that it brings' . To remain relevant and cost effective, the implementation of an Agricultural Census in coming decades must take into account this new data and technological environment. Some of the dimensions and features of the data revolution relevant for census of agriculture include: (i) The availability of multiple data sources and their potential use in agricultural census, (ii) the use of technological innovation for data collection and faster processing (iii) the availability of tools and methods for wider dissemination, faster, easier and safer access to census data (aggregated and microdata) and metadata. The sections below will discuss some aspects of these three features as they relate to the census of agriculture.

[2] The UN Secretary-General's Independent Expert Advisory Group on a Data Revolution for Sustainable Development (IEAG) report *A World That Counts: Mobilising The Data Revolution for Sustainable Development*.

1. Availability of multiple data sources and census modalities

One of the main features of the ongoing data revolution is the increasing number of data producers and data sources in general. This new data environment has an impact on data production strategies to better respond to the challenge of limited resources for statistical activities and growing demand for more quantity and diversified data to be produced more frequently. This challenging data requirement implies the use of multiple data sources and data collection modalities.

The new 2030 Sustainable Development Agenda (SDA) adopted by the UN in September 2015 defines 17 Sustainable Development Goals (SDGs) and 169 individual targets, including 19 multidimensional agricultural-related SDG targets and 25 global indicators for monitoring the progress towards the goals and targets. Many of these indicators require a diversified data sources for their compilation beyond traditional censuses and surveys. Administrative records for example will be an important source of some data requirement.

The changing environment and data requirements have led some countries to conduct a detailed review of their agriculture statistics programme, including alternative options for conducting the agricultural census. For example, a report by Statistics Canada on agriculture statistics program review (StatCan, 2012) includes a detailed evaluation of several options for conducting the census of agriculture within the agriculture statistics program. The evaluation uses criteria based on the six main dimensions of data quality[3] and other criteria such as respondent burden, cost, operations[4] and acceptability by user community and respondent community. The study concludes that the census of agriculture remains relevant and necessary but, regardless of the option chosen, changes are needed in order to render the program more efficient. The changes proposed include increasing the use of administrative data to gradually replace survey data and increasing utilization of remote sensing applications with a view to replacing survey items or entire surveys over the longer term.

The guidelines for the new World Programme for Census of Agriculture 2020 prepared by FAO also takes into account the new data and technology environment and proposes various modalities and data sources for conducting the census of agriculture in the next decade. The guidelines recommend that the census of agriculture may be carried out using different approaches and in various ways, depending on available resources and national conditions. The following four modalities for conducting a census of agriculture are recommended:

(i) *classical approach* in which a census is conducted as a single one-off operation in which all the census information is recorded;

(ii) *modular approach with a* core module by complete enumeration and supplementary module(s) on sample basis, using information collected in the core module as the frame for the supplementary module(s) and conducted shortly after the core module ;

(iii) *integrated census and survey modality* with a census core module (which could be lighter than in the modular approach) to be carried out on a complete enumeration basis and rotating thematic modules to be conducted annually or periodically on sample basis over a longer period between two censuses (usually ten years) through a permanent agricultural survey programme such as the Agricultural Integrated Survey programme (AGRIS); and

(iv) *combined census with* use of registers and administrative records as a source of part of the census data.

It is considered that regardless of the census modality and way of conducting a census, the crucial principle is to achieve the main objectives[5] of a census of agriculture depending on country

[3] Relevance, accuracy, coherence, timeliness, interpretability, and accessibility

[4] Defined as ability to react quickly to new demand, ability to conduct large occasional surveys, timing and compliance with StatCan Business architecture.

[5] To provide **data on the structure of agriculture, especially for small administrative units**, to provide **data to use as**

statistical capacity, national preferences and the availability of resources and data sources. The increasing use of multiple sources in census programmes is one of the changes that will be adopted by many countries in coming agricultural censuses.

