ARPHA Preprints

PENSOFT

Guidelines

Author-formatted document posted on 14/02/2023

Published in a RIO article collection by decision of the collection editors.

DOI: https://doi.org/10.3897/arphapreprints.e101949

Essential Biodiversity Variable workflows: designing the freshwater, marine and terrestrial EBV workflows from data collection to modeling

W. Daniel Kissling, Maria Lumbierres

ARPHA Preprints *Author-formatted document posted on 14/02/2023.* DOI: https://doi.org/10.3897/arphapreprints.e101949

EUROPABON

Essential Biodiversity Variable workflows

Designing the freshwater, marine and terrestrial EBV workflows from data collection to modeling

W. Daniel Kissling & Maria Lumbierres

Institute for Biodiversity and Ecosystem Dynamics (IBED), University of Amsterdam, P.O. Box 94240, 1090, GE, Amsterdam, the Netherlands

9. February 2023

This document serves as an input for the EuropaBON virtual workshop on EBV workflows (22–24 February 2023), with some examples of EBV workflows for integrating data streams.

Contents

ARPHA Preprints *Author-formatted document posted on 14/02/2023.* DOI: https://doi.org/10.3897/arphapreprints.e101949

Background

The EuropaBON project (https://europabon.org/) aims to co-design a European Biodiversity Observation Network with the goal to close the gap between the biodiversity data needs of policy-makers on the one hand and the existing reporting streams and data sources on the other hand. In this context, Essential Biodiversity Variables (EBVs) are used by EuropaBON as a core concept to integrate different reporting streams, data sources, and monitoring schemes, and to measure biodiversity change across multiple dimensions in space and time. Through an extensive methodology (incl. rapid surveys, stakeholder and expert workshops, standardized surveys, semi-structured interviews and internal reviews), EuropaBON (with its deliverable D4.1, revised version 31/01/2023) has identified a list of EBVs for a European-wide biodiversity observation network. This list contains 70 EBVs across the terrestrial, freshwater, and marine realms, covering all six EBV classes (Figure 1).

