European Commission

DIRECTORATE-GENERAL

ENVIRONMENT

Directorate D – Natural Capital

Unit D2 – Biodiversity

EU project No. 07.0202/2019/815700/SER/ENV.D.2

European monitoring of biodiversity in agricultural landscapes (EMBAL)

Survey Protocol

(v1.0)

E F T A S Fernerkundung
Technologietransfer GmbH

In cooperation with:

- Institute for Agroecology and Biodiversity – IFAB
- Environment Agency Austria – UBA / EAA

As consortium partners,

and

- Danube Delta National Research Institute
- Fundación Global Nature
- Centre for Ecology & Hydrology at Lancaster Environment Centre

As sub-contractors

Date: 24 September 2021

Document Control Information

Settings	Value
Document Title:	EMBAL Survey Protocol
Project Title:	European monitoring of biodiversity in agricultural landscapes (EMBAL)
Document Authors:	Dirk Lindemann, Michael Kobs, Rainer Oppermann, Laura Sutcliffe,
Project Owner:	Georg Altrogge
Project Manager:	Carsten Haub
Doc. Version:	V1.0
Sensitivity:	Basic, Internal use
Distribution:	European Commission, DG ENV
Date:	02/07/2021

Document Approver(s) and Reviewer(s):

Name	Role	Action	Date
Dirk Lindemann	Technical Coordinator	Review	02/07/2021
Carsten Haub	Project Manager	Approval	02/07/2021

Document history:

Revision	Date	Created by	Short Description of Changes
EMBAL2021_Survey-Protovol_v0.9	2 July 2021	Dirk Lindemann	Updates of open chapters

Applicable Documents

ID	Document Name / Content
AD 01	Terms of Reference dated 30 April 2019
AD 02	Technical Offer dated 3 June 2019

Table of content

List of tables

List of figures

Introduction

Scope of the project

EMBAL (European Monitoring of Biodiversity in Agricultural Landscapes) delivers a robust monitoring tool and contributes towards the monitoring of farmland biodiversity across the European Member States at EU-27 level. As a sampling approach based on field surveys, EMBAL serves several purposes:

- Recording of the current situation of land cover and biodiversity throughout EU-27 in order to obtain up-to-date data, constantly updated by a continuous monitoring cycle;

- Recording the types, quality and extent of land cover and landscape elements within the agricultural landscape with a common European-wide approach.

- Photo documentation of different examples of land use and landscape elements.

About this document

The EMBAL protocol describes the technical workflow in the preparation and execution of the EMBAL field survey. The main focus is on the description of the general workflow, which was tested and applied during the development of the EMBAL pilot project 2020/21.

This protocol deliberately does not go into a detailed description of individual steps in the sense of a complete set of instructions for individual tasks. Since the technical procedure for preparation, field recording and subsequent post-processing may differ from the workflow of the pilot campaign or the software and server solutions used, reference is made to the relevant existing detailed technical documents.

This protocol is divided into the following sections:

I. SURVEY PREPARATIONS

- Master plot sample and spatial levels of plot observation
- Optimal survey periods
- Qualification of personnel and training concept

II. TECHNICAL WORKFLOW

- Technical framework
- Pre-processing
- Ground documents
- Mobile data collection in the field
- Post-processing
- Data flow

III. QUALITY CONTROL

- Formal quality control
- Content quality control

IV. FINAL DATA CONVERSION

V. LIST OF TECHNICAL DOCUMENTS

I. SURVEY PREPARATION

1 Sampling

1.1 EMBAL master sample design

The sample for the EMBAL survey follows the regular LUCAS Master Sample 2x2 km grid, covering 1.090.863 points in EU-27.

The projection is ETRS89 Lambert Azimuthal Equal-Area (EPSG 3035), which is compliant with the "Data Specification on Geographical Grid Systems" of INSPIRE.

- LUCAS master grid:
  https://ec.europa.eu/eurostat/web/lucas/data/lucas-grid

- Inspire:
  https://inspire.ec.europa.eu/forum/pages/view/4989/european-statistical-1x1km-grid-etrs89-laea

Around each LUCAS point, the EMBAL plot boxes (500 m x 500 m in length) and the 5 transect starting points per plot have been calculated (see Figure 1).

According to the nature of a metric grid with an EU-wide projection is would be easy to add EMBAL plots in order to create a 1 x 1 km grid (for example, to be able to address specific issues at regional level).
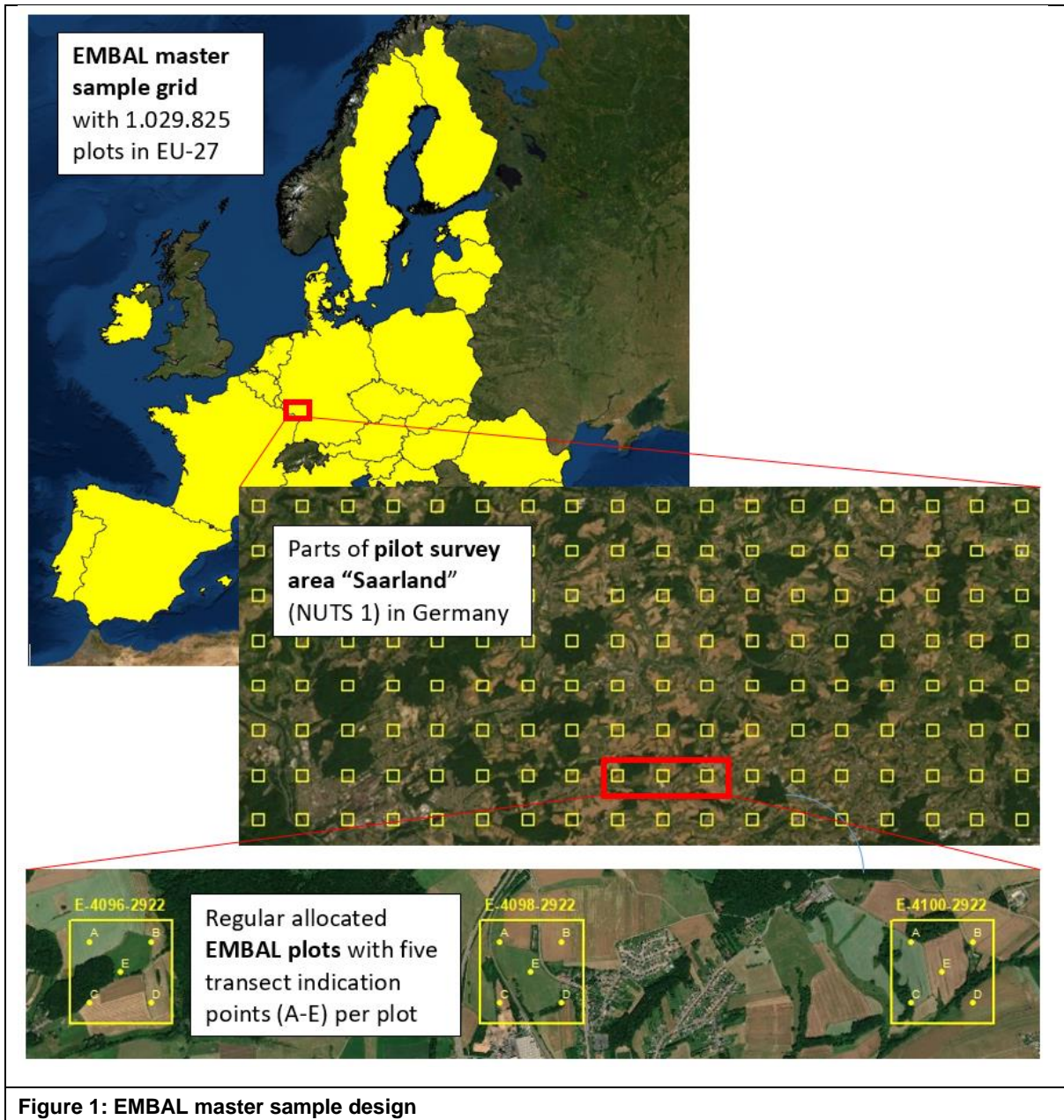
Figure 1: EMBAL master sample design

1.2 Pre-selection of potentially suitable plots for field survey

For plots to be part of a field sample, they must contain at least 10% agricultural land.

For a pre-selection of suitable plots out of the total number of 1.029.825 plots in EU-27, the master sample can be intersected with LUCAS and Copernicus data. This method was already tested and applied for the selection of the pilot survey plots. The aim of this approach is

- to exclude plots from the master sample (e.g. due to their complete location in forests, built-up areas or in water bodies)
- to calculate in advance the proportion of agricultural and other land per plot (with Copernicus data)
- to identify the presence and extend of structural elements (potentially possible with LUCAS transect data)

Especially the CORINE Landcover dataset from 2018 was tested and used for the pilot survey stratification in 2020/21 (see Table 1).

