WORLD PROGRAMME FOR CENSUS
OF AGRICULTURE 2020
Volume 2: Operational guidelines

USE OF TECHNOLOGY FOR AGRICULTURAL CENSUS
DATA COLLECTION

2018
Introduction

This technical paper is part of a series of resource papers prepared in the framework of the development of the World Programme for Census of Agriculture 2020 Volume 2 on operational guidelines. The resource papers will be accessible on the web and linked to the main publication of volume 2 of WCA 2020.

This paper discusses the use of technology for the census of agriculture data collection and draws on currently available publications and research work on existing technological tools relevant to the Census. This includes Remote Sensing (RS) and aerial photos, handheld Global Positioning System (GPS), mobile and digital devices such as smartphones and tablets.

The sections ahead are mainly drawn from relevant Global Strategy publications on the subject (Global Strategy 2017b, 2015a, 2015b). A short summary of the paper is available in Chapter 20 of Volume 2 of WCA 2020.

The availability of digital and mobile computing tools for data capture at affordable prices, such as smartphones or tablets, geo-positioning tools like handheld GPS devices, and more precise and cheaper RS images now provide new and cost-effective alternatives to traditional methods of collecting, centralizing and processing census data that are dependent on the situation in each country. However, technology is moving so fast that there may be technologies available in the next decade that are either unknown or not affordable at the time of preparing this publication. Therefore, census managers should consider the trade-off between the safety of proven systems and the benefits of using new technology. There are existing technologies that can be used in an agricultural census to support fieldwork for the control, supervision and monitoring, or to provide the tools to create a system that integrates enumeration with control and monitoring. Following are the various technologies currently available for data collection in an agricultural census:

- RS and aerial/ortho photos
- Handheld GPS devices
- Handheld and mobile digital devices

Remote sensing and aerial photos

There are five main areas of use of RS and aerial photos for agricultural statistics (Global Strategy, 2015b):

- Monitoring land cover/Land use
- Cartography and area frame construction
- Support to field work of censuses/surveys
- Crop area estimation
- Crop yield forecasting/monitoring.

The use of RS and aerial photos for cartography and frame building for agricultural censuses was discussed in Chapter 13 of Volume 2. The following paragraphs will provide an overview on the use of RS for supporting field work and crop area estimation during an agricultural census. More detailed guidelines on the use of RS and aerial photos for agricultural statistics in general, including censuses and surveys can be found in the Global Strategy (2017b).
Support to field work

Whatsoever the type of frame used (list or area) for an agricultural census, the enumerator will benefit from the availability of satellite imagery or aerial photos. Satellite images or aerial photos can help enumerators optimize their displacements and facilitate localization of holdings and fields. Aerial photos or very high-resolution imagery will help enumerators access the land and/or locate the holder. Used as paper prints or on a mobile device, imagery will also minimize the obvious declaration and measurement errors.

Area estimation

The use of aerial photography for measuring areas of fields during an agricultural census is limited by the fact that agricultural census data are collected from agricultural holdings and the use of RS or aerial photographs implies that each field of a holding covered by the census is identified by the enumerator and the holder on available images or photos. With technological advances, aerial photographs and orto photos are becoming less costly but still require substantial resources to ensure that up-to-date photos are available at the time of the enumerator’s visit. Some of the technical issues related to direct use of aerial photos are now being overcome with their transformation into orto-photos using computers (distortions in size of fields on a photo due to difficulties in keeping horizontal flight at a constant altitude and due to uneven terrain). Over the last few decades, ortho-correction algorithms have dramatically improved, which has enabled reductions in cost.

Two main methods that can be used to derive area statistics: (i) Pixel counting and (ii) Calibration methods. But few countries actually use these methods to derive their national crop area statistics. For more information, see Global Strategy (2015b).