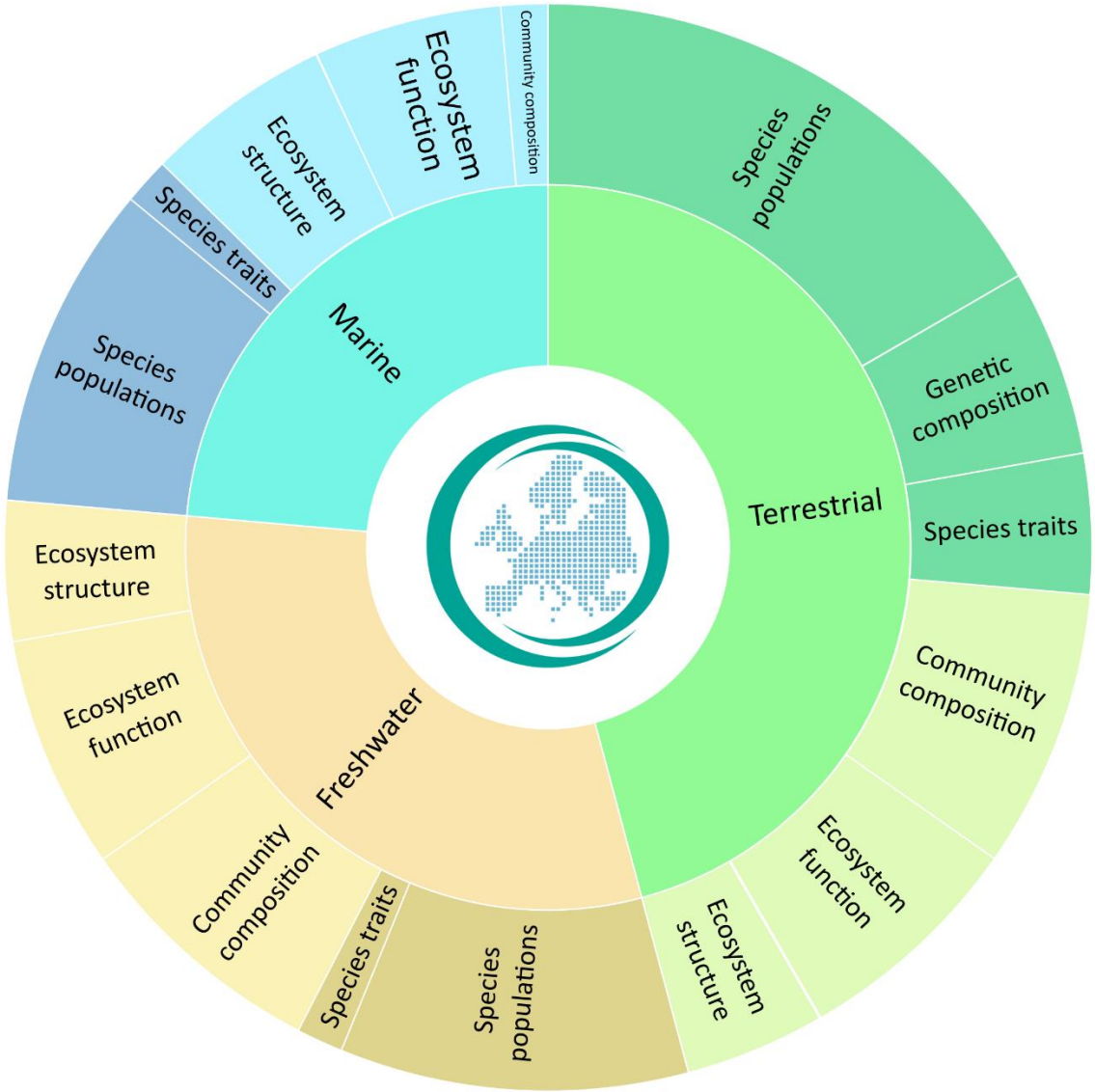

Figure 1: Distribution of 70 identified Essential Biodiversity Variables (EBVs) within terrestrial, freshwater, and marine realms (inner ring) and whether belonging to EBV classes with species (darker colors) or ecosystem focus (lighter colors) (outer ring). A total of 32 EBVs fall into the terrestrial realm, and 20 and 18 variables into the freshwater and marine realm, respectively. The EBV classes follow the classification of GEO BON (https://geobon.org/ebvs/what-are-ebvs/). Source of figure: EuropaBON deliverable D4.1 (Junker et al. 2023, revised version submitted on 31/01/2023).

This project receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101003553.

ARPHA Preprints *Author-formatted document posted on 14/02/2023.* DOI: https://doi.org/10.3897/arphapreprints.e101949

The identified EBVs cover various taxonomic groups and habitats to be monitored (Figure 2). It includes several vertebrate groups, especially birds and mammals, but also fish, reptiles, and amphibians (Figure 2a). It also covers various invertebrate groups (Figure 2b) and different plant groups (Figure 2c). Finally, a range of different habitat types across freshwater, marine and terrestrial ecosystems are represented in the EBV list (Figure 2d).

