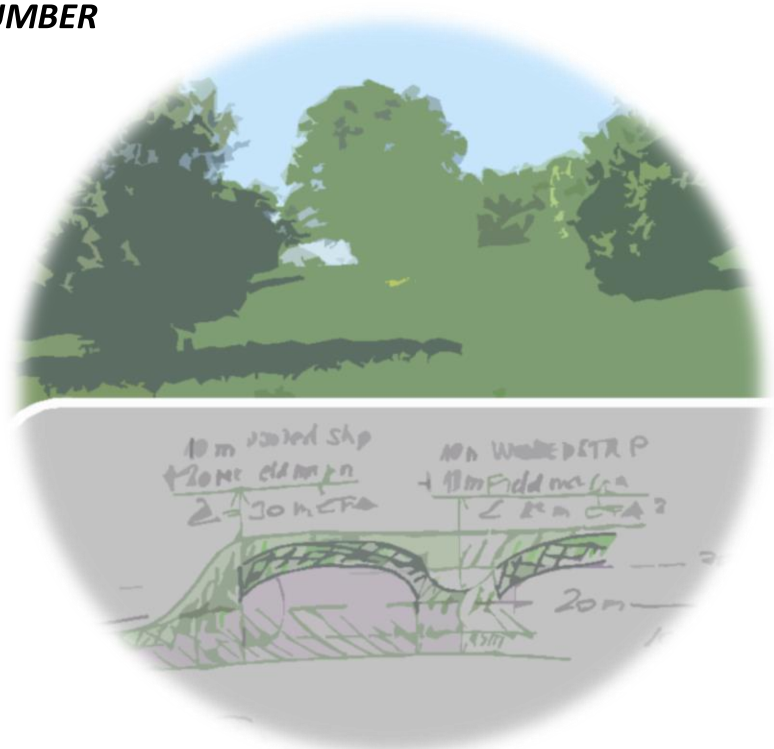
Common conclusions of the Interoperability case studies and the Data Sharing Expert Group

Report under the
EXPERT CONTRACT CONTRACT NUMBER
- CT-EX2018D319529-106

30. January 2022.



Author: Bernadett Csonka

Involved Institutions and experts:

Michal Dobíhal[1], Kateřina Hátlová[2], Lucie Šavelková[2], Ondřej Šilháček[1] - Ministry of Agriculture[1], State Agriculture Intervention Fund[2]

Alba Marcos Sanz, Ana Rita Serna Martínez, Julián Delgado Hernández, Jorge Mata Lopez - Spanish Agrarian Guarantee Fund, FEGA (Ministry of Agriculture, Fisheries and Food, MAPA). Spanish National Geographic Institute (IGN-Spain) (Ministry of Transport, Mobility and Urban Agenda, MITMA).

Jūratė Kučienė - National Paying Agency under the Ministry of Agriculture of the Republic of Lithuania

Traian Crainic and Marian Dumitrescu - Romanian Agriculture Payments and Intervention Agency (APIA)

Rositsa Ivanova-Stoyanova[1], Lora Stoeva[2] – [1] State Fund Agriculture Payment Agency, [2]Forest Research Institute, Bulgarian Academy of Sciences

Contacts: csonkadetti@gmail.com

Contents

Foreword 3
Acknowledgements 3
Abstract 3
1. Introduction and background 4
2. Challenges of the new CAP from year 2023 6
3. Data collection, sematic analysis and usability of third-party databases for updating and extending the landscape feature (LF) dataset 6
3.1. Review of available dataset 6
1.1. Semantic mapping between the proposed broad categories of LF, the existing GAEC/EFA-LF categorization and the corresponding feature types of the created LF layer 8
3.2. Common solutions regarding definition of landscape features 9
4. Comparing 3rd party data to LF-LPIS digitization 11
5. Conclusion and proposed solution 14
References 17
List of Abbreviations and definitions 18
List of figures 19
List of tables 19

Foreword

In order to facilitate interoperability of the Land parcel identification system (LPIS) and the Geospatial Aid Application (GSA) data of IACS, D.5 - Food Security Unit of the JRC has been lunching pilot projects since 2019, in domains of LULUCF, crop classification and landscape features. The scope of the pilot case studies was to assess the usability of existing IACS-GIS in other domains and to experiment the use of third-party datasets for the goals of the CAP, to create valuable input for the new, performance-based CAP framework. Pilot projects were implemented on various test sites in several member states: Spain, Lithuania, Czech-republic, Bulgaria, Romania, Germany, Austria and Bulgaria, while expertise of the Netherlands and Hungary had also been integrated. The document summarizes the common conclusions based on all deliverables and meetings of the Data Sharing Expert Group, also targeting to compare the results of integrating EU-wide datasets on different pilot sites.

Acknowledgements

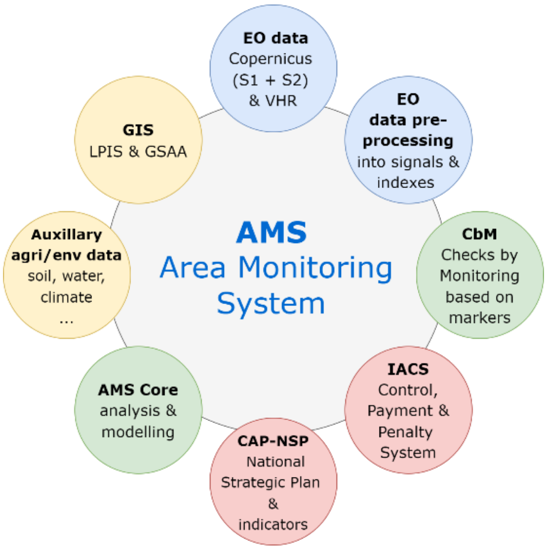
I am extending my sincere thanks to all the contributing colleagues who implemented the pilot studies in the last years, and who had participated on our workshops and brainstorming discussions. Special thanks to Katalin Tóth and colleagues from GTCAP team of EC- Joint Research Centre (JRC) for their valuable assistance.

Abstract

The main objective of the interoperability case studies for landscape features was to complete in 17 pilot sites of 4 countries the mapping of all potential landscape features linked to agricultural area, according to a harmonized and simplified feature categorization, based on phenology: woody, grassy, wet, and stony. The work is intended to address the future challenges of CAP reform, which aims to achieve higher environmental and climate objectives, also strengthening the system applied in the area of cross-compliance, which will include requirements in relation to new GAEC8. This latter also prescribes a minimum percentage of agricultural area devoted to non-productive areas or landscape features. This summary is focusing on the common lessons learnt during the pilot studies and offers the conclusions to be used by other member states. Detailed definition of landscape features, CAPI rules and comparison of the 3rd party external datasets analyzed are available in the individual reports of the pilot sites. This synthesis report consists of the following main parts: (1) Introduction (2) Challenges of the new CAP from year 2023, (3) Data collection, sematic analysis usability of third-party databases for updating and extending the landscape feature (LF) dataset, (4) Comparing 3rd party data to LF-LPIS digitization and (5) conclusions and a proposed solution.