In recent years, the potential use of drones (unmanned aerial vehicles) is being widely discussed. The ortho-rectification and mosaicking technique is sufficiently developed to produce documents with acceptable accuracy, and the dates of image acquisition can be chosen with a flexibility similar to that presented by a field survey. In many countries, drones are limited by regulations that often forbid flight beyond the sight of the operator (Global Strategy, 2015a), while coverage range is limited by the capacity of the drone battery. This technology may play a growing role in the future. Small low-altitude piloted aircrafts provide images with resolution of approximately 5 cm. They have the advantage of being more frequently authorized to fly long stripes (around 100 km by 100 m) and efficient stripe-based sampling plans can be defined for small aircrafts, especially for the estimation of nomadic livestock.

Handheld GPS

GPS is the oldest and most widely used Global Navigation Satellite System (GNSS). A GNSS is 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 geolocation (longitude, latitude, and elevation) on the Earth’s surface.

Generally, a GPS provides support to field activities: geo-referencing plots or holdings; locating the known coordinates of holdings; or measuring the area of a plot or landscape patch. The use of GPS in building frames is discussed in Chapter 13 of Volume 2. This section will provide a short overview of its use in support of census data collection.

Geo-referencing plots, holder’s housing unit or holding headquarters (for non-household holdings such as companies, etc.): When the frame has been defined on the basis of coordinates (possibly in a geographic information system (GIS) environment), points with given coordinates in the field must be located. The two main tools for locating a point are ortho-photographic documents and GPS. For each enumeration area (EA), housing units can be identified using GPS coordinates on handheld GPS devices or mobile digital device equipped with GPS. In the case of mobile digital devices, as the holders or holdings
to be visited are known, GPS coordinates can be used by a navigation option built for the device to allow enumerators to reach the holder’s house or holding headquarter easily. An extensive elaboration on both the use of GIS and GPS is presented in Chapter III B of UN (2009).

**Measuring plots with GPS:** GPS is very useful to measure the area of single plots on the field. For small fields, at the end of the twentieth century and beginning of the twenty-first century, area measurement with GPS was considered insufficiently accurate. GPS measurements used to be less precise but faster than traditional measurement by tape and compass. In an FAO pilot project, a small negative bias of GPS measurement was observed (Keita and Carfagna, 2009). However, more recent studies (Carletto, Jolliffe and Banerjee, 2015) strongly suggest that technological advancements in more recent years with moderate-priced GPS have led to significant improvements even for small plots, especially when signals of more than one satellite constellation can be combined (GPS and GLONASS currently; Galileo and BeiDou should follow). More information on the use of GPS for crop area measurement is available in a World Bank publication on the use of GPS (World Bank, 2016).

**GPS as a tool for quality control.** In some area frame surveys, surveyors will sometimes make observations from unsuitable points due to location errors or to a certain level of negligence, especially if weather conditions are bad. This type of error can be controlled if surveyors are required to record the point from which the observation is made with a protected GPS device (Global Strategy, 2015a).

**GPS for point location control in objective measurement yield surveys (when non-essential census data are to be collected).** A possible source of bias in yield surveys with objective measurements on a sample of points relates to the determination of the point in which the crop sample will be collected. The traditional approach consists of providing rules on the number of steps that the surveyor should take in certain directions. The movements are determined with the help of a random number table. In some pilot projects (Taylor et al., 1997) it has been observed that surveyors are not very rigorous in applying the rules when their supervisors are not present. The sampling process is more rigorous if coordinates are sampled before the field work and then recorded in the field with a GPS device, with a picture of the location being taken (Global Strategy, 2015a).

**Using mobile devices for census data collection**

**Using mobile devices with packages for computer-assisted personal Interviewing (CAPI)**

This is a fast-changing area and other new devices are likely to emerge in the next decade. Recent years have brought many new mobile devices into the market. Many companies produce mobile devices of varying sizes, functionalities and prices to meet consumer demand. As a result, CAPI applications are being developed to leverage these devices for data collection. The application of CAPI to mobile devices was an important evolution, as mobile devices are often cheaper than computers, contain built-in GPS, and are more portable. Current CAPI applications vary widely in terms of learning curve, features, and cost. A one-size-fits-all CAPI application or device that can meet the demands of every type of census or survey does not exist. A review of current applications with their strength and limitations can be found in IRIS (2011).

