In situ data collection for Sen2Agri system

1. Farmers declaration for their fields
2. Existing surveys for ag. statistics or insurance purpose
3. Piggyback on existing household surveys
4. Specific field campaign for Sen2Agri
In situ data collection tool:
- ground survey apps: GeoODK (WFP, Mali), ESRI app, OruxMaps 7.0.2., ...
  from motorbike, car or in the field
- drones (STARS, WFP)
- very light aircrafts (South Africa)

IT platform to compile and quality control the in situ data:
- ONA Platform (WFP)
- Google Earth or Bing combined with in season Sentinel-2 imagery
  for quality control (Mali)
Challenges for in situ data collection

To capture the entire diversity of the cropland and all the non cropland classes

To cover the full diversity of the crop types of interest and all the minor crop types (it is important to define the targeted crop types to emphasize the field data collection on them and to capture all the gradients in crop growing conditions of the crop types of interest)

To integrate the control the quality of the in situ data (labeling error, location error, editing error, etc. ) as the in data quality is of paramount importance

To ensure the spatial distribution of the in situ data across the whole study area to avoid a local overfitting of the classification model to specific growing conditions

To use spatially independent in situ data for robust accuracy assessment (closer the calibration and validation data set are in space, more the accuracy metrics represent the quality of the classification model rather than the quality of the entire map); Sen2Agri system delivers accuracy metrics for the entire map when the in situ covers the whole region of interest
In situ data collection
2 different objectives

- In situ data for **calibration (training)**: sampling to cover the diversity of situations existing in the study site (possibly the national territory) in order to represent the range of possible signatures for the different elements of interest (i.e. croplands vs non-croplands on one hand and, the five main crop types and the other frequent crops on the other hand).

- In situ data for **validation** to estimate the products accuracy (with a confidence interval) using a statistically-sound sampling to be objective and independent; for logistic reasons, sampling not strictly random but 2-stage sampling (with PSU and ESU) to assess the crop types (one field campaign).

\[ \Rightarrow \] Calibration and validation field campaigns for cropland and crop type can be all combined but the sampling design should be explicitly different to be independent.

Sometimes, 2 field campaigns are needed for early and end of season maps when all crop types can not be identified during the mid-season campaign (before some crop emergence).
The **annual cropland from a remote sensing perspective** is a piece of land of minimum 0.25 ha (min. width of 30 m) that is sowed/planted and harvestable at least once within the 12 months after the sowing/planting date. The annual cropland produces an herbaceous cover and is sometimes combined with some tree or woody vegetation**.

*the herbaceous vegetation expressed as fcover (fraction of soil background covered by the living vegetation) is expected to reach at least 30% while the tree or woody (height >2m) cover should typically not exceed a fcover of 20%.

**There are 3 known exceptions to this definition. The first concerns the sugarcane plantation and cassava crop which are included in the cropland class although they have a longer vegetation cycle and are not yearly planted. Second, taken individually, small plots such as legumes do not meet the minimum size criteria of the cropland definition. However, when considered as a continuous heterogeneous field, they should be included in the cropland. The third case is the greenhouse crops that cannot be monitored by remote sensing and are thus excluded from the definition.
1. **Stratification** according to existing agro-ecological zoning to sample the range of diversity

2. **On screen visual interpretation** to select samples (min. 1 ha) of **land cover types different than cropland**

3. **Ground survey** to delineate **crop type samples** (min. 1 ha but larger is better)
Stratification according to existing agro-ecological zoning to sample the range of diversity
ex. Ukraine: 4 zones
1. **Stratification** according to existing agro-ecological zoning to sample the range of diversity.