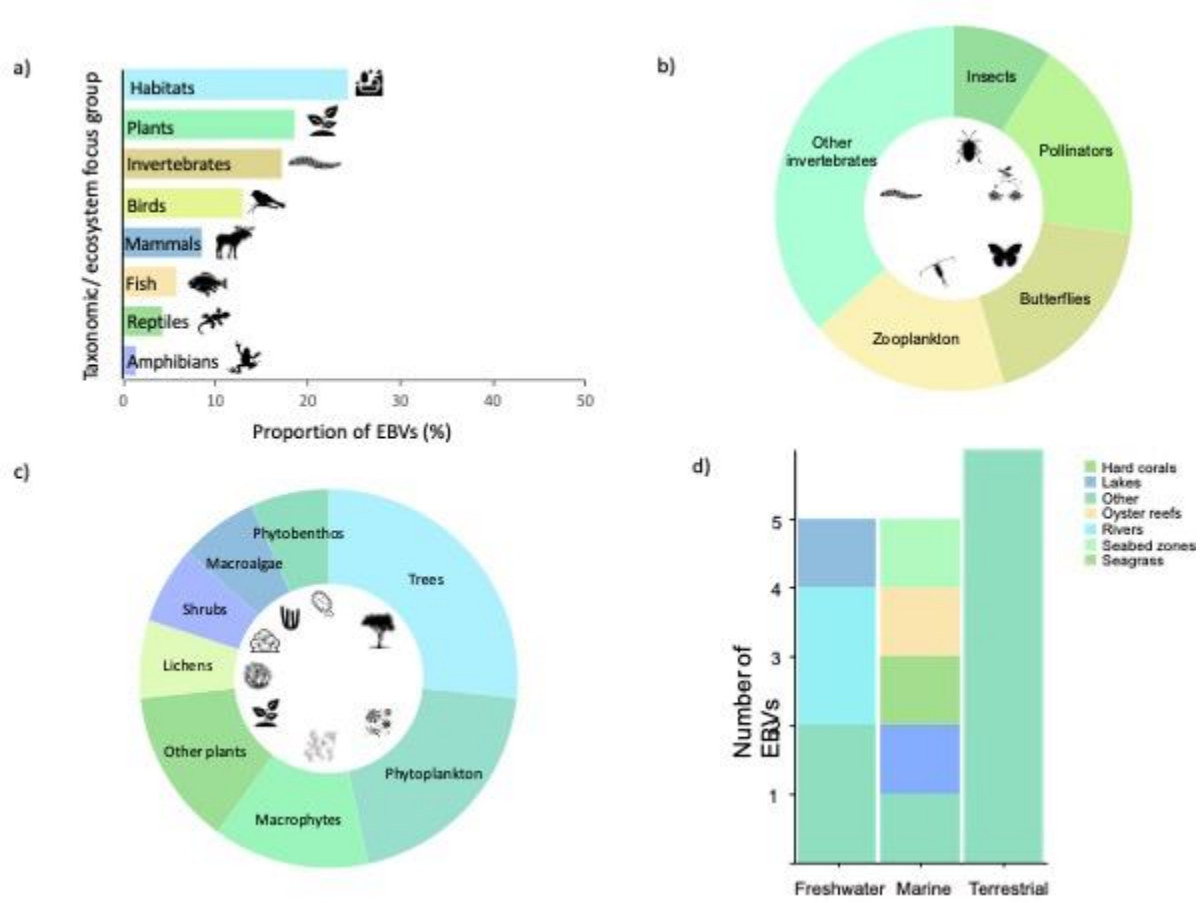

Figure 2: Representation of taxonomic groups and ecosystems in the list of 70 Essential Biodiversity Variables (EBVs) identified for a European wide biodiversity observation network. (a) Relative proportion of habitats, plants, invertebrates and vertebrate (birds, mammals, fish, reptiles and amphibians). (b) Representation of invertebrate groups. (c) Representation of different plant groups. (d) Different habitat types across freshwater, marine and terrestrial ecosystems. Source: EuropaBON deliverable D4.1 (Junker et al. 2023, revised version submitted on 31/01/2023).

For each EBV (n = 70), the EBV list provides information on the realm, the EBV class, the EBV name, and the desired spatial and temporal resolution as well as the taxonomic scope to be measured. It further includes a definition as well as metrics for each EBV. Three examples from this EBV list are shown below (Figure 3). The full version of the EBV list is openly available and can be accessed via GitHub (https://github.com/EuropaBON/EBV-Descriptions). This EBV list will form the basis for the virtual workshop on EBV workflows (22-24 February 2023) and for developing the co-design of the European Biodiversity Observation Network.

Please familiarize yourself with the EBV list from EuropaBON before the virtual workshop on EBV workflows: https://github.com/EuropaBON/EBV-Descriptions

This project receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101003553.

ARPHA Preprints *Author-formatted document posted on 14/02/2023*. DOI: https://doi.org/10.3897/arphapreprints.e101949

(a) EBV 'Aerial biomass of migrating birds, bats and insects'

Realm	Terrestrial
EBV class	Community composition
EBV name	Aerial biomass of migrating birds, bats and insects
EBV definition	Biomass flows of aerial migrants (birds, insects and bats) across Europe within contiguous spatial units (grid cells) over time
EBV metrics	Summary statistics of migration densities of birds, insects and bats derived from vertical profile time series of weather radar data (e.g. hourly averages of bird density and speed)
Spatial resolution	1 × 1 km – 10 × 10 km
Temporal resolution	1 day
Taxonomic scope	All migratory bird, bat and insect species (by size class)

(b) EBV 'Species distributions of freshwater invertebrates'