Literature comparing the use of paper and pen interviewing (PAPI) with CAPI applications shows that CAPI can reduce cost, improve data quality, and decrease the time between data collection and analysis. The primary source of cost savings is higher data quality resulting in lower data cleaning cost. Automated crosschecks and strictly enforced skip patterns are common features...
that dramatically increase data quality. Additionally, because data are entered directly on the device during the interview, the expense of data entry is eliminated (King, et al. 2013; Liesher, 2014).

**Questionnaire design, data capture, quality check and transmission using CAPI**

**Questionnaire design**

Frequently, the most challenging step of implementing a census using CAPI is programming the questionnaire. The amount of programming knowledge required varies widely which directly impacts long-term sustainability. Additionally, expensive consultants must sometimes be mobilized, which increases costs and further undermines sustainability. CAPI products that feature user-friendly interfaces and do not require knowledge of programming are more sustainable and are therefore recommended for countries with limited expertise in using new technologies.

**Box 2 Enumerator reports lower respondent burden using CAPI in Saint Lucia**

Saint Lucia has been one of the first countries in the Caribbean to implement surveys using CAPI. In interviews regarding their experience using it, field officers reported they were able to enter the data faster than on a traditional paper-based questionnaire. Also, not having to carry a paper and pen allowed them to follow the respondents around the house during the interviews allowing the respondents to complete chores. The combination of shorter interview time and the lack of interruption of household chores reduced the burden on respondents and improved the rate of response.

Faster data entry and reduced interview time is consistent with other studies comparing CAPI and PAPI in other countries as well (King, et al. 2013; Liesher, 2014; Caeyers, Chalmers and De Weedrt, 2010).

**Data capture**

CAPI software tools incorporate the traditional question types (e.g. single-select, multi-select, numeric, date, text, and list) and some novel question types not available with PAPI. Default keyboards and keypads used by the device’s operating system are utilized for data entry of the latter four question types. For single-select and multi-select questions, the interviewer selects the correct options by touching them on the screen. Additionally, some CAPI applications allow the survey designer to incorporate questions which capture the geo-coordinates of the holdings, take photos, and even scan and record barcodes of items which are present within the holding.

**Data duality**

Common tools such as skip patterns used in PAPI are strictly enforced by CAPI eliminating the potential for routing errors through ‘enablement conditions’. Enablement conditions are pre-programmed instructions, which activate or deactivate questions based on the answer to a previous question.

CAPI also provides survey designers with validation conditions which are pre-programmed instructions that anticipate an answer to lie in a certain range. If the validation condition is violated (i.e. an unlikely or impossible response is entered), the interviewer is prompted with a message.

Use of strictly enforced skip patterns, and validation conditions have been shown to completely eliminate data routing errors, and impossible or unlikely entries (Zhang, et. al 2012; Caeyers, Chalmers and De Weedrt, 2010).
Transmission

After an interview is completed, the data are transmitted to a centralized database for quality control, compilation, and analysis. The most common ways to transmit the data are over cell phone data and Wi-Fi networks. Use of Wi-Fi networks is practically a universal feature of mobile devices, and 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.

