

European Monitoring of Biodiversity in Agricultural Landscapes (EMBAL)

- EMBAL 2022 / 2023 Survey workflow -

December 2023

Date: 2023-12-13

Issue: 1.0

Authors: EFTAS

Table of Content

1. Overview	4
2. Software and tools	5
3. Pre-processing	6
4. Data collection in the field	8
5. Post processing.....	10
6. Quality control.....	12
7. Data exchange	13
ANNEX I: Software tools	13

1. Overview

This document outlines the workflow for data collection in the EMBAL 2022 / 2023 rollout. The first chapter provides an overview of the workflow, while Chapter 2 describes the software and tools utilized for data collection. Chapters 3 to 5 detail the sequential workflow steps applied during the data collection process. Chapters 6 and 7 elaborate on the quality control successfully applied at different stages of the data collection. This structured format ensures a comprehensive understanding of the methodology and processes employed throughout the EMBAL survey.

The EMBAL workflow is based on the EMBAL survey protocol developed during the EMBAL pilot project¹. Relevant documents for the EMBAL 2022 / 2023 rollout include:

- *EMBAL2023_Survey_Manual_v2.0.pdf*
- *EMBAL2023_Key_Species_Guide_v1.0.pdf*
- *EMBAL2022_2023_Sampling_Strategy_v1.0.pdf*

The workflow for recording an EMBAL plot consists of four steps.

1. The **pre-processing**. In order to facilitate the mapping of the plots in the field, each plot has to be pre-processed by means of Computer Assisted Photo Interpretation (CAPI) using QGIS. Based on previously acquired digital orthophotos, a geopackage with the initial delineation of the individual parcels and LE is generated representing the spatial delimitation of parcels and landscape elements within the plot. For the 2023 survey the pre-processing was updated with the modified geometries from the 2022 survey.
2. The actual **field survey** by the surveyors with the ODK Collect app. The required parameters are collected for all elements according to the EMBAL Survey Manual, including changes of geometry and documentation with photos. The collected data is uploaded to a central ODK server.
3. The **post processing** of the recorded data in QGIS. Adjustment of the parcel and LE geometries from the pre-processing based on the findings in the field.
4. The **data delivery and quality control** loops by the Central Office.

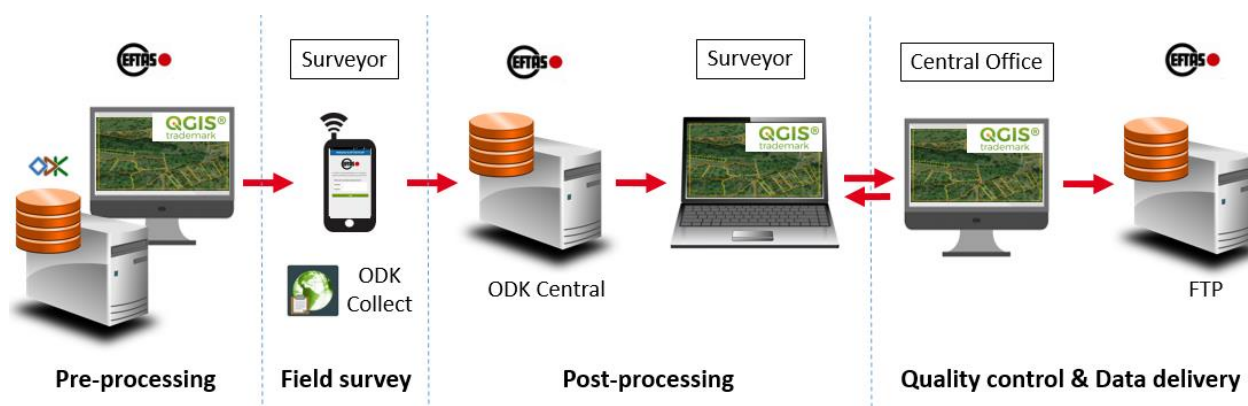


Figure 1: EMBAL workflow.

¹ <https://wikis.ec.europa.eu/download/attachments/25560696/EMBAL%202021%20Survey%20Protocol.pdf?version=1&modificationDate=1634075897799&api=v2>

2. Software and tools

Main tool for the data collection in the field is the open-source software Open Data Kit (ODK). ODK consists of two main components, a server, ODK central, and a mobile App for data collection ODK Collect. The ODK documentation is provided here: <https://docs.getodk.org/getting-started/>.