2. **On screen visual interpretation** to select samples (min. 1 ha) of **land cover types different than cropland** on recent aerial photographs, Google Earth or Bing imagery to capture the diversity of the non cropland land cover types. The sample distribution between strata could also consider the stratum size and their respective diversity (~15 samples by land cover type by stratum)

   ex. Ukraine: 720 samples for non-cropland (15 samples x 12 land cover types x 4 strata)

   => 5 days in office

3. **Ground survey** to delineate **crop type samples** (min. 1 ha but larger is better)
Selection of calibration polygons for each of non-cropland class on Google Earth

Check the image date (recent!)
1. **Stratification** according to existing agro-ecological zoning to sample the range of diversity

2. **On screen visual interpretation** to select samples (min. 1 ha) of *land cover types different than cropland* on recent aerial photographs, Google Earth or Bing imagery to capture the diversity of the non-cropland land cover types. The sample distribution between strata could also consider the stratum size and their respective diversity (~15 samples by land cover type by stratum)
   
   ex. Ukraine: 720 samples for non-cropland (15 samples x 12 land cover types x 4 strata)
   
   => 5 days in office

3. **Ground survey** to delineate *crop type samples* (min. 1 ha but larger is better): for each stratum, 75-100 samples for each main crop and 20-30 samples for each minor crop. No strict sampling design but need to capture each crop diversity. Visual delineation could use the most recent color composite. At this time of the year, summer crops are not yet visible.

   ex. Ukraine: 2000 samples for major crops (5 major crop types x 100 samples x 4 strata)
   840 samples for minor crops (7 minor crop types x 30 samples x 4 strata)
   
   => 10 days before the mid-season
National demo in Ukraine
Windshield survey

First calibration field campaign over Ukraine (21-23 March 2016):
Repartition of the in situ data set

In-situ data set – 7689 parcels
75% for algorithm calibration
25% for products validation

Second calibration field campaign over Ukraine (27-29 June 2016):
Repartition of the in situ data set

4+3 days, 1 car, 2 persons/car
2 weeks/campaigns, including days in the office

Area of interest for the demonstration
- In situ data set from the first calibration campaign
- Sentinel-2 Tiles
- Stratum 1: Woodlands
- Stratum 2: Forest and Steppe
- Stratum 3: Steppe
- Stratum 4: Mountains

Area of interest for the demonstration
- In situ data set from the second field campaign
- Sentinel-2 Tiles
- Stratum 1: Woodlands
- Stratum 2: Forest and Steppe
- Stratum 3: Steppe
- Stratum 4: Mountains

5th Sen2-Agri Webinar, 19 July 2018
1. **Stratification**

2. **On screen interpretation** of non cropland samples on GE imagery

3. **Ground survey** of 75-100 / 20-30 samples per crop and stratum

**CROPLAND MAP**

Calibration set made of non cropland (on screen) and cropland (ground survey) samples to train each stratum separately and produce the **first cropland mask** at the mid-season

**CROP TYPE MAP**

Calibration set made of crop types (ground survey) to train each stratum separately and produce the **first crop type map** at the mid-season

=> 1st field campaign data delivery before the mid-season (calibration only)
➢ Stratification  according to existing agro-ecological zoning to sample the range of diversity

➢ Two-stage sampling strategy for crop type validation
  ❖ Delineation of large Primary Sampling Units (PSU) based on ancillary data set (typically admin. regions)
    ex. : Ukraine = around 30 oblasts (districts); local site = 5 to 15 districts
  ❖ Random selection of few (2 or 3) PSU distributed in the different strata according to their cropland area (cropland area-weighted sampling probability).
  ❖ “Windshield survey” for each selected PSU to identify the crop type for each Elementary Sampling Unit (ESU) along the roads (e.g. tablet onboard of vehicle). ESU corresponds to parcels covered by the same crop/crop association. Delineation should use the best contrasted color composite for parcel identification. The road selection and the ESU selection should be as systematic as possible, unbiased with a min. parcel size larger than 0,25 ha (25 S2 pixels). Typical density of 1 ESU / 100 sq km.
    National case. Ukraine : 2 oblasts/stratum * 4 strata => 8 oblasts => 1600 crop samples
    1 day to reach the selected oblast (PSU), 200 ESUs/day by windshield survey => 16 days campaign
    Local case. 1 district/stratum * 3 strata => 3 districts => 600 crop samples
    200 ESUs/day by windshield survey => 3-4 days campaign

➢ Sampling strategy for non-cropland validation (2 options)
  1) Collection of non cropland ESU during the windshield survey
  2) If up-to-date very high resolution imagery (GE) is available for large parts of the area of interest, on screen identification of cropland/non-cropland samples of randomly selected in each stratum.
    Typically 100 – 150 ESUs of 0,25 ha per stratum. => 3 days in the office