1. Introduction and background

The environmental and climate objectives of the new Common Agricultural Policy (CAP) that shall be reflected in the strategic plans of the member states, which require specific indicators for monitoring and evaluating the implementation and the success of interventions. Parallel to this, member states have widely started implementing the Area Monitoring System, which brings a new logic of EO-data based systematic observation of agricultural activities throughout the entire year. These two initiatives together are mentioned as a paradigm shift in the Direct Payment management as they are changing the logic used in the last 2 decades: instead of focusing on controls and increasing the spatial details of eligibility inspection on parcel level, Paying Agencies now are willing to use primarily horizontal and highly automatized EO image processing methods, enlarging well-known statistical approaches to machine learning and AI direction. Paying Agencies are flagship users of recent novelties of cloud processing technologies and combined use of optical and radar sensors (Sentinel-1+2). New IoT technologies and the continuous processing of multi spectral and multi temporal EO-images offer a benefit of near real time monitoring and communication to the farmers. From this point the authority is not controlling the farmers but asking an active participation in an integrated monitoring and evaluation system. AMS data will also take part in scheme evaluation and future planning of CAP interventions. To be able to monitor the activities on the agricultural parcel, LPIS and GSA remained as core elements of the new system, ensuring a stable high resolution (M=1:5000) mapping component of IACS. The development of new image processing methods and the integrated panning approaches are slowly outlining new data inter-operability and data management challenges, such as:



- Using crop parcels of GSA as training data of image classification models. This approach requires preselection/data cleaning methods and crop grouping according to the crop phenomenon that can be detected on satellite images and a pre-validation of homogeneity of the declared parcels. The crop classification pilot project[i], several other projects (Sen4CAP, NIVA) and also national CbM and AMS implementations had created crop groups based on the GSA categorization. Experts agrees that a common EU-wide grouping could not be created with generalization, because the representation of location-specific crop occurrences are key contents of the training data.
- Using Copernicus High Resolution products – like Small Woody Features, Imperviousness etc. - to integrate land cover categories and to validate land cover changes based on the change of existing IACS geometries.
- Integrating thematic content of various 3rd party datasets into parcel-based modelling approaches such as soil and forest management data, with a multiple combination of geometrical extents.

Thanks to the INSPIRE infrastructure there is an increase of IACS data contents shared by several member states on an open and fair way. This fact itself boost the use of the land cover data and agricultural parcel data on crop level, maintained as part of IACS-GIS (LPIS and GSA), while parallelly AMS contractors are also trying to make use of the available GSA/LPIS data content to train the models deriving AMS markers. During the LF studies, our expert team had investigated how to integrate the content of external datasets into the LPIS.

In the recently past CAP period (2015-2022) the treatment of LFs was based on the cross-compliance (Regulation 1306/2013), as well as on practices beneficial to the environment (greening) defined in Regulation 1307/2013. Based on this, and following the LPIS-EFA guidance of EC (DS/CDP/2018/11 DSCG/2014/33 – draft REV 5) the following status of landscape feature data was the following:

1. Layer of potential LFs does not exists in all IACS-GIS implementations, only a rather limited number of declared GAEC/EFA LFs was digitized on ortho images and validated by the administration to become the part of a reference layer (example is Romania). The EU Regulations did not include the

obligation to digitize all potential features in LPIS, although the European Commission reported that, without digitizing these elements, it is not possible to effectively control their preservation.

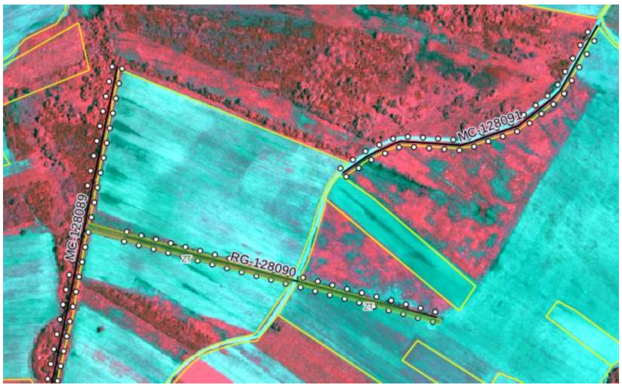


Figure 1-1: The RO pilot site: black line with white spots along represents LFs declared as line features in GSA-2021, where MC = filed margins, and RG = ditches. Yellow = physical block boundaries. M= 1:7000

2. All potential features of a given LF type are delineated on or directly adjacent to arable land, in harmonization with the LPIS reference parcels. It means that the polygon of LF is part of the LPIS reference parcel (RP), or it can form an individual RP unit. The LF always has a land cover class determined. This option is created via additional digitization activity beyond the eligible LPIS reference parcels, and usually available for GAEC LF types or buffer strips. This had been defined since 2009.

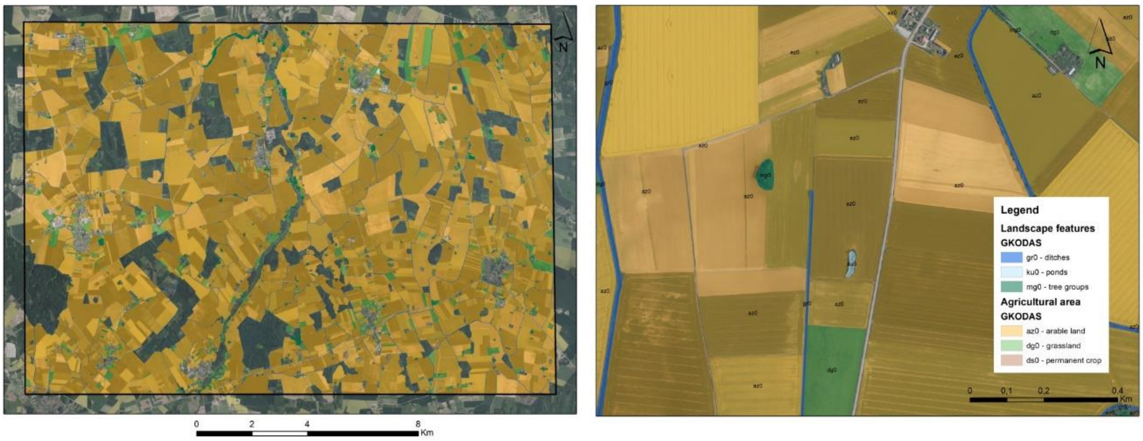


Figure 1-2: Study/test site example with existing LF and agricultural area in Lithuania. Yellow-shaded= LPIS physical blocks holding eligible area

3. The delineation and thematic selection of all potential features in several LF categories can be based on the LPIS data.This input data forms a primary base of selecting landscape features and buffer strips using certain land cover type, size, width and adjacency rules. This also requires the split of features according to the type definitions. This model is presented by the Hungarian LPIS.

Figure 1-3: The Hungarian example of categorizing LCC elements of a continuous LPIS layer for selection of LFs among non-eligible areas of LPIS physical blocks, based on different land cover categories, adjacency rules and average and min/max size limits [1]. Left image: Red lines: physical block boundaries, cyan spotted: non-eligible area, semi transparent shaded: different land cover categories inside the PBs. Right image: light green boundary: woody LFs, yellow numbers: PB IDs.