Table 1: Criteria for the selection of potential field sample plots

CORINE Land Cover 2018:	
- Particularly **suitable at European level**, as it is available for all member states, follows a **standardised coverage system** and is subject to a **regular update** - Plots will be automatically **categorised** regarding their **percentage of open land**, with focus on arable land and grassland. - This evaluation will allow a proportional coverage of agricultural land in the plots in 10% intervals. - Plots, which are completely covered by woodland, artificial land or water, will be **excluded** from the field sample. **Structural disadvantages** are the existing minimum mapping unit (MMU) of 25 ha, the geometric accuracy just below 100 m and the thematic inaccuracy up to 15%. - In order to compensate the lack of geometric precision and the thematic inaccuracy compared to EMBAL requirements, a minimum area of CLC classes 212, 242 or 243 of at least 20% within a plot (= 5 ha) is proposed to filter plots for a potential field sample.	

Other potential EU-wide spatial sources for fine-tuning of field sample selection criteria are

- LUCAS field results (2018)
- LUCAS transect information (2015)
- Copernicus High Resolution Layers (HRL)
  - Grassland
  - Imperviousness
  - Forest types
- IACS parcels / field blocks (if available)

As they played a minor role during the 2020/21 pilot campaign, they will not be further discussed here.

The final selection of a field sample takes into account the area of the member states, their percentage of agricultural land and the distribution of biogeographical regions of Europe, which are graduated according to altitude zonation.

1.3 Three spatial levels of plot observation

A detailed description of the different hierarchies of plot survey in the field and the parameters to be surveyed will be described in the EMBAL Survey Manual.

In summary, at each plot selected for field survey the following three level are used for plot survey:

1. EMBAL plot (500 m x 500 m)

Recording unit with 500 x 500 m edge length (25 ha), arranged in a regular grid of 2 x 2 km across the EU-27 Member States, based on the LUCAS master sample grid (Eurostat 2019).

Here the general character of the landscape is described and a rough estimate of the nature value is made.

2. Parcels and landscape elements

Within each EMBAL plot, all agricultural parcels and linear landscape elements are delineated and described by recording a basic set of parameters, such as land cover and land use, number and colours of flowering forbs, coverage of crop and wild plants on arable land or vigour and graminoid-forb ratio on grassland.

3. Vegetation transects

Each plot contains up to 9 vegetation transects, whose position is determined by 5 regularly arranged identification points. Vegetation transects are 20 m in length and 2.5 m wide and surveyed in pairs – one at the field margin, and one in the field interior. They are only observed in either grassland or cropland. Here, more detailed parameters, such as the presence of plant indicator species (key species), different vegetation layers or number of flowering forbs and their density will be recorded. Here also a photo documentation is made.

Figure 2 documents the different recording units in a central European plot. The map shows the status of the preparation after the aerial image interpretation in the pre-processing:

- Delineation of parcels and landscape elements
- Location of (up to) nine vegetation transects created on the basis of the five stable transect identification points (A-E).

Figure 2: Example plot with parcels/LE, location of 5 transect identification points (A-E) and related transects

A_1 – D_1: field border transects

A_2 – D_2: inner field transects

E: from E-point (EMBAL centre) 20m straight to the east (only in grassland, comparable with LUCAS grassland survey)

2 Survey period

Since the EMBAL survey consists of both the mapping of the land cover and the recording of vegetation, the **timing of the fieldwork is crucial**. Depending on the biogeographic region and the elevation of the plots, the optimum time for carrying out the vegetation transects may vary considerably. It is essential that surveys take place during the "height" of the growing season, which might already be in April for the southern Mediterranean countries or in late June or July for the Scandinavian countries and at higher elevations. This means that the vegetation is well developed, and many species are flowering, but meadows have not yet been cut and crops have not yet been harvested.

The recommended timeframe is based on the expert appraisal of phenological variation in the EU implemented in the LUCAS grassland module, shown in the following Figure 3. The table is part of the technical documentation in I.1.1_Survey period.

Figure 3: Time framing of the LUCAS- and the EMBAL-survey

Please note that there is a spatial and altitudinal variation in the survey periods according to the biogeographic zones (see figure 4 and 5 and the altitudinal zones (0-200m, 200-500m, 500-1000m, 1000-1500m).

In Figures Figure **4** and Figure **5** the survey regions 1-10 are given, Figure 4 shows an overview of the main regions and Figure 5 shows the further differentiation in subregions, e.g. region 2n, 2m and 2s. For these subregions (see I.1.2_ Survey regions) different survey times apply e.g. in region 2n (Northern Scandinavia) the survey has to be carried out later then in region 2s (Southern Scandinavia): in 2n at the same altitude the survey starts 4 pentades = 20 days = about 3 weeks later than in 2s.
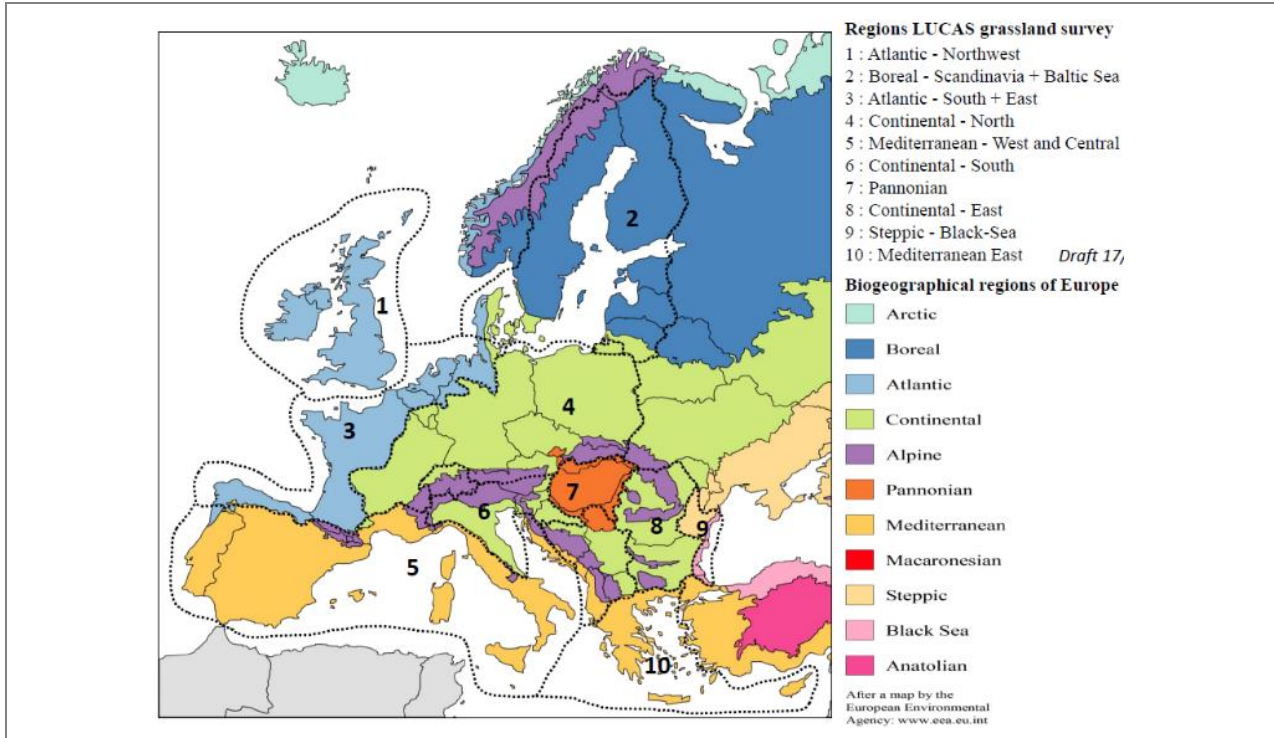