Realm	Freshwater
EBV class	Species populations
EBV name	Species distributions of freshwater invertebrates
EBV definition	The presence/absence or probability of occurrence of invertebrate species within lakes and river catchments over time
EBV metrics	• Binary presence/absence • Probability of occurrence
Spatial resolution	Lakes and river catchments as delineated in ECRINS (European catchments and rivers network system)
Temporal resolution	3 or 6 years
Taxonomic scope	• Freshwater invertebrate species listed in the Habitats Directive Annex II: Dragonflies: *Coenagrion hylas, C. mercuriale, Cordulegaster trinacriae, Gomphus graslinii, Leucorrhina pectoralis, Lindenia tetraphylla, Macromia splendens, Ophiogomphus cecilia, Oxygastra curtisii* Bivalves: *Margaritifera margaritifera* and *Unio crassus* • Freshwater invertebrate species listed in the Habitats Directive Annex IV: Dragonflies: *Aeshna viridis, Cordulegaster trinacriae, Gomphus graslinii, Leucorrhina albifrons, L. caudalis, L. pectoralis, Lindenia tetraphylla, Macromia splendens, Ophiogomphus cecilia, Oxygastra curtisii, Stylurus flavipes and Sympecma braueri* Bivalves: *Lithophaga lithophaga, Pinna nobilis, Margaritifera auricularia and Unio crassus* • Pollution-sensitive benthic invertebrates that are monitored for the Water Framework Directive: Mayflies (Ephemeroptera) Stoneflies (Plecoptera) Caddisflies (Trichoptera)

(c) EBV 'Phenology of marine spring phytoplankton bloom'

Realm	Marine
EBV class	Ecosystem function
EBV name	Phenology of marine spring phytoplankton bloom
EBV definition	The annual timing and intensity of spring phytoplankton blooms in EU's marine waters over time
EBV metrics	Phenology metrics of surface chlorophyll-a concentrations such as: • Day of start-of-blooming • Day of end-of-blooming • Blooming amplitude • Slope of the blooming up period • Length of the blooming season
Spatial resolution	10 × 10 km – 50 × 50 km
Temporal resolution	1 year
Taxonomic scope	All marine phytoplankton

Figure 3: Three examples of which details are captured in the list of Essential Biodiversity Variables (EBVs) identified for a European wide biodiversity observation network. The three examples represent an EBV description for the (a) terrestrial, (b) freshwater, and (c) marine realm, respectively. The detailed descriptions of all EBVs are available from GitHub (https://github.com/EuropaBON/EBV-Descriptions).

Co-designing EBV workflows

The virtual workshop on EBV workflows (22–24 February 2023) will focus on the co-design of EBV workflows. We here define workflows as a sequence of tasks that are needed to process a set of raw data (e.g. from in-situ monitoring or remote sensing) through data integration (e.g. standardizing, harmonizing or preprocessing) to modelling the data (e.g. statistical extrapolation or machine learning). In other words, EBV workflows are the paths

This project receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101003553.

that describe how we can go from raw observations (monitoring data) to spatially explicit (modelled) estimates that provide wall-to-wall maps (or data cubes) of EBVs across Europe. To achieve this, we consider various aspects of **(1) data collection and sampling**, **(b) data integration**, and **(c) modelling** as fundamental components of EBV workflows (Figure 4).