2. Challenges of the new CAP from year 2023

The concept of landscape features (LF) comprises the fragments of permanent non-productive areas embedded in agricultural landscapes. These small fragments have a key role in maintaining biodiversity and agro-ecosystem services, so they have become a priority focus of several EU policies. With the New Delivery Model concept in the design and management of the future CAP 2023 in mind, the main concerns are:

- Any area of the holding which is covered by LF subject to the retention obligation under GAEC standard 8 listed in Annex III are eligible for the intervention in the form of direct payments based on Article 4.4b, of Regulation (EU) 2021/2115 establishing rules on support for CAP Strategic Plans.
- Eco-schemes to be stablished by MS could cover LF or non-productive areas for the protection of biodiversity: creation, conservation or maintenance of habitats or species (Article 31.4.e of Regulation 2021/2115).
- For each reference parcel (RP) Member States shall record in the LPIS the type and location of LF that are stable in time on the parcel relevant for the eligibility of area-based interventions and for conditionality requirements (Article 2.7.d, of Commission Delegated Regulation (EU) 2022/1172 supplementing Regulation 2021/2115), as feature types or layers. Where applicable, also LF under GAEC standard 8 on non-productive areas or features (Article 2.7.e, of the same Commission Delegated Regulation 2022/1172) should be also localized and quantified.
- LPIS shall comply with delimited protected zones and designated areas, when LF under the good agricultural and environmental conditions are defined in accordance with the respective conditions (Article 66.3 of Regulation 2021/2116).
- The geo-spatial application (GSA) shall contain the type, location and, where applicable, size of LF relevant for conditionality or interventions (Article 8.3.c, of Implementing Act (EU) 2022/1173, laying down rules for the application of Regulation 2021/2116), to ensure effective support of farmers to declare different interventions.
- Creating and multi annually maintaining Performance and Monitoring Evaluation Framework (PMEF) indicators: R.34 "*Preserving landscape features*", R.31 "*Preserving habitats and species*", R.32 + R26 "*Investment related to biodiversity*" and SO6: *Contribution to halting and reversing biodiversity loss, enhance ecosystem services and preserve habitats and landscapes*. As the concept of LF is relatively abstract, which has only been recently endorsed by mainstream policy and science, there are no established quantification methods and indicators available yet, however such well-defined indicators, are critical for context and impact are laid out in the new CAP.

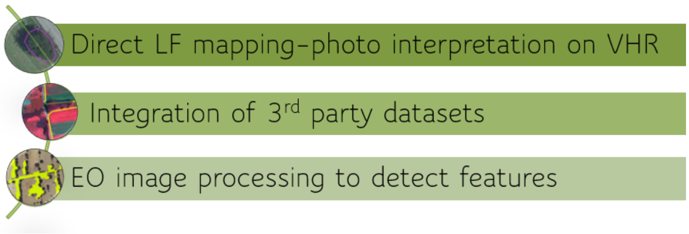
Landscape feature data able to detect and validate the change of agri-ecosystem services should cover the entire set of features on agricultural dominant landscapes. Mapping on EO-image source means that any vegetation – such as the LFs - are mapped according to the phenology and spatial extent, what does neither means that the given feature is under the disposal of a farmer, nor that the area of the feature is legally linked to the agricultural activity. On the other hand, a full coverage of features would allow the monitoring of changes, analyzing trends, would offer a set of potential features in IACS-GIS and would propose a base map to manage the retention of the features, also outside the direct payments.

3. Data collection, sematic analysis and usability of third-party databases for updating and extending the landscape feature (LF) dataset

3.1. Review of available dataset

The study found that there are many large and small scale national data containing LFs, which have been created along different specifications according to different purposes and use cases. Such datasets are usually managed by different sectoral bodies (land registry, forestry, nature conservation etc.) without pre-matching their semantics and technical specifications.

All pilot projects implemented in the 4 member states have analyzed the status of the GIS data owned by the PA and the possibility to integrate data managed by 3rd parties. The following activities have been completed by the 4 expert teams:



All IACS-GIS implementations hold high-resolution data in the member states, with outstanding quality and update frequency. In some of the member states LPIS reference parcels only covers the agricultural area – arable, permanent grassland and permanent crops – while in a few member states LPIS land cover is a continuous wall-to-wall coverage, including all eligible and non-eligible area. The latter concept, of course, offers extended functions to analyze the ecological network services linked to agricultural land. LPIS and GSA are sensitive datasets, with financial consequences, maintained according to exact rules and strict quality measures. In addition LPIS custodians are willing to keep all the responsibilities and risks related to the quality of the data.

Country	Number of test sites	Independent CAPI of LFs	Data Sorurce of CAPI - y2021	Use of LPIS RP boundaries	Copernicus -SWF	LUCAS	Use of functioning IACS EFA - layer	Number of national external datasets analyzed
CZ	9	YES	Aerial Orthoimagery	YES	YES	not available	YES	4
ES	4	YES	Aerial Orthoimagery+ Deimos	YES	YES	YES	YES	3
LT	1	YES	Aerial Orthoimagery	YES	YES	NO	YES	4
RO	3	YES	Worldview-2	YES	YES	not available	YES	0

Table 3-1: Datasets analyzed on the pilot sites

General findings of reviewing auxiliary datasets:

- Semantic anaylsis: the feature types describing LF have been defined according to national specificities.
- Datasets are created at differrent scales and with different methods of data collection, unsually more detailed spatial resolution brings more detailed classification.
- None of the national 3rd party dataset has as the specific objective of identifying landscape elements at the required level of detail as of LPIS has. Landscape features are secondary aspects or derived from the main objective of each data source.
- Pan-European datasets are more generalized while national datasets are more detailed.
- The datasets differ mainly in the method of data collection, but also in the mapping scale at which the datasets are produced. Some datasets are created by automated or semi-automated classification of aerial/satellite imagery while others are created based on visual interpretation of aerial/satellite imagery and also via field measurement.
- SWF dataset had to be transformed into the national coordinate system and the available software did not allow an accurate transformation (with an error of 1 m), which in some places, especially in mountain areas, caused a shift of Landscape Features from the reality.
- LF delineation from raster sources is fuzzy by default.
- Vector sources of information better delimit LF elements. Neverthless, the photo interpretation of the under canopy land cover is challenging, what is fundamental in case of woody features.
- Topographical map of water bodies was very useful to distingush wet features from woody ones.
- Data interoperability that backed the large-scale data fusion (data mashup) may invoke data topology problems that are not present in the original databases. For example, the accuracy of data vectorization can result heterogeneous LF boundaries (e.g., when spatial data from different data

sources is intersected). Combining large scale data for identification of LF might become an increasingly common problem.

Only a limited part of the analyzed 3rd party data was conformant to the accuracy of the LPIS (i.e. 1:5000 scale). The general outcome of the analysis is, that such **3rd party datasets shall be primarily integrated as background data for the extension and update of the IACS-GIS/LPIS/LF dataset and not as data that will directly replace the current human CAPI-based method of feature delineation**. The use of an external data by LPIS custodians to exact feature delineation is still linked to individual visual verification, to ensure the quality of data integration with sub meter precision, as required on 1:5000 mapping scale. 3rd party data content is very valuable to detect changes what can be used to spatially direct or balance the LPIS update processes, to assess risks of the system. However, if data in scale larger than 1:5000 mapping is available, it can be a valuable source of LF features, as the experience of the Netherlands shows.

1.1. Semantic mapping between the proposed broad categories of LF, the existing GAEC/EFA-LF categorization and the corresponding feature types of the created LF layer

As part of the spatial analysis, a landscape feature dataset was created with direct mapping (CAPI) on VHR ortho images by the national expert teams. This dataset was taken as a reference, i.e. one that reflects the latest situation of Landscape Features accepted as ground truth.. For the dataset itself, the following four functional LF (FLF) classes were used focusing only on ecologically relevant distinctions between features. They represent different broad vegetation and ecological network types. Functional LF typologies tested on the pilots were proposed by the recently published "*Landscape features in the EU Member States"* technical report of the JRC [2]. Using the categories of the above-mentioned study and based on the current mapping experiences and on digitizing GAEC/EFA LFs in different EU countries the following definitions are proposed:

Woody features (A): Linear or island-like natural and semi-natural individual biotope features covered dominantly by arboreal (perennial woody) vegetation, such as trees and bushes, can be natural or planted, but never integrated part of a larger forest vegetation.

Grassy features (G): Individual patches or linear areas dominantly covered (> 50%) by permanent herbaceous vegetation, where permanent grassland dominates, embedded in an agricultural landscape, can be natural or planted, but never integrated part of a larger permanent grassland vegetation.