Figure 4: The 10 survey regions 1-10 are shown against the background of the biogeographical regions in Europe.
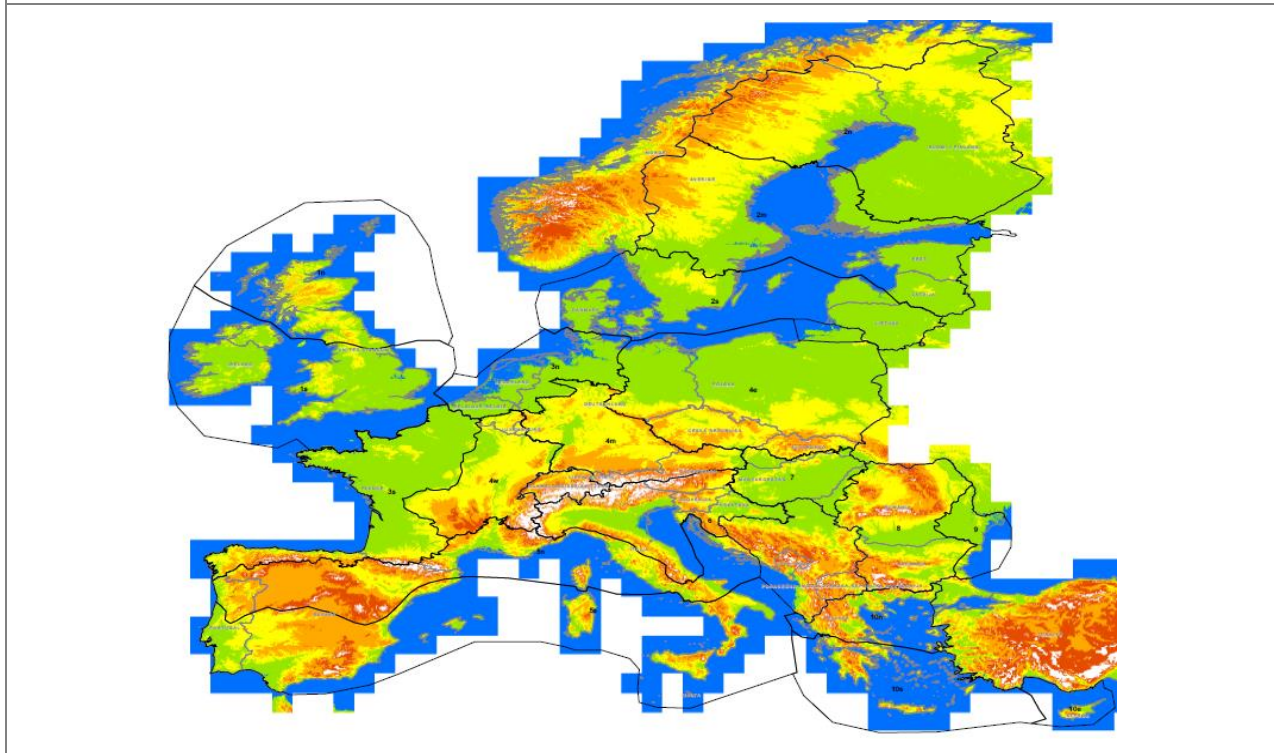

Figure 5: The survey regions 1-10 are divided into subregions (e.g. 1n, 1s, 2n, 2m, 2s, etc.). These subregions are relevant for the relevant execution times (time framing) of the survey.

3 Qualification of personnel and training concept

3.1 Qualification of personnel

For the complete execution of an EMBAL survey, employees with different technical qualifications are necessary.

In the following, the essential technical tasks and the necessary employee qualifications are listed:

Pre-processing of plots

- Aerial photo analysis
- Basic GIS knowledge (digitisation, attribution)
- Map creation with GIS (production of ground documents per plot)
- Image data conversions

Server and data collection application

- Setting up a data server on which the collected data is stored (in pilot campaign first realised with ODK Aggregate (survey in DE in 2020), then with its successor ODK Central (surveys in AT, ES, RO in 2021))
- Creation of digital entry application including data forms (in 2020/21 realised with ODK Collect)
- Data management, data export, spatial data conversions (in 2020/21 realised with ODK Briefcase)

Field Survey

- Rapid field approach with non-botanists, but surveyors should have an environmental science background (biology / geography / ecology / agriculture / forestry or similar) and be experienced in field surveys in general

- GIS knowledge is clearly an advantage for post-processing of captured field data with delineated objects from pre-processing (parcels, landscape elements)

3.2 Training concept for surveyors

The field recording and further digital processing of EMBAL data requires the knowledge and training of various project-specific contents. Table 2 lists the most important aspects that must be prepared and carried out for a training session.

Table 2: Content for training of surveyors

Tasks	Content
Methodology	• Sample design • Survey method and requirements • Introduction of the parameters
Digital data collection	• Overview digital workflow • Installation and use of digital data collection app • Exercise: Data collection in the field of a training plot
Data post-processing	• Knowledge of the individual data components • Update of object geometries (parcels / LE) in GIS • Merging the tabulated data collected in the field with updated object geometries in GIS • Exercise based on collected training data
Data quality	• Formal data quality • Content data quality
Data flow	• FTP server structure • Delivery of data via FTP
Administrative	• Survey areas (plot assignment to surveyors) • Survey period • Mobility and accommodation • Communication • Safety in the field • Access rights • Training evaluation

It is recommended that the different training components be presented by different trainers, each specialising in different tasks.

The scope of the training should be at least 3 full working days, with 1 day reserved for the survey of a specially prepared training plot.

The surveyors are to be provided with the essential training materials prior to the training (survey manual and technical workflow documents).

II. TECHNICAL WORKFLOW

4   IT concept

4.1 Technical framework

In order to facilitate the implementation of EMBAL, a set of digital tools has to be setup and concerted in a specific data workflow. It enables the requested collection of various parameters on biodiversity for monitoring purposes.

- The underlying IT concept ensures the following key characteristics:
- Time-efficient and accurate data collection in the field
- Thorough and reliable data transmission procedures
- Open source solutions
- Implementation of adequate quality control methods

In general, the workflow comprises three main stages:

**Pre-processing** of the mapping units (see chapter 4.2 Pre-processing):

> In order to facilitate the survey of the plots in the field, each plot will be pre-processed by means of Computer Assisted Photo Interpretation (CAPI) using QGIS (or similar, see https://www.qgis.org/en/site). Based on previously acquired digital orthophotos, two shapefiles will be generated, representing the spatial delimitation of parcels and landscape elements within the plot and the location of the vegetation transects

Field survey (see chapter 4.4 Data collection):

> The data recording in the field will be based on mobile devices. The implemented digital survey tool (in 2020/21 ODK, see https://getodk.org) will ensure efficient data entry, while the pre-processed shapefiles serves as a basis to record the spatial extent of land cover and landscape elements.

Post-processing of recorded data (see chapter 4.5 Post-processing):

> Data from the field will be further processed by the surveyors via QGIS (or similar) to create clean mapping results, which are in line with the EMBAL survey methodology. This also comprises quality control procedures to ensure data plausibility by applying a dual control principle. A File Transfer Protocol server (FTP) will serve as an interface to ensure data exchange that is customized for each user.

Figure 6: Overview on IT workflow
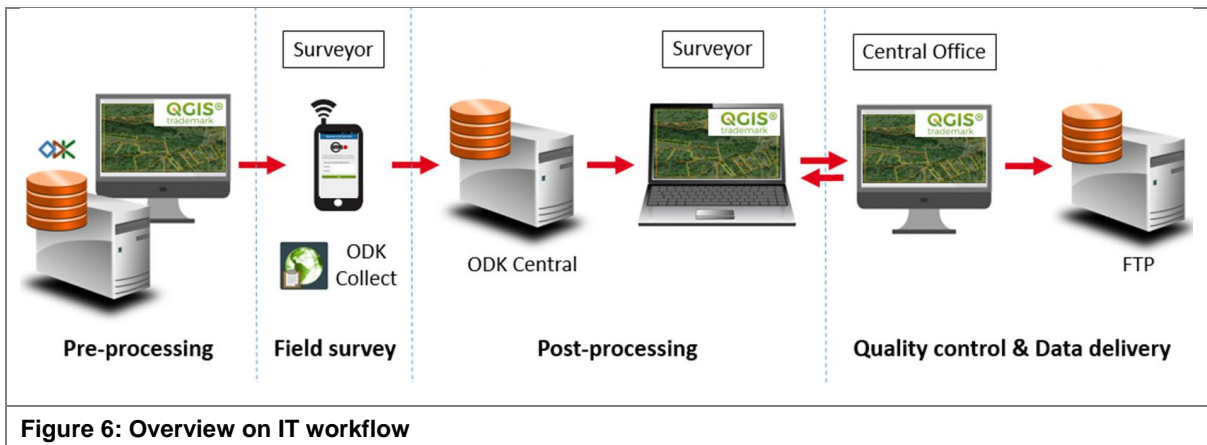
All proposed and tested digital tools are open-source and therefore free of charge and modifiable. After being utilized in the pilot surveys 2020/21, these can be further fine-tuned based on the practical experiences (see Figure 6).

4.2 Pre-processing

Based on the regular LUCAS Master Sample 2 x 2 km grid, a shapefile is generated providing 500 x 500 m polygons to mark the boundaries of each potential EMBAL plot. Also, 5 regularly distributed transect identification points are assigned to each plot.