<table>
<thead>
<tr>
<th>Box 3 Transferring data in areas of low connectivity in the United Republic of Tanzania</th>
</tr>
</thead>
<tbody>
<tr>
<td>The Ministry of Livestock and Fisheries Development (MLFD) in the United Republic of Tanzania implemented a baseline survey of veterinary livestock officers using SuSo during July 2015. The enumerators were required to travel to remote areas of three regions to conduct interviews. Accordingly, the MLFD took a few steps to mitigate the risk posed by unreliable connectivity to cell phone networks and internet connectivity. For example, the enumerators were given SIM cards from two different companies providing a wider range of access to cell phone networks. Furthermore, enumerators were trained to wait to transmit the data until they returned to more populated area where cell phone coverage was better. Finally, field supervisors were given back-up paper questionnaires if no coverage could be found. In the end, not a single interview was conducted using PAPI (Rahiia and Mtui, 2016).</td>
</tr>
</tbody>
</table>

Using mobile devices equipped with GPS for geo-referencing holdings

Many mobile devices contain GPS that use satellites to determine the location of the mobile device. Many CAPI products integrate this feature into data entry applications allowing the enumerator to capture geo-references during interviews. The accuracy of the geo-references depends on the quality of the GPS in the mobile device, satellite coverage, and other factors during the interview (Keita, Elisabetta and Mu’Ammar, 2010).

If accurate geo-references are obtained, they can be used to optimize logistics through enumerators’ navigation and planning transportation, monitor census progress, and provide opportunities to conduct geospatial analysis of the variables obtained.

Using mobile devices equipped with GPS for optimizing logistics and supporting enumerators

Geo-reference information on holdings can be especially useful when enumerators are required to make more than one visit to the holding. Examples of this include:

- Listing exercise is undertaken and enumerators must return to conduct the census enumeration;
- Longitudinal surveys;
- Questionable responses that require follow-up;
- Agricultural surveys which require pre-harvest and post-harvest visits;
- Agricultural surveys involving crops with continuous harvesting; and
- Agricultural censuses using modular approach or integrated census survey modality requiring more than one visit.

Census managers can use geo-referenced information to create maps using free open-source software.

These maps can be shared with enumerators through mobile devices, allowing the enumerator to navigate to the holding. Notably, this can also have the positive effect of minimizing the likelihood of the enumerator going to the incorrect holding. Finally, by using a map, enumerators can plan their best route, which overall maximizes the efficiency of logistics and reduces data collection time.

Mobile devices with GPS and paradata for monitoring census progress

GPS for monitoring

Fraudulent interviews and fabricated data are potential problems when conducting agricultural censuses. When enumerators are paid based on the number of interviews they conduct, there is an incentive for
enumerators to inflate the number of completed interviews by filling in questionnaires with fake information. Accordingly, census managers conduct various quality checks and field supervision to minimize this risk. Requiring enumerators to geo-reference the location of the interviews gives census managers another tool to prevent fabricated data.

Furthermore, plotting geo-references on a map allows census managers to visualize the progress of their census. As a result, they can see differences between areas with regards to the number of interviews conducted and make better informed decisions.

**Paradata for monitoring**

Paradata is information captured about the interview (e.g. date and time of start, completion, approval, etc.). The daily tabulation of this information allows census managers to closely monitor the progress of activities and even detect fraudulent interviews.

Table 1 below shows the start time, end time, and duration of interviews conducted during a day. This table indicates that there is a lot of variation in the time takes, but it is clear that an entire interview taking under thirty minutes would be cause for suspicion.

**Table 1: Tabulation of paradata for census monitoring**

<table>
<thead>
<tr>
<th>id</th>
<th>Interviewer</th>
<th>Names</th>
<th>Starttime</th>
<th>Endtime</th>
<th>Duration</th>
</tr>
</thead>
<tbody>
<tr>
<td>be15</td>
<td>les_inter9</td>
<td>removed</td>
<td>2015-07-21 08:03:13</td>
<td>2015-07-21 09:11:23</td>
<td>0:1:01:09</td>
</tr>
<tr>
<td>44a59</td>
<td>les_inter18</td>
<td>removed</td>
<td>2015-07-21 09:30:30</td>
<td>2015-07-21 10:48:38</td>
<td>0:1:18:08</td>
</tr>
<tr>
<td>ae76</td>
<td>les_inter6</td>
<td>removed</td>
<td>2015-07-21 07:52:08</td>
<td>2015-07-21 08:58:18</td>
<td>0:1:55:10</td>
</tr>
<tr>
<td>98id</td>
<td>les_inter16</td>
<td>removed</td>
<td>2015-07-21 09:06:31</td>
<td>2015-07-21 09:37:21</td>
<td>0:0:30:50</td>
</tr>
<tr>
<td>8917</td>
<td>les_inter3</td>
<td>removed</td>
<td>2015-07-21 07:54:14</td>
<td>2015-07-21 08:48:19</td>
<td>0:0:54:05</td>
</tr>
</tbody>
</table>