ODK Central manages user accounts and permissions, stores the form definitions for data collection and the data. The user interface of ODK central allows for some basic management and visualisation of collected data. There are different ways to set up ODK Central server depending on the server infrastructure and operating system. A detailed description how to set up ODK Central is provided here: <https://docs.getodk.org/central-setup/>

ODK Collect app is an Android app which enables the user to perform survey-based data gathering. It can be easily customised and supports the collection of a variety of parameters. The app displays survey forms as a sequence of input prompts, which cover a wide range of data types: besides free text, predefined answers and numeric answers, it also supports the collection of images or locations. ODK Collect connects to the ODK Central server to download forms for data collection and to upload collected data, photos, etc... The connection from ODK Collect to ODK central server is established via URL and individual username and passwords provided to the surveyors. Further information is provided here: <https://docs.getodk.org/collect-intro/>

The data entry in ODK Collect is structured via customised data entry **forms** created in xlsx format. Forms define the parameters to be collected and be adjusted via entry constrains, entry conditions, repeating sub-structures or other logical sequences, etc. The forms are uploaded to the ODK central server and are from there downloaded to the Collect App to start filling the relevant fields. Details how to create forms for ODK is described here: <https://docs.getodk.org/form-design-intro/>.

The forms used for the EMBAL survey are provided as deliverable:

- ***EMBAL_ODK_FORM_1_Parcel_Survey_2023_v1.xlsx***
- ***EMBAL_ODK_FORM_2_Transect_Survey_2023_v1.xlsx***
- ***EMBAL_ODK_FORM_3_Overall_Plot_Survey_2023_v1.xlsx***

Each form consist of four tables:

- *survey* – contains the parameters to be observed in the field and rules and structure for data entry as described in the EMBAL 2023 Survey Manual.
- *choices* - contain list of predefined data entries
- *external_choices* – contain the information on surveyor id, plot id and parcel ids. It needs to be filled with:
 - *SU_ID* – list of all surveyor ID in Column B and C.
 - *Plot_2023* – list of all EMBAL_IDs selected for the survey in column B and C, and the Surveyor ID assigned to the plot in column D.
 - *n_id* – list of all polygon ids created in the pre-processing in column B and C and the respective EMBAL_ID in column E
- *settings* – define the name of the form, the url for the odk server and versioning

QGIS (<https://qgis.org/en/site/>) is a free and open source Geographic Information System (GIS). The great advantage of qgis is on the one hand its wide range of functions and on the other hand its cost-free use, which is of great importance especially on the level of the surveyors. At least version 3.22 is required. QGIS is used for the pre-processing of the EMBAL plots and for the post-processing of the survey data, it is further recommended for the quality control of the data.

ODKImport is a plugin for QGIS, developed by EFTAS GmbH within the framework of the EMBAL project. This plugin facilitates the downloading of survey data from ODK Central in geopackage format and seamlessly imports them into QGIS software. Optionally, the plugin supports the download of photos captured during the survey. The plugin is provided as a .zip file and can be easily installed in local QGIS installations through the plugin menu.

- **EMBAL_QGIS_odk_import_plugin.zip**

The plugin requires the URL of the ODK server, E-Mail, Password, and Project ID defined in ODK Central and an output folder to download the data Figure 2. In a further step the data download can be selected per form, surveyor Id or EMBAL_ID.

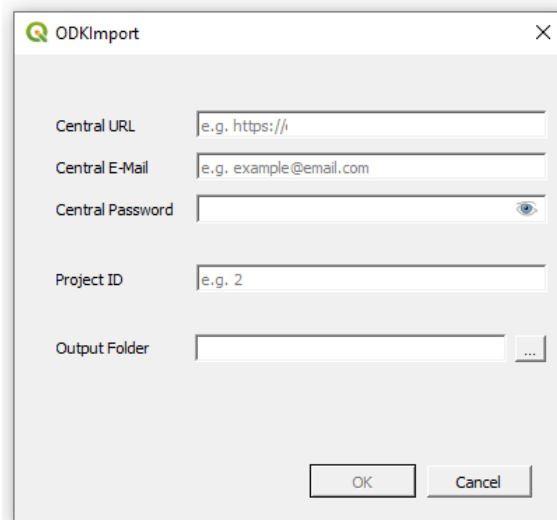


Figure 2: ODK Import plugin

Please be advised that the plugin has been exclusively developed by EFTAS GmbH for the purpose of EMBAL rollout 2022 / 2023. It is not an official QGIS plugin and is not available in an official repository. Additionally, please note that the plugin is provided 'as-is,' and there is no warranty associated with its usage or performance.