4.2.1 Parcels & LE

In each EMBAL plot selected for the field survey, all parcels and landscape elements (LEs) will be delineated based on Computer Assisted Photo Interpretation (CAPI). It is recommended to perform the operation using QGIS. The delineation will be based on visual interpretation of previously acquired Digital Orthophotos (DOPs). All parcels and landscape elements will be delineated and saved as individual polygons in one shapefile (Capi.shp) per plot. It will contain all relevant attributes for the field survey, including a consecutive numbering (ID) of all polygons. The combination of Plot ID and Parcel / LE ID forms the unique key identification for the polygon.
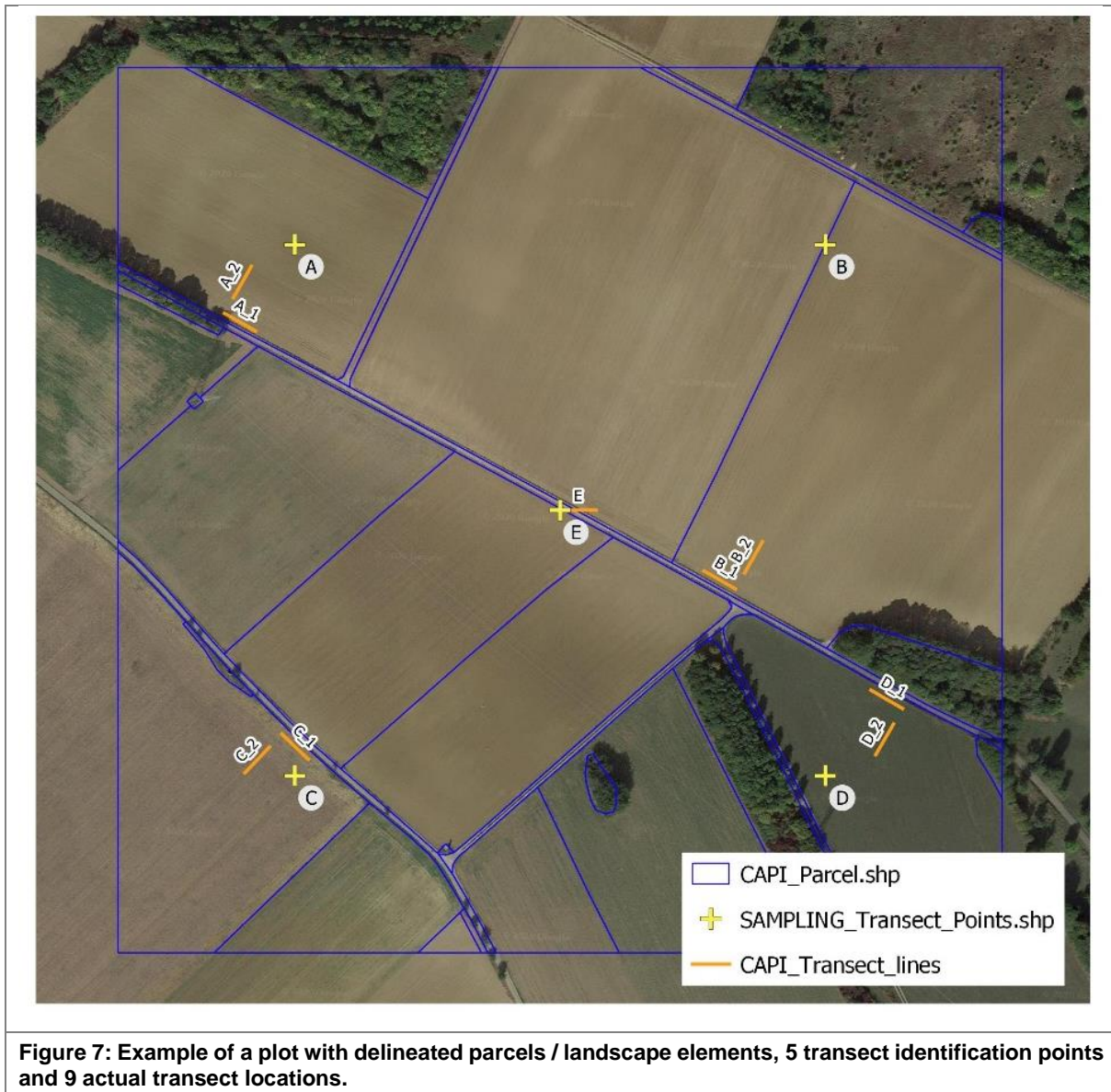
4.2.2 Transects

During the CAPI the five transect indication points are verified, if they fall on agricultural parcels. According to the rules laid out in the EMBAL Survey Manual, the actual vegetation transect lines (4 transect pairs + one transect in the centre of the plot) will be placed on the edge and in the interior of each indicated agricultural parcel. In case a transect indication point is located on a landscape element or on a parcel with non-agricultural use, the transects will be shifted to a suitable agricultural parcel following a predefined selection rule.

The complete workflow (using QGIS) is documented in technical document II.1.1 Pre-processing of plots.

A topological result is shown in Figure 7.

Generated datasets should also be checked for completeness, correct topologies etc. via automated routines.

Figure 7: Example of a plot with delineated parcels / landscape elements, 5 transect identification points and 9 actual transect locations.

The shapefiles will be made available to the surveyors and form the basis for the survey in the field. It will speed up the data collection, as the survey does not start from scratch. Parcels and landscape elements are already created in the shapefile, can be specifically described by the surveyor and attributed with relevant parameters. Obviously, the spatial extent always needs to be verified in the field, but digitization efforts are only needed in case the pre-processed geometries differ from the current situation on the ground (see chapter 4.4 Data collection for further details).

4.2.3 MBTiles

During the pilot surveys in 2020/21 ODK Collect was used for mobile data collection. As the ODK app can display map layers only saved as a set of tiles in the MBTiles format, the shapefiles need to be converted accordingly. QGIS can be used in order to create these files. In fact, the surveyors in the field will be displayed an *image* of the plot and parcel/LE polygons in ODK, and not the original geometries itself (see technical document II.1.2 Compilation of MBTiles for ODK).

The pre-processed data will be made available online to the surveyors via FTP and need to be copied individually on the respective mobile devices (see II.3.2 ODK Collect application for surveyors).

4.3 Ground documents

Next to the digital infrastructure it is important to generate so called ground documents per plot in PDF or JPF format. They are used for the orientation of the surveyors in the field and have to contain the following information:

- EMBAL ID
- Country ID
- Year of survey
- Aerial photo with parcels and landscape elements incl. their IDs from pre-processing (scale approx.1:3.000)
    - Scale bar
    - Image projection and copyright/source
    - North arrow
- Map extract for general orientation (scale approx.1:50.000)

    - Scale bar
    - Copyright/source
- Legend
- Logos of GD Environment and the involved companies/institutions

An example ground document is shown in Figure 8.