2. Use of technological innovation for more relevant and cost effective census

Another important feature of the data revolution is the explosion of technological innovations. The following tools and emerging technologies are some of the innovations that are directly relevant to the design and conduct of agricultural census. They have the potential to allow faster, better quality data collection and processing during an agricultural census at reduced cost. The upcoming FAO guidelines and many recently published guidelines under the Global Strategy initiative and other sources provide practical guidance on the use of these tools and technology. This section discusses some of the main tools and technology relevant to the conduct of agricultural census.

Remote sensing (RS)

Remote Sensing refers to 'images acquired with a conventional camera or electronic sensors from aircraft or satellites. The techniques applied to process and interpret remote sensing imagery include visual photo-interpretation and a wide range of numeric algorithms' (Global Strategy, MSF, 2015).

According to (Delince, J, 2015), there are five main areas of use of Remote Sensing (RS) and aerial photos for agricultural statistics:

- Monitoring Land cover/Land use
- Frame construction
- Support to field work of censuses/surveys
- Crop area estimation
- Crop yield forecasting/monitoring**.**

The most relevant uses of RS and aerial photos for the census of agriculture are cartography and frame building, supporting field work and crop area estimation.

<u>Census frame construction and updating</u>*:* Many countries use enumeration area (EA) maps derived from the cartography work of the population and housing census (PHC) as basis for building the frame for the agricultural census. One of the main products of the PHC cartography and mapping work is a set of EA maps which are large scale maps delineating accurately the area of work of enumerators and covering the entire country. Also, a listing operation is usually conducted during the cartography of PHC. Recent and high resolution satellite images or aerial photos can provide enough details (population density, infrastructure etc..) to support the preparation and updating of EA maps. They facilitate observance of the usual enumeration area properties such as completeness, non-overlap, respect of administrative divisions, accurate delineation and allocation of attributes (Delince, J. 2015). These EA maps from PHC cartography, together with all other information collected (for example number of households or households practicing agriculture) can be used for building the frame for agricultural census based on list frame when this operation is conducted not too far from the PHC.

Satellite images can be also used to identify and subdivide the agricultural land (stratify) by intensity of land use and other land cover characteristics and therefore help in the construction of area sampling frames when this type of frame is used in the agricultural census.

---

benchmarks and to provide **frames for agricultural sample surveys**.

As indicated in (Global Strategy, MSF, 2015) recent developments have seen the emergence of some companies in the information technology sector that produce public-access images with global coverage **(Google Earth, Bing)**. A major advantage is these images are available and easily accessible, with an efficient interface. Most agricultural areas of the world are covered by Very High Resolution (VHR) images which is a significant advance, especially for countries that could not access such images in the past.

Support to field work*:* Recent advances regarding Satellite images or aerial photos can help the enumerator to optimize his travel and facilitate localization of farmers and fields during the field work of the agricultural census. Aerial photos or very high-resolution imagery can help the enumerators to access the land and/or locate the holder. Used as paper prints or on a tablet, imagery can also be used to check on farmer's declaration and minimise measurement errors.

Crop area estimation**:** The two main methods used to derive crop area statistics from remote sensing (Delince J, 2015) are: Pixel counting which is the more direct way of area estimation and Calibration methods for integration of field surveys data and image classification results. However, todate, the operational use of these methods remain limited.

During the recent census of agriculture of Morocco (2015/2016), an extensive use of orto-photos was made during data collection as these photos were considered precise enough (spatial resolution of 30 cm) to identify boundaries of holdings and their plots[6]. The scale of the orto photos was 1/5000, with photos covering an area of 3 km X 4.5 km (about 1350 ha). Below is an example of orto-photo displaying the delineation of one holding composed of one parcel and plot.