3. Pre-processing

The EMBAL plots chosen for the field survey are selected from the EMBAL Master Sampling Frame. For further details on the selection of EMBAL plots, please consult the provided *EMBAL2022 2023 Sampling Strategy*. Each selected plot is represented by a square polygon measuring 500x500 meters, clearly delineating the boundaries of the EMBAL plot. In addition, five transect identification points are systematically distributed.

Delineation of Parcels and LEs:

In the pre-processing each all parcels and landscape elements (LEs) within the selected EMBAL plot are delineated based on Computer Assisted Photo Interpretation (CAPI). The operation is performed using QGIS and delineation is based on visual interpretation of previously acquired Digital Orthophotos (DOPs). All parcels and landscape elements are delineated in compliance with the EMBAL definitions and according to the defined minimum mapping units:

- Minimum mapping width: 1m
All parcels/landscape elements being mainly narrower than 1 m are not recorded, those elements have to be included in the neighbouring elements.
- Minimum mapping unit: 25m²
Only parcels/landscape elements with a minimum area of 25 m² are to be delineated and classified as a separate polygon. If the feature is smaller, it is to be considered as a part of a neighbouring polygon. The threshold of 25 m² also applies for small fragments of actually bigger parcels at the edge of the EMBAL plot.

The delineated polygons are saved as individual polygons in one geopackage per plot (*EMBAL_ID_CAPI_Parcel*). It contains all relevant attributes for the field survey, including a consecutive numbering (*n_id*) of all polygons. The combination of Plot ID (*EMBAL_ID*) and Parcel / LE ID (*n_id*) forms the unique key identification for the polygon (*FEATURE_ID*).

Allocation of transect lines:

Up to 9 vegetation transects, representing an area of 20 x 2.5 m, are recorded in each EMBAL plot, comprising four transect pairs (A-D) and one central transect (E). Transect pairs A-D include margin and interior transects, while the central transect E starts at the center point of the EMBAL plot, the LUCAS point, and is recorded only if it is located on grassland, harmonizing with the LUCAS Grassland Module.

During the CAPI, the five transect indication points are verified to ensure they fall on agricultural parcels. If a transect indication point is situated on a landscape element or a non-agricultural parcel, the transects are shifted to the nearest agricultural parcel. An exception is made for the central E transect, which is not shifted and is only recorded if it is located on grassland.

Within the selected parcels, the transect pairs are located at the nearest edge of the parcel. The position of the transects adheres to specific conditions:

- Should be accessible by the surveyor.
- Should be at a true land cover border (not borders between two types of crops).
- Should be with a minimum of 10 m distance from the nearest parcel corner.
- For arable parcels, the transects should be located in the headland and follow the seeding rows.

Margin transects are positioned parallel to the edge of the parcel, while the interior transect is located with a 10m distance to the border at preferably a 90° angle to the margin transect. More specific rules for special cases are described in the EMBAL Survey Manual. Figure 3 shows an example of an EMBAL plot after the post-processing with delineated parcels and LEs (blue) and transect position lines (orange).



Figure 3: Example of a plot with delineated parcels / landscape elements, 5 transect identification points and 9 actual transect locations.

The generated polygons and transect lines serve as essential references for the surveyors and are compiled into printed Ground documents and digital background maps for ODK (MBtiles), forming the foundation for the field survey. Any disparities between the pre-processing delineations and the actual field conditions are meticulously documented during the survey, leading to subsequent adjustments to the corresponding polygons in the post-processing phase. While field verification is a standard practice to ensure spatial accuracy, digitization efforts are only necessary when deviations arise between the pre-processed geometries and the current on-site conditions.

4. Data collection in the field

The field data collection in EMBAL is conducted by trained surveyors. Having reached an EMBAL plot, the surveyors collect the data via smartphone using the ODK Collect Survey App and the customised entry forms for parcel, plot and transect. Geometry changes are additionally noted by hand on the printed ground document.