Wet features (W): Natural or human made linear or island-like individual features, where the ecosystem – such the phenology of the vegetation – is dominantly determined by the permanent or regular presence of water.

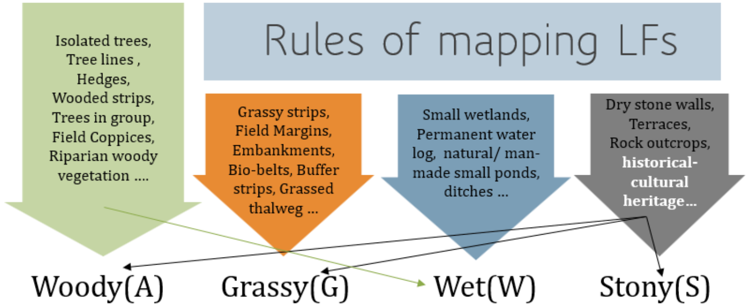
Stony features (S): Natural or human-made appearance of individual stone features or the presence of rock layer or non-productive soil fragment on the land surface.

The categories defined according to the four functional simplified LF classes and the definitions used for digitization of features on VHR images are summarized in the following tables.

Definition

LF type (MS specific) detailed category - features and maintenance rules are defined

LF class (EU-common): simple **functional typology** by phenology (Bálint Czúcz et.al. – 2022)



Def. of agricultural landscape: LCC adjacency	AL+PG+PC: inner features + partly located within a boundary buffer of 20 m, including directly secondary LFs + **distinguishable from the neighboring fields and natural ecosystems** - Understanding the landscape structure is fundamental.
Size limits: min, max. width, area - to comply with other regulations	Max 5000 m2 and **approximate average width** of 20 m or more if suitable, focus on habitat spatial structure: skip the old rules of 2 meters gaps – only consider real artificial sealed elements Dominancy (>50%) of grassy or woody vegetation of the entire feature – not to cut a linear feature to small parts
What not: excluded as a LF:	LFs fully on yards and recreational areas, narrow PG parcels

Figure 3-1: Summarizing the 4 main simplified LF categories and the main rules of the detailed definitions.

3.2. Common solutions regarding definition of landscape features

During the actual vectorization of the LF dataset, it was found that it is quite difficult to distinguish some LF types from each other (Group of trees, Isolated tree or even Hedge, Riparian, Woody strip and Group of trees). For this reason, all the expert team proved that it would be more appropriate to simplify the LF-EFA classification to only four main groups of Landscape Features as proposed by the JRC (Woody, Grassy, Wet and Stony Features), which would reduce the number of possible mistakes made by the operators determining the specific LF type.

To be able to delineate the functional features of the agri-landscape on a logical way, the following main aspects have been commonly identified, what has a direct relation to the definitions and to the mapping. Solutions are also proposed based on the experiences of the pilot studies:

- A LF should be **independently identifiable** “by it is own extent”, holding a distinct physical appearance in relation to the surrounding landscape during a significant period of the year. They should never be an organic part of a neighboring larger natural vegetation. Typical example of such challenge is, when a wooded strip along a ditch is bordering the parcel and a natural wetland, where similar groups of arboreal vegetation are spread. In this case the wooded strip has to be identifiable as a distinct feature from the side of the natural wetland as well. I it is not the case, the trees will form part of the wetland and there will be no LF identified.
- Width can only be exact if the actual land cover boundary is captured as the **limit of the feature**, excluding the canopy extent overlapping a neighboring land cover. This requires a canopy independent delineation method, in most cases visually estimating (CAPI) the real under canopy boundary.
- An **average width** of a feature is suitable to determine the limit of maximum extent. Natural features might vary in their width, even extending the predefined maximum of 20 meters. This should be taken into consideration with appropriate flexibility, not to exclude any of the meaningful linear ecological network elements and to avoid geometrical separation of a functionally continuous feature. The average width should be defined according to the typical nature of a certain linear feature type, taking into consideration the local landscape structure and historical background. It can even lead to different rules per zones.
- The **dominant woody or grassy land cover phenomenon (over 50% share)** in relation to the spatial extent of a given LF should be taken into account. This method itself determines the decision on

the extent/length of the feature. The borderlines of the categorization of features are: (1) the dominant extent of canopy cover of arboreal vegetation, and (2) the decision on the limit of an individual object. While deciding the dominant extent with CAPI the human decision is able to take into account the annual growth/invasion capacity of the local bush and tree species in relation to the update cycle and the human contribution to the maintenance (potential cut of edges). While off-leave winter images support the under canopy interpretation, it would lead to under estimation of leaf fall vegetation. The relative homogeneity of the woody vegetation spread will determine the extent of a feature as an independently identifiable single unit. In case of linear features, a minimum length of a feature is also recommended to be defined (30-50 meters) to avoid meaningless splitting of small units.

- Not all types of LF can be identified by photo-interpretation on all land use categories, e.g.: trees in PC parcels brings risk of incorrect identification of cultivated trees as LF.
- **Dominancy among feature types,** in case multiple presence of categories in a single feature is detected, the order is based on the strength of ecosystem contribution:
    1. wet (in case of continuous presence throughout the year)
    2. woody
    3. grassy
    4. stony nature.

  It means that the part followed by trees will remain a wet feature even though the water body is covered by canopy and in case of a natural habitat where bare rocks and grass species are in mosaiced pattern, the dominancy of the grass will define the category.

Distinction of 100% woody and 100% grassy features from those, where dominant share is coded will contribute to better implementation of:

- to run image classification/machine learning algorithms for automatized validation of the features,
- prioritization of update procedure, focusing CAPI on primarily on dynamically changing elements, while the more stable ones can be validated with EO-image analyzing,
- monitoring the encroachment of woody vegetation.

