



# Guideline for Testing the Standard Agricultural Ground Reference Data Collection ODK Form

## Background

As the applications of artificial intelligence (AI) and machine learning (ML) in crop analytics become more readily accessible to public and private sector actors, the need for a core set of standardized ground truth data is growing. While a number of organizations have developed effective tools for the collection of in field ground truth data, no generic and standardized tools exist on common mobile data collection platforms. Mobile data collection tools are commonly used by agricultural development projects supporting millions of smallholder farmers around the world. The opportunity exists to leverage these tools to rapidly expand the collection of core ground truth data and to dramatically improve the availability of this data for crop analytics. At the same time, data collection and sharing hubs offer platforms for aggregation, sharing and use of ground truth data in ways that multiple actors can easily access. These factors create an exciting opportunity to accelerate innovation and scale up of crop analytics applications for smallholders across a range of use cases.

## Goal

This test is designed as part of the Enabling Crop Analytics at Scale (ECAAS) project of the Bill & Melinda Gates Foundation. The goal is to create a standard ODK form for collecting core agricultural ground truth data for advanced crop analytics.

## Partners

Tetra Tech was awarded a contract by the Gates Foundation to implement the first phase of ECAAS initiative. ODK and Radiant Earth Foundation are working with Tetra Tech to design the standard ODK form and post-processing data, Q&A and ingestion of the data catalog to enable development of analytics for agricultural monitoring.

## Access the Form

### Data model:

You can see the full list of form questions in [this document](#).

### Form Overview:

You can watch this [short video](#) to see the form in action.

## Download the Form

This form requires Android 5.1 or later. If you are already using ODK, make sure you are not actively collecting data and have sent all existing data to your server before trying this new form, because scanning the QR code below will change your device settings.

1. Go to the Play Store and download the [latest version of ODK Collect](#).
2. When the app launches, tap the ☰ menu in the upper right corner, then tap Configure via QR code.
3. Scan the following QR code:
4. Now press Fill Blank Form and you should see the form named “ECAAS Field Mapper”



## Community Forum

We have established an online community forum that you can access on the ECAAS website: <https://community.cropanalytics.net>. We encourage you to join the forum to connect with the ECAAS team and other members of the community, and share feedback or ask questions about the ODK field mapper form that you will use in this survey.

## How to collect data

The goal of this data collection is to capture polygons on the ground and their corresponding crop properties. With that in mind, you can collect a geo-boundary that is smaller than the actual field boundary if:

- a) The whole perimeter of the field is not accessible.
- b) The perimeter of the field is very large and not walkable.
- c) There are multiple crops within the same field, and you can distinguish their boundaries. In this case, it is requested that you collect separate entries for each crop rather than using the intercropping option.

Follow these steps to record an entry in the form:

- The first question asks how the consent of the farmer was collected to record his/her field's data. This question will guide how the data can be shared later on. Therefore, you need to inform the farmer about the intended use or publication of data and then record that in the form. The actual consent doesn't need to be recorded for each entry, and can be shared with the whole dataset at the end. If no consent is given by the farmer or a Village Head/Chief you cannot collect data.
- The next question will ask whether you are allowed to walk around the field and collect the GPS coordinates of the boundary. If you are not allowed to do so, you will collect one point in the center of the field.
- Next, you are asked if the field is fallow (and then the questions regarding crop type, etc are skipped).
- Next, you will be asked whether this form is being recorded post harvest. (this option can be used to record data such as crop type, water source etc for the field during the previous growing season).
- The next question is about the class of the primary crop type. This form supports data collection for Cereal and Food crops, Perennial crops and Tree crops. Depending on your response, the rest of the questions will be customized.

For Cereal and Food Crops as well as Perennial Crops the questions will be:

- Crop type(s) is the next set of questions. First, you are asked if this is an intercropped field. If the field is intercropped, you will be asked to enter the percentage of the field that is intercropped. Then, the primary crop type is asked, and if it is intercropped, the secondary crop(s) are recorded. The list contains some existing crops (most common), and there is a "Other" class if the crop you want to enter is not among the pre-populated ones. You can select only one crop as the Primary. Please use a consistent naming convention to enter any new crops. We recommend using [FAO AGROVAC](#) ensure cross-domain specificity. The list/taxonomy of the crop names should be provided to surveyors in advance.
- [for cereal crops only] The next question will be about the crop development stage.
- Planting and Harvest Date are the next two questions. For both of them you can either enter a full date or a Month and Year. If the dates are not known, you can select "Date not known". It's recommended to capture the full date for both of these if they are known. This will support developing better in-season forecasting models.

For Tree Crops the questions will be

- The primary crop will be asked first. If it's coffee, the next question will be about the variety of the coffee.
- Next question will record the number of trees in the field.
- Then the year that the majority of the trees have been planted is asked.
- Finally the Harvest Date is recorded. You can either enter a full date or a Month and Year. If the dates are not known, you can select "Date not known". It's recommended to capture the full date if it's known. This will support developing better in-season forecasting models.
- Next, the water source is recorded.
- The next questions will depend on your response about the possibility of recording the field boundary. First, you will be asked to collect a point in the center of the field. Then, if you are able to capture the field boundary coordinates, you will be shown the "GeoShape" module to collect a polygon. You can use either of the three methods to capture the field boundary. Ideally, using the "Automatic location recording" and walking around the boundary is preferred. This method will record a point every 10 sec (default but can be changed) as you walk around the boundary, and the resulting polygon is the optimal data for developing labeled data on satellite imagery and generating improved analytics.

## Feedback

We would like you to try collecting several (7-10) entries using this form, and provide us with your feedback. Questions to keep in mind for the feedback:

- Does the questionnaire address the data requirements for your use case?
  - What questions are missing from the form?
  - Is there information that you cannot capture from the existing questionnaire, and would rather exclude that?
- Is the recommended geospatial boundary collection (Automating location recording) a feasible method for you?
  - Would inclusion of a pre-populated field boundary from satellite imagery facilitate the collection?
- Any other suggestions or feedback related to this exercise.
- This form doesn't contain a question about crop yield. Are you able to collect crop cut measurements to estimate the crop yield if it's included in the form?