In the ODK Collect App the three different digital survey forms are used to collect the data:

- **(1) EMBAL_1_Parcel_Survey_2023:** Collection of parameters for all agriculturally used parcels, landscape elements (LE) and non-agricultural areas within the plot. The individual features are identified by the unique FEATURE_ID and by recording a GPS point (parcel position) inside the feature. In addition to capturing parameters, the surveyor documents whether adjustments to the geometry are necessary.

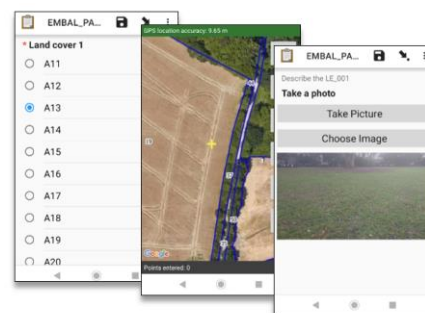


Figure 4: ODK Collect form

- This includes specifying if new features need to be created or if existing features should be merged (deleted).
- **(2) EMBAL_2_Transect_Survey_2023:** Collection of more detailed parameters for all proposed vegetation transects (line of 20 m length and 2.5 m width) in **grassland or arable parcels**. Within one EMBAL plot, up to 9 transect recordings are required. This includes 4 transect pairs, each consisting of a field margin transect and a field interior transect, along with an additional transect at the center point of the EMBAL plot. A separate form needs to be filled for each individual transect entering all the parameters and taking photos. While transect locations are generally predefined on the Ground Document, adjustments may be necessary in the field due to specific circumstances, adhering to predefined rules for shifting transects. For every transect, a GPS point is recorded at the general transect position, and a GPS line is documented along the actual observed transect.
- **(3) EMBAL_3_Overall_Plot_Survey_2023:** Collection of parameters and photos describing the whole EMBAL plot. form is intended to be completed only once for each EMBAL plot. During the form-filling process, a single GPS point is recorded, marking the surveyor's position at that moment. This ensures that a geospatial reference is associated with the collected parameters and photos.

Upon the completion of the survey for all features within a plot, the surveyor proceeds to upload the finalized forms, along with associated photos to the ODK Central server. This upload ensures that the collected data, including the comprehensive documentation and photos, are securely stored and accessible for post- processing and quality control.

5. Post processing

Post-processing in the EMBAL workflow is carried out by the surveyor upon completing the field survey. Following the data collection in the field, each EMBAL plot will require further post-processing by the surveyor in QGIS. In this process the data from pre-processing and data collected during the field survey will be combined to create final datasets per EMBAL plot. Only the parcel/LE data requires further post-processing, the other datasets generally only need to be checked for completeness and correctness (Overall plot data, Transect data).

The following tasks have to be completed:

- For Parcel and LEs:
 - Geometrical adjustment of pre-processed polygons based on field findings.
 - Integration of Parcel and LE information, gathered in the field as points, with the pre-processed polygons.
- For all Elements
 - Comprehensive quality control checks.

The required data for the post-processing are the polygons from the pre-processing provided to the surveyor:

- *E-xxxx-yyyy_CAPI_Parcel.gpkg*

the raw field data which can be imported with ODK Import plugin in QGIS in the following format:

- *E-xxxx-yyyy_Parcel_Position.gpkg*
- *E-xxxx-yyyy_FIELD_Transect_points.gpkg*
- *E-xxxx-yyyy_FIELD_Transect_lines.gpkg*
- *E-xxxx-yyyy_FIELD_Overall_plot.gpkg*

and the orthophotos used in pre-processing as background image. The orthophotos are provided either as individual datasets or as a WMS/WMTS layer. Additionally, supportive layers, such as open access image services like Google Satellite and BING Maps, are incorporated to enhance the geospatial context and visualization of the surveyed area.

In the first step the **geometries from the CAPI_Parcel are modified** according to the findings in the field and based on the handwritten notes on the field map and digital notes in the ODK forms. This is only required if there are actual geometry changes observed in the field. Geometries can either be extended, reduced, extended and reduced, newly created or deleted. This can be realized with the standard geometry editing tools in QGIS:

- **Split Features**  : Splits selected features into two or more new features.
- **Fill Ring**  : Creates a new polygon inside another one.
- **Merge**  : Merges two or more polygons to a single polygon



Figure 4: Example of a geometry change. On the left is the initial state, on the right the final state after the change in QGIS.