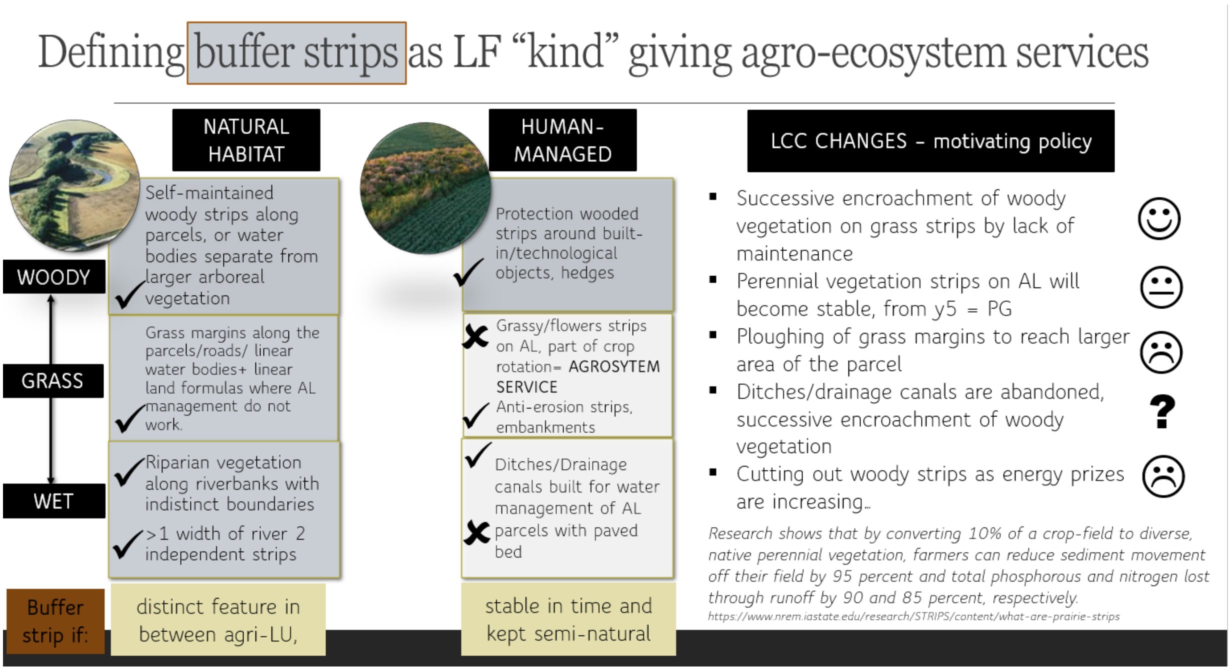


Figure 3-2 Categorization of buffer strips

It hardly matters neither for biodiversity, nor for ecosystem services if there is a gap under 5 meters somewhere in the regular pattern of trees, or if a feature of irregular shape consisting of a mixture of shrubs and grass. What really matters is the total extent and the spatial network of the features, the “connection” to the surrounding semi-natural larger spots and to agricultural fields. Both can be reasonably well limited with simple land cover classes, and width and area measures for features embedded among agricultural fields.

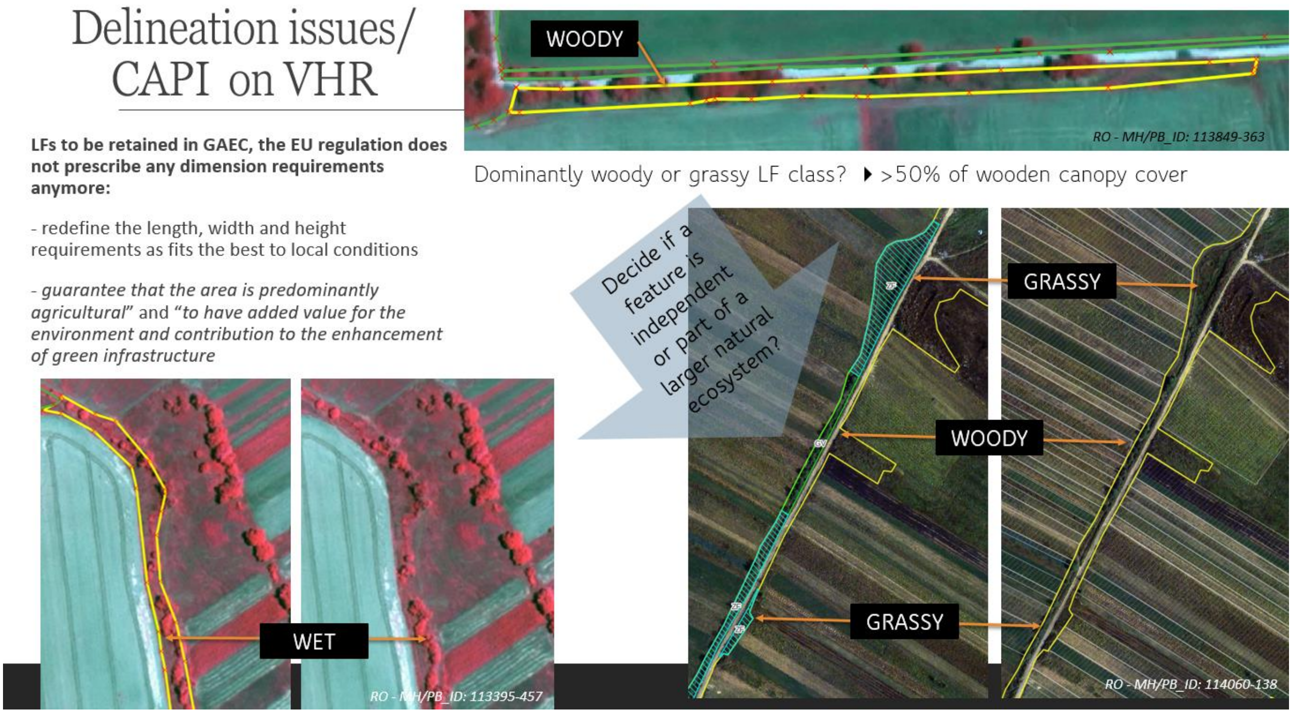


Figure 3-3 Examples of buffer strips categorized as LFs or not, according to the average width, along running water bodies on AL landscape

During the spatial analysis it was found out that some types of Landscape Features can be considered more as Landscape Features at the boundaries of the reference parcels (Hedge, Riparian woody vegetation, Group of trees, Woody strips), i.e. they are more linear type features. On the other hand, features that represent small areas (represented by a point in some databases) are more likely to be located within the reference parcel. It was also found out that in all datasets there are more Landscape Features present in permanent grassland than in other agricultural land use types, while in the greening period of CAP 2015-2022 only features on arable land had been delineated. At the same time, it should be noted that test sites with higher elevation contain more Landscape Features than test sites located in lowland or mid-elevation areas, where the vulnerability of the features is anyway more critical.

Our study analyzing the status of LFs mapping in several EU-member states shows why the direct mapping on a VHR image is a safe method for the Paying Agencies. The practice of the last 10 years also pushed this technological solution. The current status is, that member states were selecting among the possible LF types. Therefore, not all the types of features occurring in a reference parcel were systematically delineated. On the other hand, it must be admitted, that the initial extension of the LPIS layers with all the directly adjacent LFs requires significant resources. Based on the experiences to complete the delineation of all potential features in a 20 meters buffer of all agricultural area would require approximately 1 year full time work of 21 person for the agriculture area of 5 000 000 ha (only the CAPI team, excluding organization and IQC activity). Our calculation holds that in average 1,7 minute is spent per feature during a 6 hours of active CAPI per man-day.

4. Comparing 3rd party data to LF-LPIS digitization

The best result of 3rd party data integration of course can be reached with datasets created by photo-interpretation on VHR images, especially of woody features, as they are all relatively permanent and well-distinguishable from the neighboring land cover. In this respect the area overlap (intersect) of the 2 datasets shows the correct identification of the SWF existence, what determines the type of the feature as woody, but not the correct delineation of the feature's entire extent. This means, that the accuracy measures based on feature overlaps will describe the identification of the woody vegetation inside a given feature, this is called as "matching features", while the accuracy of the boundary delineation of the feature itself cannot be derived from the overlap analysis.

Errors of commission, omission, semantic and geometric were identified while comparing the 3rd party LF data to the digitized reference LF layer. The best match is with datasets created by photo-interpretation on VHR images, while the lack of thematic inter-operability significantly decreases the result. Difficulties leading to poor data integration result:

- Matching of semantic classification – level of depth, level of generalization
- Differences in interpretation rules: width/independency/phenology, like woody or wet feature?
- Delineation: false negative (omission)/false positive error in spatial extent
- Geometrical shift

Matching share of LFs delineated during the pilot studies and of existing LF-EFA/LPIS and 3rd party LF datasets had been analyzed with GIS overlaps. "Matching %" is the ratio of features also captured in the 3rd party dataset, while "Missing % " means the omission of features compared to the reference LF dataset delineated during the pilot studies. The overall rate of matching is very low, hardly reaching the 30%, while in average 68% of the features are missing compared to the amount digitized on ortho-photographs.