Source: **Rahija and Mtui, 2016**

Furthermore, paradata can be used to create daily figures showing progress across enumerators and teams. Figure 1 clearly shows that Teams A and C are performing the best, while Team B is lagging behind. It also indicates that interviewer_10 did not perform any interviews.

**Figure 1: Number of interviews conducted by team and interviewer (simulation)**

Source: **Rahija and Mtui, 2016**
Real time monitoring of field staff can be also very important in contexts where there may be some potential risks for their safety and security may be in danger. Therefore, mobile devices equipped with GPS, by monitoring the geo-location of each field staff can help in addressing security threats.

**Limitations of using mobile devices with CAPI for census data collection**

Censuses and surveys vary widely in terms of size, target population, and variables to be collected. Accordingly, a one-size-fits-all CAPI technology that is ideal for every single type of census or survey does not exist. There are also circumstances in which using a traditional PAPI is more appropriate than using CAPI. The main considerations when deciding between PAPI and CAPI should be size and cost, access to cellular data networks and Wi-Fi, and safety and security.

**Size and cost**

Use of CAPI for census or survey requires the up-front fixed cost of purchasing equipment, and training staff. However, many of the variable costs associated with PAPI such as printing questionnaires, data entry, and data cleaning are completely eliminated or significantly reduced (Caeyers, Chalmers and De Weedrt, 2010). Accordingly, as the population size increases, the marginal cost of conducting a census using CAPI declines. Additionally, if the CAPI will be used for future censuses or surveys, then fixed costs are spread across multiple data collection operations.

**Access to cellular data networks and Wi-Fi**

Though examples of methods to offset the risk of under coverage posed by cellular data networks in remote areas were listed above, there may be circumstances in which the risks posed by lack of coverage are unsurmountable. The transmission of data using CAPI relies upon frequent access to cell phone data networks or Wi-Fi. It may be the case that in certain areas of enumeration, the interviewers go for long periods without access to cell data networks or Wi-Fi. Additionally, mobile device GPS functions sometimes using data networks to improve the accuracy of the geo-referenced information. If cell phone networks are not available, low accuracy of the geo-references may render it unusable. Significant cost increasing measures may be required under these circumstances.

**Access to power source**

Power sources may be non-existent in some rural settlements in some developing countries which may add a major constraint on the use of CAPI. Back-up batteries and/or solar chargers may be needed adding to the cost of use of CAPI.

**Safety and Security**

As with planning any census or survey, the safety and security of enumerators should be a major priority of census or survey managers. Enumerators carrying expensive equipment like mobile devices could make them a target for thieves.

**Other issues**

In some countries, issues related to culture have been mentioned such as reticence of older farmers when young enumerators using CAPI are performing interview. Also, there may be a need for frequently updating the device due to rapid advance of the technology.
References and suggested readings


King, K., Kang, H., Jin, M. & Lew, D. 2013. Feedback control of Swe1p degradation in the yeast morphogenesis checkpoint, Volume 24, number 7, pages 914-922. (also available at http://www.molbiolcell.org/content/24/7/914.full)


Taylor, G., Liu, Y., Baskerville C. & Charbonneau, H. 1997, The Activity of Cdc14p, an Oligomeric Dual Specificity Protein Phosphatase from Saccharomyces cerevisiae, Is Required for Cell Cycle Progression,