Source: Ministère de l'agriculture et de la pêche du Maroc (2015)

[6] The average farm-size is 1.6 hectares in Morocco, according to (USAID Country profile for Morocco: http://www.usaidlandtenure.net/sites/default/files/country-profiles/full-reports/USAID_Land_Tenure_Morocco_Profile.pdf)

Global Positioning System (GPS)

The global positioning system (GPS) is the oldest and best known global navigation satellite systems (GNSS). A GNSS is a system based on a network of navigation satellites that is controlled by ground stations on Earth which continuously transmit radio signals – captured by receivers – to determine the receiver's geographical location (longitude, latitude, and elevation) on the Earth's surface (Global Strategy, MSF, 2015).

The expanded availability, low cost and improved accuracy of handheld GPS devices provide an opportunity to support field activities during a census of agriculture: geo-referencing plots, household or headquarters of holdings; locating sample units or measuring the area of a plots.

Geo-referencing **:** In addition to its use during the cartographic operations to geo-reference boundaries of EAs, GPS can support enumerators during data collection. For each enumeration area, headquarters of holding units can be identified using GPS coordinates on handheld GPS devices or mobile digital device equipped with GPS. When integrated in EA maps, this information can significantly facilitate the field work of the enumerators. A more extensive elaboration on the use of GPS is available in the (UNSD, 2009).

Measuring plots with GPS: GPS is very useful to measure the area of plots on the field and recent studies (Carletto et al., 2015) suggest that technological advances with moderate-priced GPS have led to significant improvements in accuracy of measurements even for small plots. GPS can also be used as a tool for quality control, for point location control in objective yield surveys when used during the census.

Computer Assisted Personal Interview (CAPI) on laptops or notebooks, tablets and Smartphones:

With increased capacity of mobile electronic devices (such as laptop computers, tablets, Smartphones, etc..), large geographical coverage of internet and mobile phone networks and availability of suitable packages, data collection activities during censuses and surveys does not have to rely only on paper questionnaires. In fact, a comprehensive study conducted by (Caeyers, B. and all, 2010) using a randomized survey experiment among 1840 households, designed to compare pen-and-paper interviewing (PAPI) to computer-assisted personal interviewing (CAPI) found 'that PAPI data contain a large number of errors, which can be avoided in CAPI'. A growing number of countries are now using CAPI in data collection activities. An overview of the use of CAPI for agricultural census is provided below[7].

Data capture: Mobile electronic devices with adequate CAPI software can be used to directly capture data in the field, using electronic questionnaires. The electronic questionnaires usually contain the same questions types like on paper questionnaires, but also some new question types not available with PAPI. Some CAPI applications allow the capture of the geo-coordinates of the statistical units, take photos, and scan and record barcodes of items which are present within the statistical unit.

Field data editing and validation: The use of CAPI facilitates the implementation of skip patterns through 'enablement conditions' which are pre-programmed instructions, activating or deactivating questions based on the answer to a previous question. CAPI also provides census and survey designers with validation conditions which are pre-programmed instructions that anticipate an answer to be in a certain range and showing a message in case the answer does not fall in the range.

[7] The main source of information used for this overview is a technical paper prepared in May 2016 by Michael Rahija, Research Officer at the Global Office of FAO Global Strategy as input to World Programme for Census of Agriculture 2020 Volume 2 (forthcoming).

Data transmission and centralization: The use of CAPI, allows a quick transmission of census data captured to a centralized database for further processing. Data is usually transmitted over cell phone and Wi-Fi networks (available in most tablets) . Some tablet devices offer a slot for a SIM card to access cell phone data networks. Finally, some CAPI applications also offer tools for offline data transmission.

Other use of CAPI include: Geo referencing statistical units using tablet with GPS and use for optimizing logistics and supporting enumerators, using GPS and paradata for monitoring survey progress.

An important benefit of CAPI is the savings in costs related to the use of paper questionnaire: paper and printing cost, logistics and storage cost, cost of computers and personnel for data entry. However, equipment and software costs for CAPI should be considered when comparing to paper questionnaires. This cost may be substantial for a comprehensive operation such as a census of agriculture. Equipments may be also re-used in subsequent surveys resulting in additional savings. Data quality should also be improved thanks to field editing and validation features. Finally, timeliness should be also enhanced with quick transmission and centralization and the absence of any further data entry work.