Geometry changes are indicated in the *FIELD_Parcel_Position* in the column *geo_mod*. Planned changes are called “*extent*” or “*reduce*”, “*extend and reduce*”, or “*delete*”. New geometries to be created are indicated by an entry in the column *new_n_id*. Each *FIELD_Parcel_Position* entry is indicated as a point on the map.

- After modification of the *CAPL_Parcel* for the entire plot, every polygon must have an unique *FEATURE_ID* and it must be the same as the associated *FEATURE_ID* of the point in *FIELD_Parcel_Position*.

The QGIS function ***Join attributes by field value*** is used to join the information from both layers using the *FEATURE_ID* as unique identifier for the join.

The output is the *FINAL_Parcel* geopackage contain the adjusted geometries with the collected parameter from the field survey.

After post-processing all datasets for Parcel, Transect, and Overall Plot data, are checked for obvious errors or geometry issues following defined control steps. In the case of incorrect or missing parameters, these are corrected using the attribute table of the geopackages.

In the finally step the final datasets are forwarded to the Central Office for Quality Control via FTP.

6. Quality control

This chapter provides an overview of the implemented quality control measures in the EMBAL 2022 / 2023 rollout. Quality control is systematically conducted at different stages throughout the EMBAL survey process.

During the Field Survey:

- Quality control measures are integrated into the ODK entry forms during field recording to enhance data accuracy and completeness. These measures include features that prevent incomplete data entry and detect inaccuracies, triggering warnings when implausible parameter combinations are encountered.
- *First Flush*: The "First Flush" phase involves immediate quality control loops following the first two initial surveyed plots from each surveyor. The Quality Control team provides prompt feedback, preferably within 24 hours, to swiftly identify and correct any misunderstandings or systematic errors at the onset of the survey phase. This proactive approach not only ensures data accuracy but also boosts the surveyor's confidence in their decisions contributing to an improved survey speed.

In the post-processing phase:

- Multiple quality control measures are implemented. The surveyor conducts checks for missing data, corrupted geometries, and the plausibility of entered values. Additionally, photos taken during the survey can be downloaded and utilized to verify the recorded parameters.

Following the upload of the post-processed data, the central office and the quality control team perform several quality control steps, including:

- Verification of the formal completeness of the delivered data. This involves ensuring that all plots, parcels, and transects have been recorded, and all required datasets are included. For elements that could not be observed (e.g., due to inaccessibility in the field), an explanation and the reason for non-observation must be provided. This step ensures a thorough assessment of the completeness and transparency of the delivered datasets.
- Quality control on the survey parameters is conducted by a dedicated and qualified team. Parameters are cross-checked with ground documents and photos. This control is executed on a sample of plots from each surveyor. Any identified issues are promptly reported back to the surveyor for correction. If necessary, further checks are applied to other plots surveyed by the same surveyor.
- In the final stage of data processing, a thorough data cleaning and harmonization process is carried out. This involves removing geometry errors, eliminating data fragments, addressing obvious errors in transect lines, and standardizing the naming of photos.

Depending on the identified errors and the need for correction, multiple iterations of quality control loops may be initiated, involving collaboration between the surveyors and the quality control team. This

iterative process ensures that identified issues are addressed, corrections are made, and the dataset undergoes thorough scrutiny until an acceptable level of data quality is achieved.

7. Data exchange

Within the workflow of EMBAL, data is distributed between the administration/coordinators, quality control and the surveyors via FTP. Therefore, the use of an FTP client is mandatory, which allows up- and downloads at surveyor's level. Each surveyor has access to a personal folder, which contains the relevant data for the survey of the assigned plots (ground documents, lot overviews, ODK access data, background images, style files) and to a public folder which contains survey manuals, instructions, survey sheets and data for the training.

ANNEX I: Tools developed for EMBAL

Filename	Description
• <i>EMBAL_QGIS_odk_import_plugin.zip</i>	QGIS Plugin to import field data from the ODK server to QGIS in geopackage format.
• <i>EMBAL_ODK_FORM_1_Parcel_Survey_2023_v1.xlsx</i> • <i>EMBAL_ODK_FORM_2_Transect_Survey_2023_v1.xlsx</i> • <i>EMBAL_ODK_FORM_3_Overall_Plot_Survey_2023_v1.xlsx</i>	ODK Forms for data collection in the field. The forms have to be initially uploaded to the ODK server to create the data structure and to allow surveyors to download the entry from in the ODK Collect mobile App..