The layout of ground documents needs to be approved by DG ENV before starting the production for all plots. A model is presented in technical document II.1.3 Model of ground document.
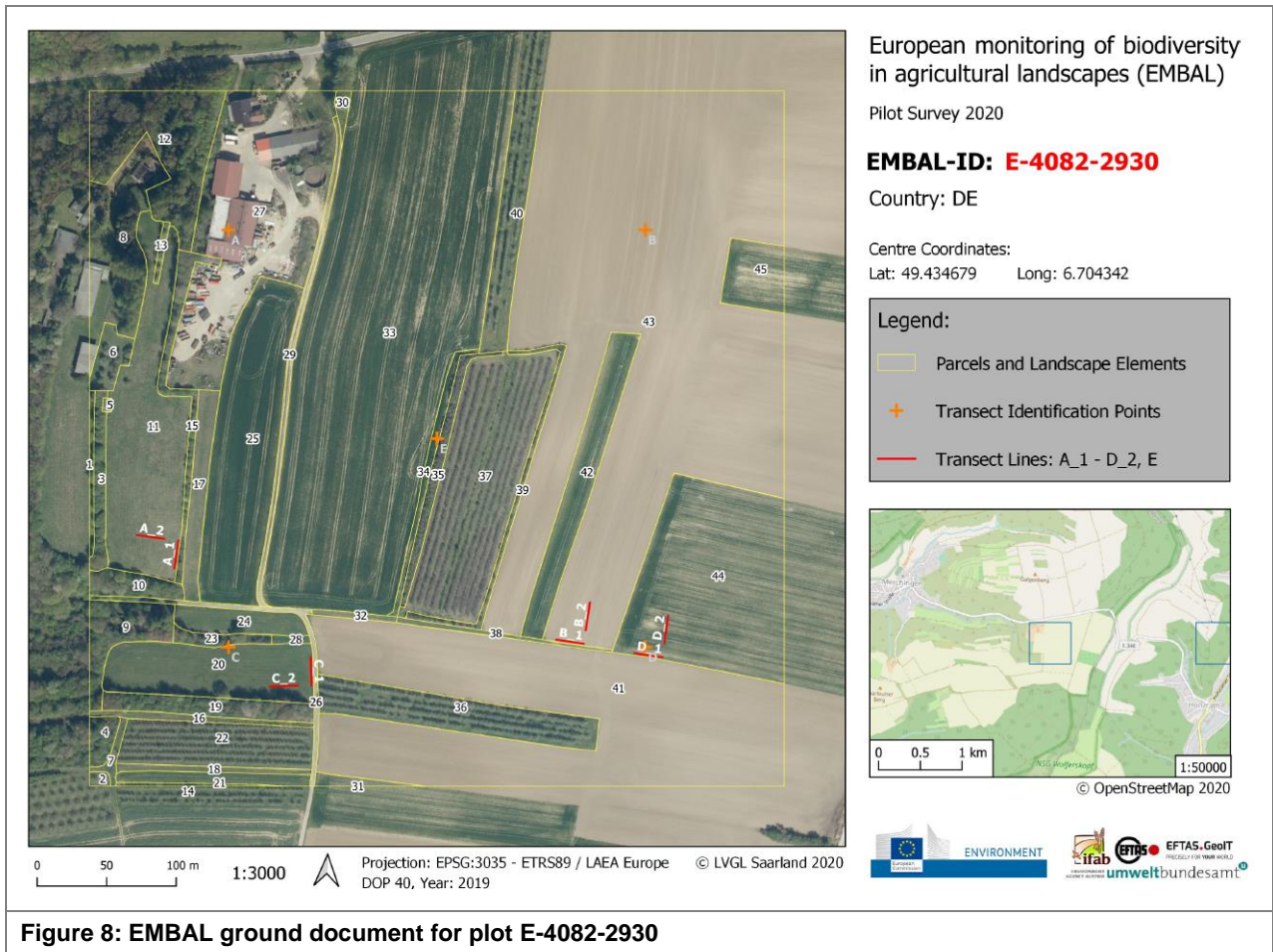
Figure 8: EMBAL ground document for plot E-4082-2930

4.4 Data collection

The data collection in the field is described via the survey applications of ODK (Open Data Kit), which runs on Android smartphones or tablets. However, surveyors should also be equipped with hardcopies of ground documents and field forms as backup solutions.

ODK is a suite of open source tools to collect and manage data. A selection of these tools will be part of the proposed workflow and further described below:

- ODK Collect (Android app to collect survey data in the field)
- ODK Central (a server to save, export and visualize collected data as well as host blank survey forms)
- ODK Briefcase (a desktop tool to pull and export data from ODK Central)

4.4.1 ODK Collect app

Main tool for the data collection in the field during the pilot surveys 2020/21 was the ODK Collect app. This open-source Android app 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. Forms can be adjusted via entry constrains, entry conditions, repeating sub-structures or

other logical sequences. The detailed workflow is described in technical document II.2.2 ODK Collect application for surveyors.
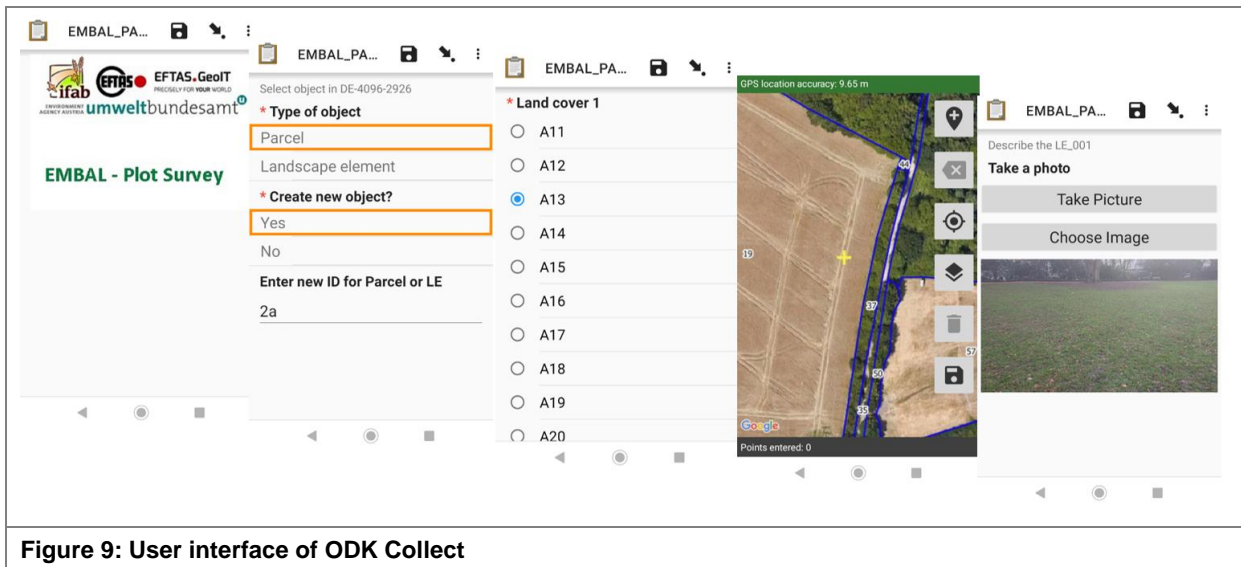

Figure 9: User interface of ODK Collect

Indicative changes on the geometry of existing parcels / LEs or new geometries are recorded through the ODK Collect mapping function. *The ODK mapping function does not allow editing shapefiles directly*; only new features can be recorded as overlays to indicate the changes:

In Detail: Customised field maps are used as background map containing the digital orthophotos and the pre-processed (*Capi.shp*) with the respective IDs of the parcels/landscape elements and their (potential) borders. In case polygons need to be edited (as the orthophoto does not reflect the current situation in the field), the surveyor maps the required spatial changes as a spatial feature within ODK Collect. These changes will be edited into the final dataset at a later stage using QGIS (see chapter 4.5 Post-processing) on a laptop or desktop computer.

The mapping function also allows the surveyor to get oriented in the field, as the current position of the user is displayed in the centre of the customised background map.

The customised field maps with the DOPs and *Capi.shp* will be generated for each plot, converted into MBTiles format (see Chapter 4.2.3) and stored locally on the mobile device.

The app is designed to work well without internet connectivity. Before the survey, all required blank digital field forms have to be downloaded once from the connected server (ODK Central). As the survey for each EMBAL plot consists of several sub-surveys on different spatial levels, several specific digital field forms will be needed:

- Field form for describing the whole plot
- Field form to survey individual parcels / landscape elements
- Field form to survey arable transects
- Field form to survey grassland transects

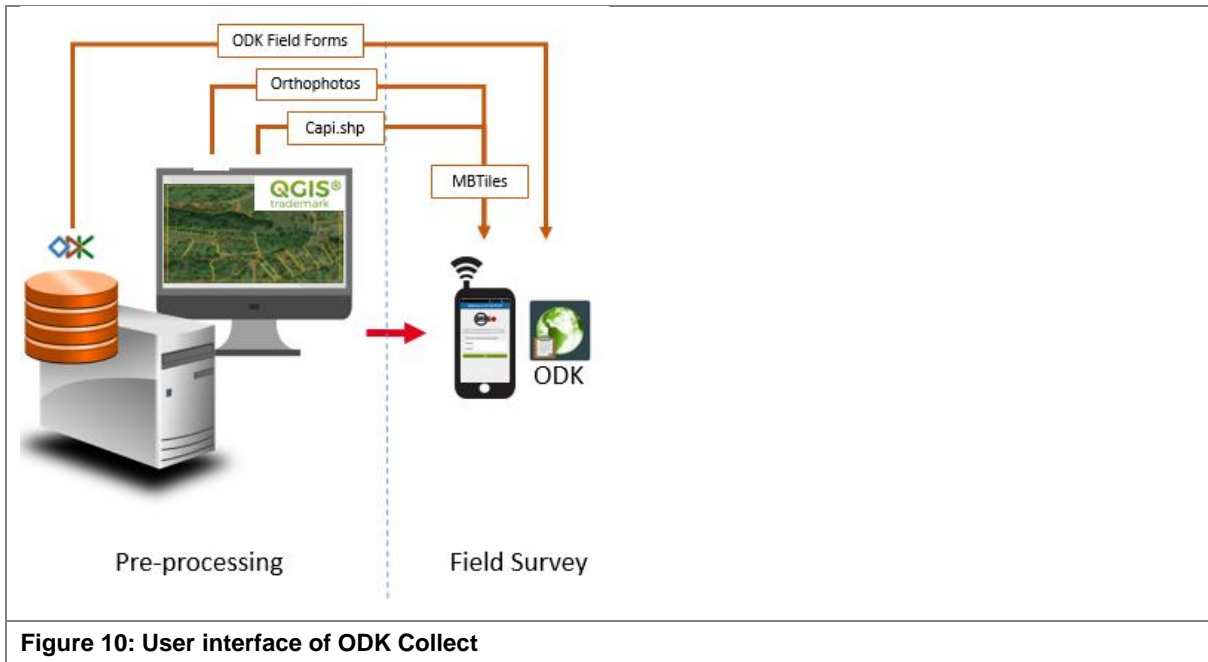Each filled field form will be saved as an individual dataset.