There may be also some limitations in the use of CAPI for census of agriculture in terms of operational constraints, particularly in developing countries, such as limited field personnel able to operate the electronic devices, poor internet and cellular phone network coverage, poor electricity availability in rural areas for recharging batteries and difficult operating conditions (sun, dust, rain poor light etc.). Other limitations reported include poor visibility of screens, slow processing when volume of data increases. A detailed study and testing is recommended before adopting CAPI in a wide scale operation such as a census of agriculture.

Iran was one of the first countries that have successfully implemented its agricultural census using CAPI on tablets; this was during its fourth Iranian Census of Agriculture carried out from 27 September to 9 November 2014 [SCI, 2014]. In order to collect data from the estimated 4.3 million holdings of Iran, 5550 enumerators were mobilized and 7500 tablets were procured. It is reported that there was substantial cost savings and considerable reduction in time lag between data collection and data availability (preliminary results were available within 6 months). Also, the tablets were to be re-used for several other future surveys (Keita, N. 2015).

Computer Assisted Telephone Interview (CATI) and Computer Assisted Web Interview (CAWI[8])

With the considerable developments in ICT during the past decades, remote data collection methods are increasingly being used in countries where the conditions are favourable, reducing considerably the cost of data collection. CATI and CAWI are two such methods. In CATI, data is collected via telephone interview (respondents are contacted by phone and data is entered into electronic questionnaires) while in CAWI, census notice is sent to respondents with instructions on how to access the web (online) questionnaires with their secure access code, phone number to call for help and how to complete it online. The web questionnaire usually includes navigational help information, drop-down menus and online edits similar to the one based on CATI questionnaires. Skip patterns are built in the web questionnaires so only questions related to the type of farm operations are presented to respondents.

[8] Also known as CASI: Computer Assisted Self Interview

Several countries such as Iceland, Spain, Canada, Poland have used these remote data collection methods in a mixed mode (combined with other data collection methods) during their census of agriculture.

3. Wider access to aggregated data, safe access to micro data and open data

The current environment is characterised by a phenomenal development in computer and software capacity, wider internet connectivity and emergence of standards, protocols, methods and tools for data archiving, safe and on-line access to macro-data and micro data. This new environment has opened a new era for potentially accessing census data both in aggregated form and micro data on an unprecedented scale.

There are now digital preservation standards that make it possible to archive and manage digital data over the long term. More detailed guidance on these standards are available from the International Household Survey Network (IHSN) [9] Principles and Good Practice for Preserving Data.

In addition to the traditional paper publications of aggregated census results, a wide variety of electronic products can be made available and easily accessed as stand alone or via internet to satisfy the needs of a wide range of users.

Macro-data aggregated in statistical tables, in graphical or cartographic forms can be made available to the general public online in databases or on the web pages of the census agencies. Aggregated data may also be disseminated through optical disks (CD-ROM or DVD-ROM) and/or on a pen drives. Data can be visualised (often interactively) using appropriate GIS packages and on cartographic supports such as atlases in printed or electronic format.

There is a growing demand for more open data with possible access to micro-data. For a long time this was not possible in many countries, due to the constraints resulting from the statistical laws (confidentiality of individual information) and lack of standards and technical tool to avoid the risks of statistical disclosure. Nowadays, guidance is available for adopting an adequate legal and policy framework which respects the confidentiality provisions in the statistical law for safe access to micro-data (see see http://www.gsars.org/providing-access-to-agriculture-microdata-a-guide/). Moreover, a lot of progress has been made in terms of technical tools for controlling the risks of statistical disclosure through anonymisation of micro-data to respect confidentiality requirements. Finally a variety of methods for formatting data files for safe access to micro data is available, including Public Use Files, Licensed use files, Remote Access Facilities, Data enclaves, Deemed Employee. Further details on these methods of access can be found at: http://www.icpsr.umich.edu/icpsrweb/content/deposit/guide/chapter5.html