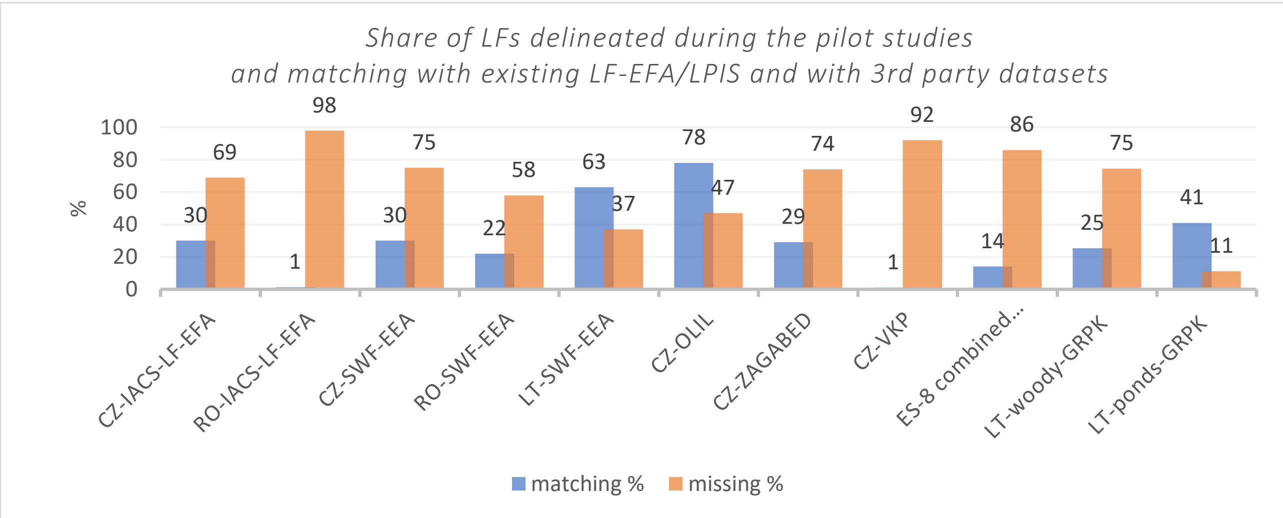


Figure 4-1: Example of matching features of the Czech 3rd party datasets: Zagabed = Land Survey Office,VKP = Significant Landscape Features-Ministry of Environment, OLIL= dataset of wood cover - Forest Management Institute. The LT GRPK is a Georeferential Base Map, created by direct mapping from aerial orthoimagery.

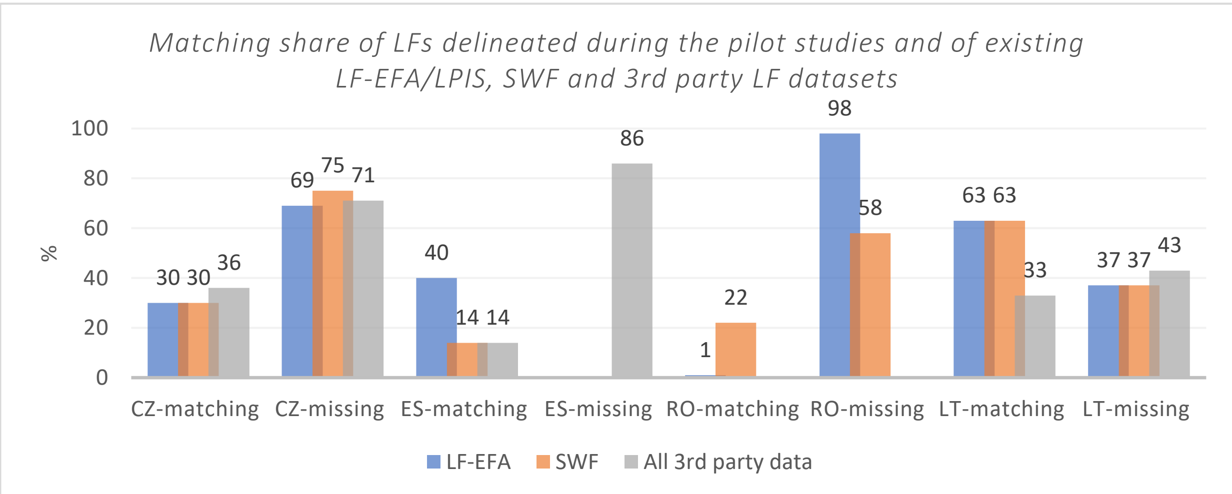


Figure 4-2: Matching share of LFs delineated during the pilot studies and of existing LF-EFA/LPIS, SWF and 3rd party LF datasets.

The numeric results are aligned with the visual impression that the SWF data is not suitable for direct LF delineation along exiting IACS rules. The accuracy measures reflects that the SWF data is not usable to delineate functional features of the ecological network on 1:5000 scale. SWF data is generally suitable to detect the presence of woody vegetation, while some narrower linear strips surrounded by AL seems systematically missing. One of the main reasons is that for LF delineation a functional definition of a continuous habitat is implemented primarily, and the share of woody/grassy vegetation inside the feature is a 2nd level attribute. During LF CAPI, features surrounded by other land uses and visible boundaries are delineated and classified based on the vegetation phenomenon (woody/grassy/dominancy), ecological conditions (wet) and size limits. Different size limits will determine the potential function of the feature inside the ecological network: linear elements, patchy island elements etc. Thus, the content of the SWF is just the arboreal (woody) vegetation type, what can contribute to a LF as the dominant vegetation category or also as a minor type.

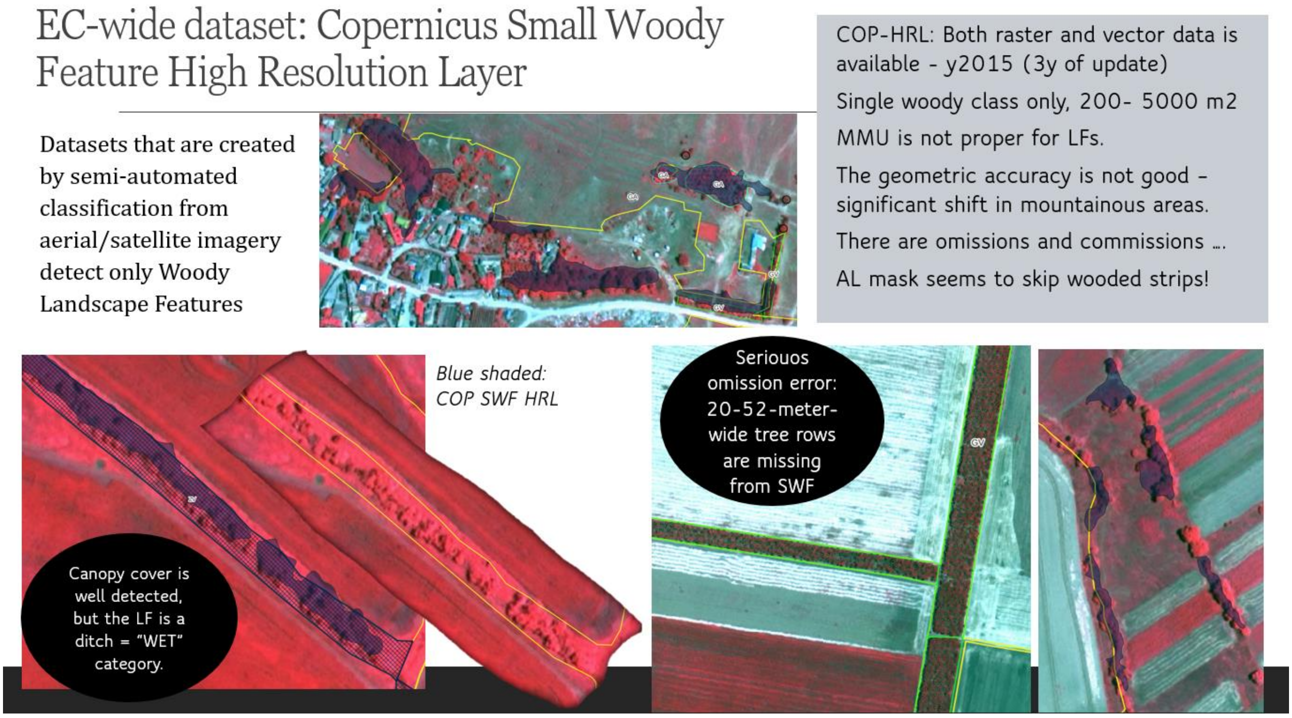


Figure 4-3: Romanian example of analyzing the use of Copernicus Small Woody Feature dataset

The Spanish study also strengthens, that the automatic LF obtaining assists in the identification but does not replace their vectorial delimitation nor all its location. In order to achieve an automatic delimitation with good geometric precision of the LF, there should be more specific, automatable and higher resolution data sources, which do not yet exist at the national or continental level.