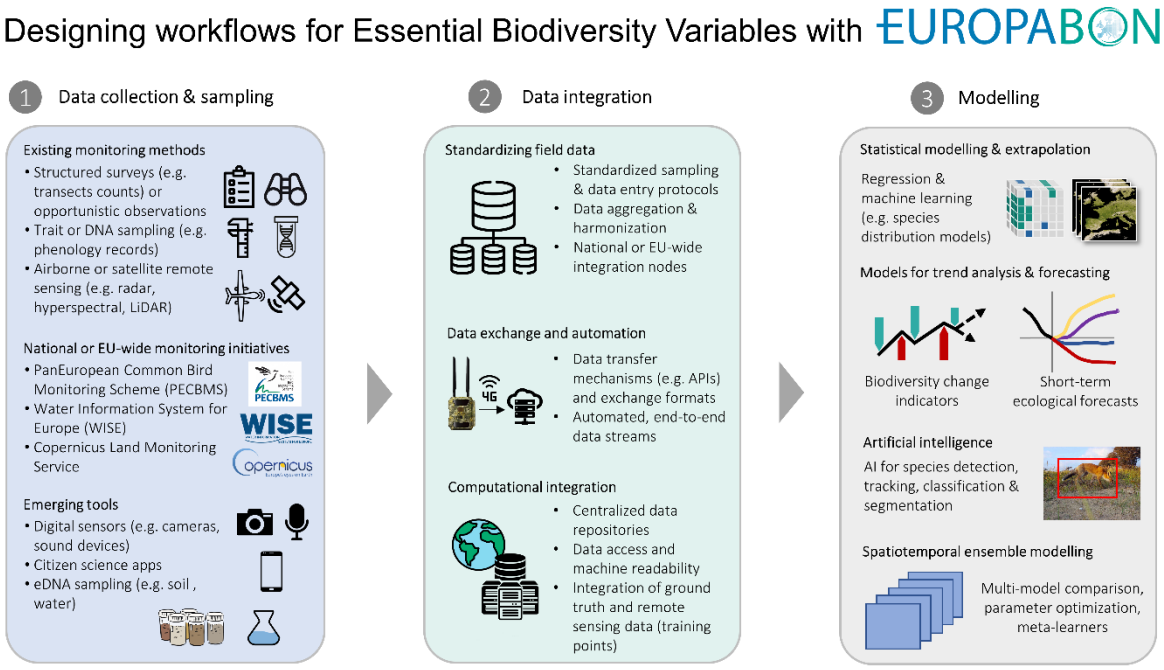

Figure 4: Designing workflows for Essential Biodiversity Variables (EBVs) in the context of an EU-wide biodiversity observation network requires information on (1) data collection and sampling, (2) data integration, and (3) modelling for each EBV. We need to identify existing monitoring methods, national or EU-wide monitoring initiatives, and newly emerging tools for data collection and sampling. Data integration requires to standardize and harmonize field data, to identify potential data exchange and automation mechanisms, and to specify what is require for computational integration. Modelling aspects might include statistical and geospatial extrapolation, trend analysis and short-term forecasting, the use of artificial intelligence for knowledge extraction from digital files (e.g. images or audio), or spatiotemporal ensemble modelling of remote sensing data.

Besides data collection, data integration, and modelling, the co-design of EBV workflows also needs to identify aspects of **interoperability** (i.e. the capacity of computers and software to exchange and make use of data and information) and the needs for **IT infrastructure** (e.g. platforms for data integration, storage and processing). This currently often limits the operationalization of EBVs on trans-national and cross-infrastructure scientific workflows (Hardisty et al. 2019, https://doi.org/10.1016/j.ecoinf.2018.11.003). We therefore include interoperability and IT infrastructure aspects in the co-design of EBV workflows (orange box in Figure 5), and hope to get input and insights on those aspects from the participants of the virtual workshop on EBV workflows (22-24 February 2023). Overall, the co-design of EBV workflows is a key component of the co-design of the European Biodiversity Observation Network (Figure 5), building on and contributing to other EuropaBON work packages and tasks (Figure 5).

This project receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101003553.