Figure 10: User interface of ODK Collect

All entered parameters will be stored initially on the local device. The data (e.g. free-text, numbers, coordinates) is saved in single *Extensible Markup Language (XML)* files on parcel, LE or transect level. Photos will be saved as .jpg-files, while their file sizes can be managed by defining the maximum image resolution.

4.4.2 Server exports to surveyors

To export the collected data, two options are implemented in the workflow:

1) Upload to ODK Central: with given internet connectivity, all XML-files and photos can be directly uploaded from the mobile device to the central server. This serves as a backup of all collected data. It also allows tracking the mapping progress of individual surveyors (as long as survey data is uploaded regularly).

2) Export via ODK Briefcase: As the surveyor will have to further process the collected field data within QGIS, this tool will be used to export the data from the mobile device to a local computer.

This data conversion step is usually performed by the administrator of the ODK Central server and made available to the surveyors via FTP data exchange.
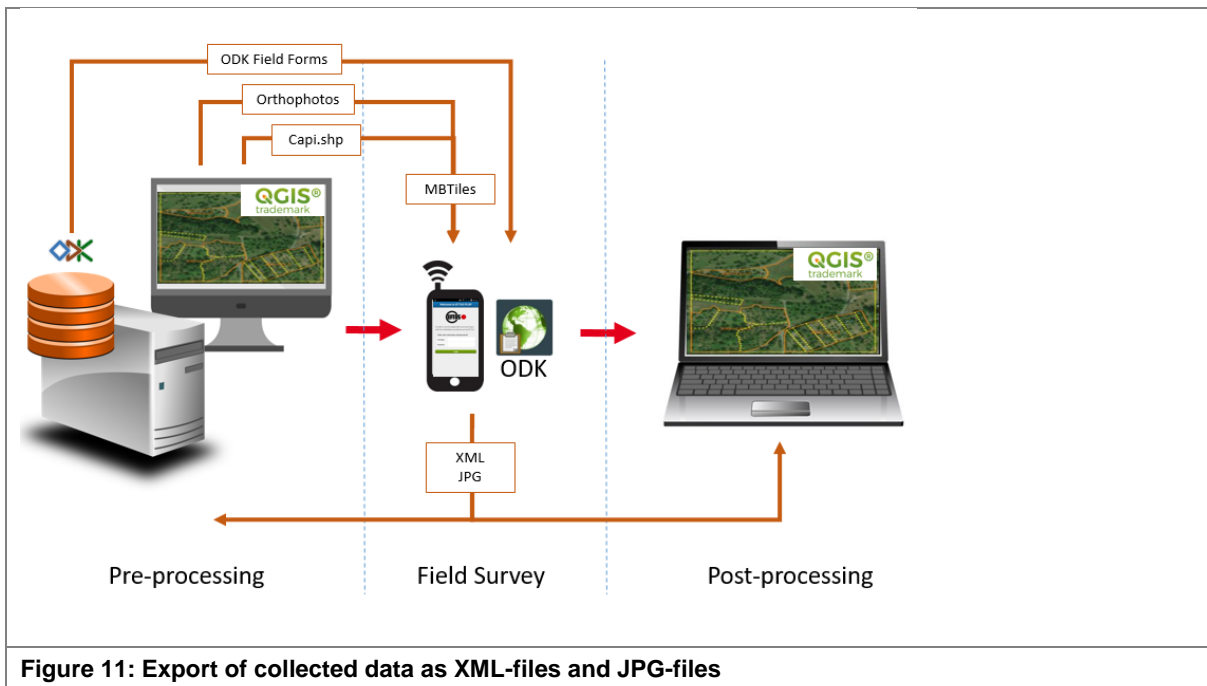
Figure 11: Export of collected data as XML-files and JPG-files

4.5 Post-processing

After having collected all relevant parameters in the field, the data needs to be compiled and edited by the surveyor in order to generate one final dataset for each EMBAL plot.

In detail, the following workflow is foreseen in order to create the finalized datasets:

- **Upload to the central server:** The surveyors will upload the collected field datasets to ODK Central (see technical document II.2.2 ODK Collect application for surveyors)
- **Processing of server datasets**: The raw datasets uploaded to the server will be converted to several shapefiles per plot. These will be made available to the surveyors via FTP. Some shapefiles will be used to carry out the post-processing (data on Parcels/LEs), others only need to be checked for completeness and correctness (Overall plot data, Transect data)
- Execution of Post-processing::
    - **Loading all datasets** (originating from Pre-processing and Field survey) into *QGIS*
    - **Modification of geometries**: Incorporation of all geometry changes indicated in the field by editing the CAPI-shapefile from pre-processing. The edited shapefile will represent the actual delineation of parcels/LEs for each EMBAL plot.
    - **Joining attributes** (collected parameters on parcels/LEs) to the previously modified CAPI-shapefile (containing the delineation of the plot from pre-processing). The shapefile will now represent the actual delineation of parcels/LEs and contain all collected parameters for each feature.
- **Perform quality control**: Implementation of data checks for completeness, correctness etc.
- **Data delivery**: Forward data to the Central Office for Quality Control

4.5.1 Exporting field data from ODK Central server

After having completed the data collection in the field with ODK Collect, the surveyors will upload their datasets to the ODK server. As of 2021, ODK Central was used as the latest server solution for this task; the following documentation therefore focuses on this most recent workflow. The predecessor ODK Aggregate was only used in 2020 for the EMBAL Pilot surveys in Germany.

The raw datasets uploaded to ODK Central will be converted to several shapefiles per plot by the Central office, using the desktop applications ODK Briefcase and QGIS. These shapefiles will be forwarded to the surveyors via FTP and serve as the basis for post-processing.

The following steps are explained in more detail in the technical document II.3.1_Data_export_ODK_Central:

(1) ODK Central – Data hosting

ODK Central is the current ODK server application. Being connected to the data collection client ODK Collect, it serves as an interface providing two main services:

- Hosting digital blank survey forms
- Storing uploaded submissions

Via ODK Central, surveys can be organized in individual projects, including a dedicated user management with role-based permissioning. Besides a fully featured REST API, data output is also compatible to ODK Briefcase.

(2) ODK Briefcase – Import and transformation of data

ODK Briefcase is a desktop application which can be used to pull, push and export forms from ODK servers.

It can be downloaded from GitHub (via https://github.com/opendatakit/briefcase/tags). Before first use, ODK Briefcase has to be setup by providing relevant ODK Central server credentials and specifying a local storage location.

ODK Briefcase is used to:

- import the collected field data (in XML format) from ODK Central to a local computer and
- further transform and export the data as .csv-files (collected parameters) and .geojson-files (geometries).

(3) QGIS – Creation of shapefiles

Being transformed to .csv- and .geojson-files, the collected field data can now be imported into QGIS in order to create shapefiles. This task basically includes two steps:

- Joining the attribute data collected in the field (stored in csv.) to the respective geometries (stored as .geojson)
- Exporting all .geojsons as shapefiles.

As the survey for each EMBAL plot is based on four basic field forms (Parcel, Arable transects / Grassland transects / Overall plot), the shapefile generation will take these four respective datasets into account. Depending on the underlying field form, up to three different shapefiles will be created out of these datasets:

- Parcel survey data

- Arable transect / Grassland transect data

- Overall plot data

(4) Photo renaming

Having exported the survey data via ODK Briefcase, all photos taken during the survey will be stored in the media-folder. Taken via ODK Collect, all photos will initially be named using a unique number, consisting of 13 digits (e.g. 1621253787050.jpg).

To be able to assign the photos directly to a plot, these should be renamed using an explicit name (e.g. E-xxxx-yyyy_13_D_interior_start.jpg for a transect photo or E-xxxx-yyyy_Plot_1.jpg for an Overall plot photo). The initial numerical names of the photos as well as all needed parameters for renaming (e.g. Plot_ID, Transect_ID etc.) are stored in the corresponding .csv-files. This data can be used accordingly to set up an automated batch process for renaming the photos. Finally, the renamed photos will also be provided to the surveyors via FTP.