Copernicus SWF data can be integrated at the following data development and data management steps:

a. Validating the encroachment of arboreal (woody) vegetation on the area of the already delineated features, and updating the woody feature categorization according to the dominancy rule.
b. Automatic verification of existence and enlargements of woody features, and only assigning a woody feature to CAPI in case there is any change of the extent is foreseen.

5. Conclusion and proposed solution

The following main conclusions are found as a result of the pilot exercises:

- **The broad categories of LFs types (woody/grassy/wet/stony)**, as proposed by the JRC, **are fully adequate, and presents enough details to detect the changes of the agroecosystem services**, while reduces the burden on sophisticated feature categorization.
- 3rd party datasets shall be primarily used as background data for the extension and update of the LF-EFA dataset **not as data that will replace the current method of creating the LF-EFA dataset,** as visual inspection (CAPI on VHR) is necessary**.**
- The LF should be defined based on it's entire extent, with categorization of land cover dominancy and average feature width.
- To store the share of woody/grassy information on feature level is valuable for updates based on EO-image sources. Although the 1:5000 scale on-screen visual mapping of LFs cannot be substituted the thematic categorization can be based on ML-derived EO products.
- The best approach is to create a geometrically fully suitable LF layer as part of the LPIS data, extending the LFs adjacent to AL/PG/PC parcels (taking into consideration some limitations on PG and PC), as this supports the management of elements in IACS and PMEF monitoring solutions at the same time.
- Regional pattern of the ecological network elements could also be evaluated using different spatial biodiversity indexes. Such index could indicate the critical limits of feature loss and become the base of planting new features.
- In case the full coverage of potential LFs is not chosen to be maintained, technically the presence of woody features can be monitored by the Copernicus HRLs based on the agricultural parcels adjacency and size, taking into consideration the serious omission of smaller features in agri-dominant zones and the lack of annual data updates.
- If a full LF coverage exists, it would be essential to **harmonize the retention requirements among the different policies**, building on the logic of cross-compliance. It is evident that CAP cannot handle the proper maintenance rules of all LFs as some are under the forestry law, some features are the result of spontaneous spread of arboreal vegetation and some are handled by owners or users that have no relation to CAP. To prohibit a clear cut of woody vegetation and to require sustainable forest management is one of the easiest rule to integrate equally by the different policy branches. Requirements of buffer strips along built-in objects and linear infrastructure is again a well-regulated area of spatial planning. Preservation of wetlands can be fully covered by the CAP rules, where arable land ploughing holds the highest risk. This short list already highlights the beneficiaries and users of a complete and continuous LF GIS data.

LF vector and raster data representation, data completeness, semantic definition and classification of the relevant feature types of the 3rd party datasets as well as the level of details (spatial resolution and the granularity of thematic information) may not be sufficient and thus, may not be fully compatible with the concepts recognized in IACS-GIS/LPIS. There seems to be 2 main logical directions to use any data, including those derived by image classification (like the SWF) for new CAP implementation and monitoring, which are not fully independent from each other:

2. The bottom-up approach: **Keeping the administration of LFs as part of IACS-GIS,** developing the data on VHR image sources, continuing to fulfil the 1:5000 mapping scale and 100 m2/2m of linear minimum mapping unit (MMU) **as required in LPIS**. In this case meaningful monitoring how the agri-

related ecosystem functions and biodiversity is changing will only be possible if all the LFs related to agricultural land are digitally mapped. A positive effect is, that complete LF mapping itself strongly supports the conservation of existing LFs (see the Hungarian example of wall-to-wall LPIS), independently from the fact if the feature is disponible and declared by the farmer or not. Further advantage is, that already mapped features can be offered for the farmers in GSA avoiding LFs validation under a campaign year before the payments. This concept will ensure the adequate input data to derive indicators for CAP to monitor ecosystem services (SO6), preserving landscape features, habitats and species (R34, R.31). This solution is evaluated as the most trustable regarding the data quality and PA-driven data governance, as it offers the highest support for the farmers in the declaration process.

3. The top-down approach: **Deriving indicators for the CAP monitoring system on the base of** other continuous datasets than IACS and LPIS. Most likely the requirements of the refresh time will determine **remote sensing-based datasets**, where the annual or 3-4 yearly update can be managed with the same image classification method, ensuring the type and spread of errors to be similar. A possible solution is either to run dedicated EO data processing models, or to use an EO-seamless external data, like the Copernicus SWF data. The weak point of such external (IACS-independent) dataset is the minimum mapping unit (MMU), where features on agricultural landscape < 300 m2 are not entirely detected, while these forms the most vulnerable elements of an agri-ecosystem. This can be solved with the fusion of higher resolution images than Sentinel2 (see the Planet Fusion approach as an example). Apart from the known weaknesses, analyzing the change of SWF in relation of LCC and land use combinations stored in LPIS is possible and would give clear indication of trends. Automating LF identification process requires additional: (i) LF semantic mapping (ii) data specification mapping [5] (iii) data matching and (iv) data fusion algorithm development efforts. The advantage is that the developed algorithms can be shared and used by other users, especially in those cases when non-national data is used.

The capability of the ecological network is a serious agri-environment issue. The spatial indicators of enhancing ecosystem services, preserving habitats and landscapes should be derived continuously to allow annual comparison and trend analysis. The combination of the two approaches seems to bring the fastest solution. To detect the loss, the analysis must focus on the existence of the features primarily, while the accurate delineation has more importance for IACS implementation. To evaluate the existence of the small (under 300 m2) spotted features a point representation in GIS is already a valuable data.

While the IACS-related quality and the clearance aspects would support the slowly-growing bottom-up approach, **the increased loss of features in intensive agri-dominant regions would require a fast and adequate policy answer**. The network of larger natural/semi natural vegetation spots, the linear and spotted LFs practically determine the spatial distribution of the habitats. This is strongly a region-dependent issue, thus main typologies of landscapes with different problems and different interventions could be defined. It is pointed as an important direction of further studies. For example, CAP interventions could have significantly higher effect with increased sensitivity towards regional and specific aspects of these problem. The proposed method of collecting LF monitoring data, using the 4 broad LF categories as proven by the current study, is the following:

1. Using all LFs already delineated in LPIS by GSA progress or systematically as part of GAEC/EFA, mapped to the 4 broad LF categories.
2. Determining the woody/grassy/wet/stony nature of all inner non-eligible elements delineated already in the LPIS reference parcels with image classification and with the data integration of EO-thematic HRLs.
3. Integrating additional 3rd party national data content in 20 meters buffer of agricultural area, excluding permanent crops and permanent grasslands with scattered features. Deriving generalized boundaries if features are available as point or line, with the same typology as used in point 2, verified by the same EO data content.
4. Adding small spotted features under 300 m2 with a “generalized extent” from EO data image analysis, preferably integrating VHR/orthophotographs into the image processing.