4. Concluding remarks

There are features of the current data revolution that can be an opportunity for conducting cost effectively agricultural censuses in this 21st century to respond to the needs of a wide range of data users. Depending on country conditions, tailored census modality may be considered and technological innovations can facilitate timely data collection and processing. Developments in computer and software capacity, wider internet connectivity and emergence of standards, protocols, methods and tools for data archiving, online access to macro-data and safe access to micro data have created a completely new data and technological environment. This new environment has

[9] International Household Survey Network (IHSN) 2009, Working Paper No 003, Principles and Good Practice for Preserving Data

opened a new era for potentially accessing census data both in aggregated form (macro-data) and micro data on an unprecedented scale. Detailed guidelines are also available on the use of these innovations for agricultural censuses and sample surveys.

REFERENCES

- Caeyers, Bet (University of Oxford), Neil Chalmers (EDI) and Joachim De Weerdt (EDI) (2010): *A Comparison of CAPI and PAPI through a Randomized Field Experiment*. November 2010
- Delince J. (2015): Cost-Effectiveness of Remote Sensing for Agricultural Statistics in Developing and Emerging Economies. Global Strategy Technical Report Series GO-09-2015. Rome, FAO December 2015.
- Food and Agriculture Organization of the United Nations (1996): *Conducting Agricultural Censuses and Surveys* (SDS#6 Rome, 1996).
- Food and Agriculture Organization of the United Nations (2015): *World Programme for the Census of Agriculture 2020 -VOLUME 1 –Programme, Definitions and Concepts- Final draft*, FAO, Rome, 2015.
- Global Strategy to Improve Agricultural and Rural Statistics (2015) *Handbook on Master Sampling Frames for Agricultural Statistics*, by F. Vogel, N. Keita, J. Galego, C. Ferraz and M. lmes.
- José Luis Hernández Rodríguez, (2015): *Exploitation of Administrative Data from RAN and PROCAMPO to Strengthen the Geo-statistical Framework of the Agricultural Census in Mexico*. Country paper presented at an FAO Global Strategy, Expert Group Meeting on administrative data for agricultural statistics, Rome, April 2015.
- Keita N. (2015): *Relevance and cost-effectiveness of agricultural censuses in the 21st Century- Lessons learned from experiences in selected developing countries*. Invited paper presented at ISI 2015
- Ministère de l'agriculture et de la pêche du Maroc (2015): Registre National Agricole-Recensement Général de L'Agriculture : MANUEL DE LA METHODOLOGIE
- PARIS21 (2014): *KNOWING IN TIME, How technology innovations in statistical data collection can make a difference in development*, Espen Beer Prydz. PARIS21 Discussion Paper No. 2 January 2014
- Rahija M. (2015): *Technical paper on use of CAPI for agricultural census (internal)* prepared as input to World Programme for Census of Agriculture 2020 Volume 2 (forthcoming).
- Statistics Canada (2012): *Agriculture Statistics Program Review*, August, 2012.
- Statistical Center of Iran (2014): *Implementation of the Fourth Iranian National Census of Agriculture*, August 2014. https://www.amar.org.ir/Default.aspx?tabid=972&articleType=ArticleView&articleId=1632
- UN Secretary-General's Independent Expert Advisory Group on a Data Revolution for Sustainable Development (IEAG) (2014) report *A World That Counts: Mobilising The Data Revolution for Sustainable Development.* Nov 2014
- United Nations. (2016). Sustainable Development Goals: The Sustainable Development Agenda. Accessed at http://www.un.org/sustainabledevelopment/development-agenda/
- UNSD (2009), *Handbook on Geospatial Infrastructure in Support of Census Activities, United Nations, Series F No. 103.*
- World Bank and Food and Agriculture Organization of the United Nations (2011): *Global Strategy to Improve Agricultural and Rural Statistics*. Report no 56719-GLB. World Bank, Washington DC, 2011.