4.5.2 Joining data from the field with data from pre-processing

Using ODK Briefcase, field datasets can be pulled from ODK Collect (stored on the mobile device) or ODK Central (if previously uploaded to the server) as CSV files for the attributes and Geojson (or KML) for the indicative geometry changes recorded in the field. The datasets also include the photos taken in the field (as JPG files).

Out of these files with indicative geometry changes, one new shapefile (*Changes.shp*) will be generated by the surveyor for each EMBAL plot.

The *Capi.shp* and the *Changes.shp*, linked with the collected data in CSV format, will now be used to create a final *Field.shp* in QGIS. This shapefile will be an edited version of the *Capi.shp*: all necessary spatial edits (indicated by the *Changes.shp*) will be applied, resulting in one shapefile per EMBAL plot representing the actual situation in the field as observed by the surveyor. In combination with all other collected data, the *Field.shp* will form the final Parcel/LE dataset for each EMBAL plot (see Figure 12 and Figure 13).

The whole process is described in a separate technical document II.3 Post-processing.

Figure 12: Capi.shp (blue lines) and Changes.shp (red line) will be used to create a Field.shp (green lines)
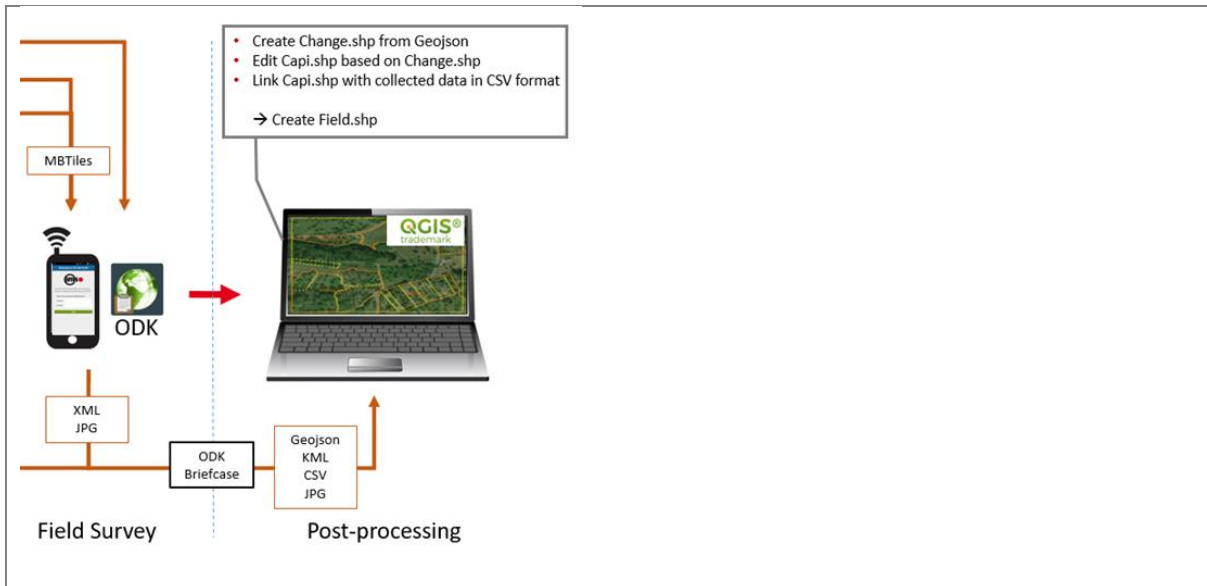
Figure 13: Creating a Field.shp on surveyor level

In a simplified form, this procedure is also carried out for the vegetation transects and the overall plot observation.

The datasets can now be forwarded by the surveyors to be subjected to quality control (see also chapter 5 Quality Control).

4.6 Data flow

Within the workflow of EMBAL, data will be distributed between the administration/coordinators, quality control and the surveyors via FTP.
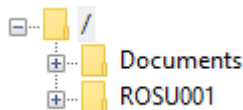
Therefore, the use of an FTP client is proposed, which allows up- and downloads at surveyor's level. The survey coordination will provide each surveyor with individual login data.

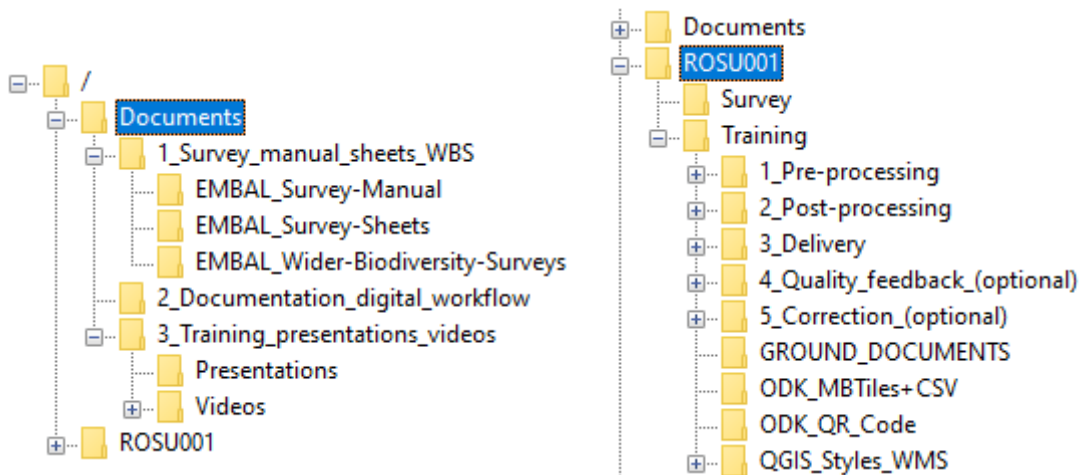The detailed description of login and folder structure used in the pilot surveys in 2021 are described in technical document II.4 Data flow.

4.6.1 FTP Training and survey folder structure

Each surveyor has access to a personal folder, which is both containing all documents and all relevant data. The data folder provides all necessary files to execute the survey (data from pre-processing, MBTiles etc.) but also contains folders for the exchange of data regarding quality control and for final delivery of finalised plots.
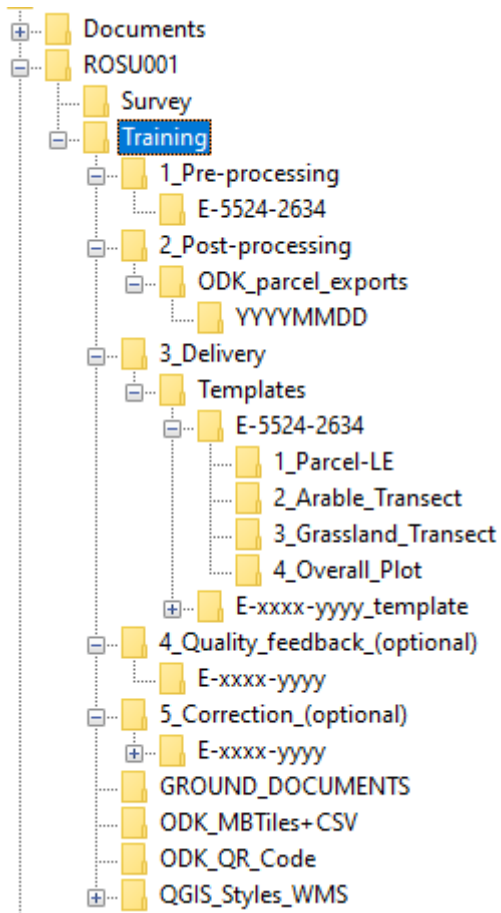
Main folders per surveyor login

```
/
├── Documents
└── ROSU001
```

Basic structure of the main folders (in Romanian pilot survey 2021)

```
/
├── Documents
│   ├── 1_Survey_manual_sheets_WBS
│   │   ├── EMBAL_Survey-Manual
│   │   ├── EMBAL_Survey-Sheets
│   │   └── EMBAL_Wider-Biodiversity-Surveys
│   ├── 2_Documentation_digital_workflow
│   └── 3_Training_presentations_videos
│       ├── Presentations
│       └── Videos
└── ROSU001
```

```
├── Documents
└── ROSU001
    ├── Survey
    └── Training
        ├── 1_Pre-processing
        ├── 2_Post-processing
        ├── 3_Delivery
        ├── 4_Quality_feedback_(optional)
        ├── 5_Correction_(optional)
        ├── GROUND_DOCUMENTS
        ├── ODK_MBTiles+CSV
        ├── ODK_QR_Code
        └── QGIS_Styles_WMS
```

Detailed structure of /training (or /survey)

```
⊞ 📁 Documents
⊟ 📁 ROSU001
   ┊    📁 Survey
   ⊟ 📁 Training
      ⊟ 📁 1_Pre-processing
         ┊    📁 E-5524-2634
      ⊟ 📁 2_Post-processing
         ⊟ 📁 ODK_parcel_exports
            ┊    📁 YYYYMMDD
      ⊟ 📁 3_Delivery
         ⊟ 📁 Templates
            ⊟ 📁 E-5524-2634
               ┊    📁 1_Parcel-LE
               ┊    📁 2_Arable_Transect
               ┊    📁 3_Grassland_Transect
               ┊    📁 4_Overall_Plot
            ⊞ 📁 E-xxxx-yyyy_template
      ⊟ 📁 4_Quality_feedback_(optional)
         ┊    📁 E-xxxx-yyyy
      ⊟ 📁 5_Correction_(optional)
         ⊞ 📁 E-xxxx-yyyy
      ┊    📁 GROUND_DOCUMENTS
      ┊    📁 ODK_MBTiles+CSV
      ┊    📁 ODK_QR_Code
      ⊞ 📁 QGIS_Styles_WMS
```

4.6.2 Data flow structure

The folder structure for **training** and **survey** are identical.