Features in point 1 will have the most correct spatial and thematic accuracy. That is why this data is directly used in IACS and is fully proven and updated as a reference. Data of point 2 has the same level of spatial accuracy but is not a fully proven thematic grouping: less error in identifying wooded vegetation, while the mismatch of other features is high, but the type of semi-natural/natural vegetation is fully proven. Data in points 3-4 will only highlight the existence of the feature, but that is enough to monitor the significant changes in a region. The Spanish pilot showed that handling the entire data as a raster is a good approach, which allows an easy to manage annual comparison. By this method the quality of data can be increased on a sensitive way from year to year, also data quality measures can be assigned time independently - omission and commission errors as output of EO-based validation can be determined at any later stage. A seamless data would allow the categorization of regions along any well-known biotope indexes what would lead to prioritizing policy interventions.

This would mean, that the data used for monitoring the existence of the features would be different than the LPIS data used to implement the scheme, up to the point the entire set of features will be mapped in LPIS. The approach is proposed because the systematic long term data integration of different sources of different data quality would still be able to highlight the critical regions of loss and the efficiency of feature restoration, what enable fast policy reaction for a highly critical question.

References

[1] *Bernadett Csonka, Managing EFA declarations and the development of LPIS in year 2015. GAEC WS - Prague Source:* https://marswiki.jrc.ec.europa.eu/wikicap/images/d/dd/GAECWS_Prague2015_HUCSD_v04.pdf

[2] Czúcz B, Baruth B, Angileri V, Prieto Lopez A, Terres JM,: Landscape features in the EU Member States: A review of existing data and approaches. EUR 31063 EN, Publications Office of the European Union, Luxembourg, 2022, ISBN 978-92-76-52324-6, doi:10.2760/101979, JRC128876.

Technical Guidelines on IACS spatial data sharing. Part 1. Data discovery (2020). https://marswiki.jrc.ec.europa.eu/wikicap/images/f/f8/DS-CDP-2019-04_REV2_TechnicalGuideline_IacsSpatialDataSharing_Part1_Final.pdf

Methodological report on resolving interoperability issues in reusing IACS data in LULUCF https://marswiki.jrc.ec.europa.eu/wikicap/index.php/File:Methodological_Report_on_Resolving_Interoperability_in_Reusing_IACS_Data_in_LULUCF.pdf

Data Interoperability Pilot: Analyzing the usability of generalized GSAA data in context of crop classification https://marswiki.jrc.ec.europa.eu/wikicap/index.php/File:Classif_pilot_Report_JRC_CsD_final.pdf

The INSPIRE data models (https://inspire.ec.europa.eu/Data-Models/Data-Specifications/2892)

The INSPIRE data specifications (https://inspire.ec.europa.eu/dataspecifications/2892

European Environment Agency Copernicus Land Monitoring Service – High Resolution Layer Small Woody Features – 2015 reference year Product Specifications & User Guidelines, Tobias LANGANKE, version 3.7 of 2020-02-26 https://land.copernicus.eu/user-corner/technical-library/hrl_lot5_d5-1_product-specification-document_i3-4_public-1.pdf

IMAP stands for ***I****ntegrated* ***M****odelling platform for* ***A****gro-economic and resource* ***P****olicy analysis*. With IMAP the Joint Research Centre provides scientific support and tools to DG Agriculture and Rural Development for implementation, monitoring and evaluation of the CAP - https://wikis.ec.europa.eu/display/IMAP/Landscape+features_GENERAL

GUIDANCE DOCUMENT ON THE LAND PARCEL IDENTIFICATION SYSTEM (LPIS) UNDER ART. 5, 9 AND 10 OF COMMISSION DELEGATED REGULATION (EU) NO 640/2014 AND ON THE ESTABLISHMENT OF THE EFA-LAYER REFERRED TO IN ART. 70(2) OF REGULATION (EU) NO 1306/2013 CLAIM YEAR 2018 ONWARDS https://marswiki.jrc.ec.europa.eu/wikicap/images/a/ae/DS-CDP-2018-11-LPIS-guidelines2018-clean.pdf

Amy H. Pickens, Matthew C. Hansen, Matthew Hancher, Stephen V. Stehman, Alexandra Tyukavina, Peter Potapov, Byron Marroquin, Zainab Sherani, Mapping and sampling to characterize global inland water dynamics from 1999 to 2018 with full Landsat time-series, Remote Sensing of Environment, Volume 243, 2020, 111792, ISSN 0034-4257, https://doi.org/10.1016/j.rse.2020.111792 (https://www.sciencedirect.com/science/article/pii/S0034425720301620)

List of Abbreviations and definitions

AL	arable land
AMS	Area Monitoring System
APIA	Agenției de Plăți și Intervenție pentru Agricultură, Agricultural and Intervention Agency - the Romanian Paying Agency:
CAP	Common Agricultural Policy
CAPI	Computer Aided Photo Interpretation
CbM	Checks by monitoring
EC	European Commission
EO	Earth Observation
FLF	functional LF classes
GIS	Geographical Information System
GSA	Geospatial Application (used as GSAA before 2022)
HRL	High Resolution Layer
IACS	Integrated Administration and Control System of direct payments
IoT	Internet of things
JRC	Joint Research Centre
LCC	Land Cover Class
LPIS	Land Parcel Identification System
LUCAS	Land Use/Cover Area Frame Survey
LU	land use
LF-EFA	Landscape Features categorised as Ecological Focus Area (CAP requirement from 2015-2022)
LF	Landscape Features
MMU	Minimum mapping unit
PA	Paying Agency
PB	physical block = reference parcel type of LPIS
PC	Permanent crop
PG	Permanent grassland
QC	Quality Control
RO	Romania/Romanian
RP	Reference Parcel of LPIS
SWF	Small Woody Features
VHR	Very High Resolution satellite images, with sub-meter spatial resolution

List of figures

Figure 1-1: The RO pilot site: black line with white spots along represents LFs declared as line features in GSA-2021, where MC = filed margins, and RG = ditches. Yellow = physical block boundaries. M= 1:7000 5

Figure 1-2: Study/test site example with existing LF and agricultural area in Lithuania. Yellow-shaded= LPIS physical blocks holding eligible area 5

Figure 1-3: The Hungarian example of categorizing LCC elements of a continuous LPIS layer for selection of LFs among non-eligible areas of LPIS physical blocks, based on different land cover categories, adjacency rules and average and min/max size limits [1]. 5

Figure 3-1: Summarizing the 4 main simplified LF categories and the main rules of the detailed definitions. 9

Figure 3-2 Categorization of buffer strips 10

Figure 3-3 Examples of buffer strips categorized as LFs or not, according to the average width, along running water bodies on AL landscape 11

Figure 4-1: Example of matching features of the Czech 3rd party datasets: Zagabed = Land Survey Office,VKP = Significant Landscape Features-Ministry of Environment, OLIL= dataset of wood cover - Forest Management Institute. The LT GRPK is a Georeferential Base Map, created by direct mapping from aerial orthoimagery. 12

Figure 4-2: Matching share of LFs delineated during the pilot studies and of existing LF-EFA/LPIS, SWF and 3rd party LF datasets. 12

Figure 4-3: Romanian example of analyzing the use of Copernicus Small Woody Feature dataset 13

List of tables

Table 3-1: Datasets analyzed on the pilot sites 7