There are five main folders for the data flow:

>   1_Pre-processing
>
>   2_Post-processing
>
>   3_Delivery
>
>   4_Quality_feedback_(optional)
>
>   5_Correction_(optional)

/1_Pre-processing

>   Each folder contains three SHP-files with the results of the pre-processing (CAPI) for each plot.
>
>   The files are needed for the post-processing.

/2_Post-processing

>   The administrator of the ODK Server will provide server exports on a daily basis here. For each new daily export, a new folder (named by the date of export) is created.
>
>   The file(s) provided here can contain the raw data of more than one plot in SHP format and need to be joined with the data from pre-processing (CAPI), which is in detail explained in this manual in chapter 3.

/3_Delivery

>   Provided by surveyors: Each plot delivery folder contains the following sub-folders:
>
>   1_Parcel-LE
>
>   2_Arable_Transect
>
>   3_Grassland_Transect
>
>   4_Overall_Plot
>
>   In each of them, the post-processed and controlled data (one SHP file) needs to be provided.

/4_Quality_feedback_(optional)

>   All delivered data will be quality controlled.
>
>   If quality issues are found, they are documented in the form of point shapefiles per plot and the surveyors need to correct the errors in their local data.

/5_Correction_(optional)

>   Corrected data will be provided here in the same way as in /3_Delivery.

III.   QUALITY CONTROL

5   Quality control

Within EMBAL, a three-stage quality control process is envisaged.

1) In the field (surveyor): In-app control during data entry in ODK collect

- **Filter check** (inner plausibility)**:** The parameter dependencies in the survey manual were taken into account during app development by means of filter queries. Therefore, on the basis of previous entries in a form, only those parameters that refer to the previous entry are displayed in the course of the following data input (e.g. specific sub-selection for nature value parameters on parcels / LE depending on the object's land cover).

- **Value range check**: Only a specified set of allowed values per parameter are provided (exception: free text).

- **Completeness check:** An entry for parcel/LE, transect or overall plot can only be saved and sent to server if it has been filled in completely in the related form.

2) After data delivery (central office): Formal data control per plot

1. **Completeness** of the data sets supplied **per plot** (number of shapes and photos)

2. **Completeness** of the data sets **within the individual data sets** (all relevant parcel/LE, transects, overall plots have been surveyed)

3. **Geometrical integrity** (especially Parcel/LE)

3) After data delivery (central office): Content related checks per plot

- Correct addressing of the **land cover classes**

- Correct application of the **parameters**

- Comprehensibility of the delimitation of objects and the final positioning of the vegetation transects

After the execution of these Quality control processes, all respective data components can be stored in Geodatabases per EMBAL plot and forwarded to DG Environment.

5.1 Formal quality control

The formal control of the data provided by the surveyor is divided into the following parts:

1. Documentation of survey per plot
   o Date of plot survey
   o Date of delivery via FTP (parcel/LE, arable & grassland transects, overall plot)
   o Completeness / No. of photos (overall plot, transect photos)

2. Correct assignment of data to a plot (parcel/LE & transects)

3. Parcel/LE:
   o No. of total objects
   o No. of modified or new objects (compared to pre-processing)
   o No. of documented not observed objects
   o No. of objects with formal inconsistencies (post-processing not completed, missing objects without comment)

4. Transects:
   o No. and type of transects
   o Reasons for not surveyed transects
   o Transect shifts

The individual control results are documented in an Excel table according to the date of the day. A template is provided in III.1_Formal-quality-control-table.

A detailed description of the individual test steps, which are filed in the table, can be found in III.2_Formal-quality-control-of survey-data.

5.2 Content data control

Beside the formal check that the data has been delivered, the quality of the content should be inspected. A plausibility check of the data has to be carried out covering following issues:

- How many parcels/transects were not recorded or had missing data, and what reasons were given for this in the plot record?
- Which combinations of land cover codes were recorded, and do they all follow the rules for LC1/LC2 and can be translated into plausible land cover types?
- Are there any implausible parcel sizes (i.e. too large or too small for the land cover type), and if so, why?
- Which are the most prevalent land cover categories, and does this fit the assumptions for the region?
- Are there unlikely outliers in the parameters recorded (e.g. extremely high flower densities in cereal fields)?
- Are there unlikely combinations of parameters (e.g. very high flower density but no flowering forbs)?
- Do surveyors show different distributions of records, and is this plausible because they surveyed different regions and habitat types or could there be a systematic error, e.g. due to a misunderstanding of the instructions?

Individual parcels or transects flagged as implausible can be checked by experienced landscape ecologists using satellite images and the photos taken by the surveyors to determine the source of the error. In addition, a very small proportion of plots per surveyor can be checked thoroughly by comparing the records to the satellite image and photographs to detect possible errors.

IV. FINAL DATA CONVERSION

The final dataset for the EMBAL pilot survey consists of the ESRI geodatabase "**EMBAL_2020-2021_Field Data.gdb**".

The following components were recorded for each plot in the field:

1. Overall plot information
2. Information on parcels and landscape elements
3. Arable transect information
4. Grassland transect information

During post-processing, four individual ESRI shape files were created for each plot, since all four components have an individual data structure and geometric configuration. In the course of the data compilation, the quality-controlled individual data sets per plot will be combined into four global feature classes, which will contain the field data for all individual plots in EU-27. These SHP feature classes are to be transferred to the geodatabase as feature layers (see IV.1_Generated-database).

Results on	Feature layers in GDB	Geometry
Overall plot	RESULTS_EMBAL_<campaign>_**Overall_Plot**_<date>	Points
Parcels and LE	RESULTS_EMBAL_<campaign>_**Parcel_LE**_<date>	Polygons
Arable transects	RESULTS_EMBAL_<campaign>_**Transect_Arable_Lines**_<date>	Lines
Grassland transects	RESULTS_EMBAL_<campaign>_**Transect_Grassland_Lines**_<date>	Lines

The data structure of the individual RESULT feature layers is described in the technical document IV.2_Structure-of-parameters. Each feature layer is documented here in a single folder.

It should be noted that there are also parameters listed that were necessary for mobile data collection with ODK or for data post-processing, but have been removed from the final data structure. These are marked with "NO" in the column "Part of the final delivery".

For further documentation, additional thematic feature layers can be added to the geodatabase, which are not part of the results.

CAPI (interim results from pre-processing)

- CAPI_EMBAL_*_Transect_Lines
- CAPI_EMBAL_*_Parcel_LE

These are the geometries of the parcels and LE from the pre-processing and the location of the vegetation transects to be surveyed (CAPI = Computer assisted photo-interpretation).

SAMPLE (plot boundaries + transect indication points

- SAMPLE_EMBAL_*_Transect_Points
- SAMPLE_EMBAL_*_Plots

Plot boundaries (as part of the EMBAL master sample) and their individual five transect identification points per plot.

CAPI and SAMPLE information are the basis for the ground document production and form the starting point for field recording with ODK collect.

V. LIST OF TECHNICAL DOCUMENTS

Below are the detailed technical documents produced for the development of the digital workflow for EMBAL pilot surveys in 2020/21 based on the open source components ODK and QGIS.

- I. SURVEY PREPARATION
  - I.1 Survey period & regions
    - EMBAL2021_I.1.1_Survey period
    - EMBAL2021_I.1.2_Survey regions
- II. TECHNICAL WORKFLOW
  - II.1 Pre-processing
    - II.1.1 Pre-processing of plots
    - II.1.2 Compilation of MBTiles for ODK
    - II.1.3 Model of ground document
    - II.1.4 Data naming converntion
  - II.2 Data collection with ODK (Open Data Kid)
    - II.2.2 ODK Collect (data collection application)
  - II.3 Post-processing
  - II.4 Data flow
- III. QUALITY CONTROL
- IV. GENERATED DATABASE