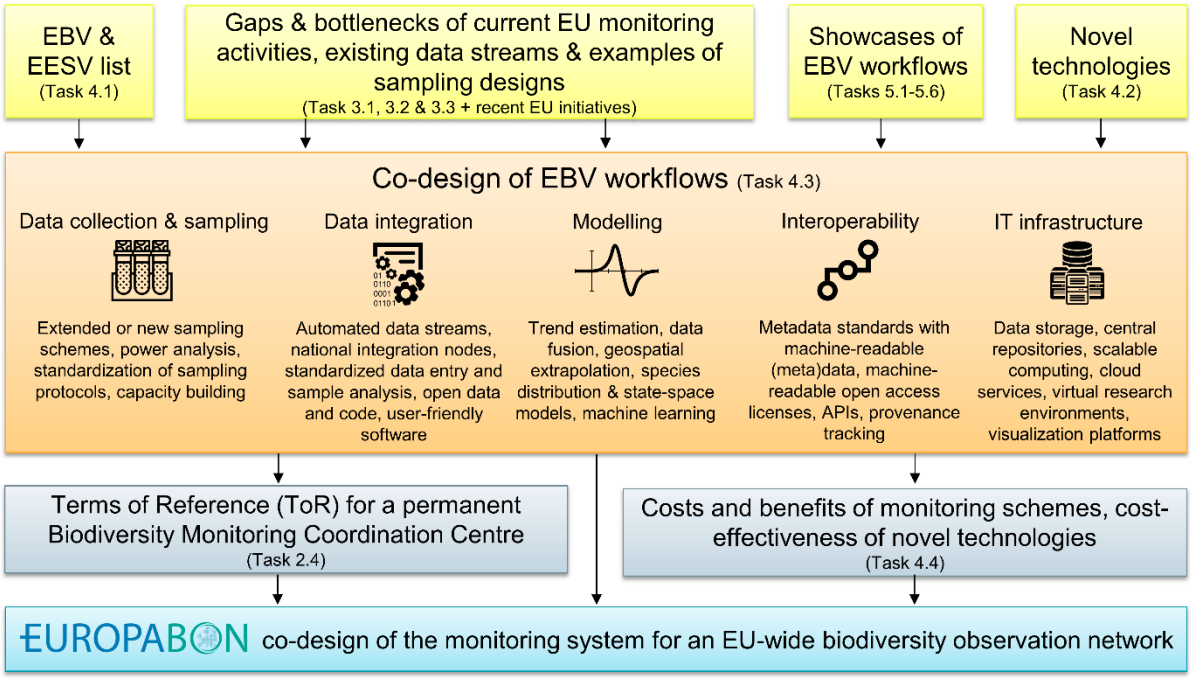

Figure 5: Co-designing the monitoring system for an EU-wide biodiversity observation network. Tasks from various EuropaBON work packages (yellow) feed information into the co-design of EBV workflows (orange). The co-design of EBV workflows includes aspects of data collection & sampling, data integration, modelling, interoperability and IT infrastructure. Together with the Terms of Reference (ToR) for a permanent Biodiversity Monitoring Coordination Centre and a cost-benefit analysis (gray), an overall EuropaBON co-design of the monitoring system for an EU-wide biodiversity observation network will be developed.

For describing the EBV workflows, we are particularly interested in gathering information from you on the following aspects of each EBV workflow:

- Data collection and sampling
    - Which primary observations (e.g. national species monitoring programs, citizen science, eDNA sampling, remote sensing) are currently collected (and by which initiative) that could enable the generation of a specific EBV at a European scale?
    - What are important or emerging national or EU-wide monitoring initiatives that are relevant for a specific EBV?
    - Which new tools (e.g. sensor networks, apps, eDNA sampling) and new monitoring schemes (e.g. from new EU initiatives or pilot projects) are currently emerging to create a specific EBV with European coverage?
- Data integration
    - Which methods are available or emerging for data integration and harmonization in the context of a specific EBV?
    - Which data exchange mechanisms or automation methods can be used for a specific EBV?
    - What is needed for computation integration (e.g. centralized repositories, preprocessing procedures, ground truth and remote sensing data)?
- Modelling
    - Which types of models are needed to generate a specific EBV, e.g. in terms of knowledge extraction (e.g. AI species identification from images), geospatial prediction (e.g. species distribution modelling or spatiotemporal ensemble modelling) or short-term ecological forecasting?
    - How can uncertainties be quantified?

This project receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101003553.

ARPHA Preprints *Author-formatted document posted on 14/02/2023*. DOI: https://doi.org/10.3897/arphapreprints.e101949

- Which emerging or new ways of modelling should be developed for creating an EBV with European coverage?

To describe EBV workflows for an EU-wide biodiversity observation network, we need to understand what is already established, what is newly emerging, and which future needs are required for operationalizing the EBV at a European level. We therefore aim to describe for each EBV and each workflow component (data collection and sampling, data integration, and modelling) these three aspects separately (i.e. current initiatives, emerging tools and projects, future needs):

- **Current initiatives:** Which data collection and sampling, data integration, and modelling is currently done in the context of a specific EBV, e.g. by existing and ongoing (national or EU-wide) monitoring initiatives?
- **Emerging tools and projects:** Which data collection and sampling, data integration, and modelling is emerging (but not yet fully operational), e.g. in the context of new EU (pilot) projects?
- **Future needs:** Which data collection and sampling, data integration, and modelling is additionally needed to establish operational workflows that facilitate EBV generation at a European level?

On the following pages (Figures 6–8), we provide three examples of workflow descriptions to illustrate these components. We also provide a template (in Appendix A) for filling in and collecting details for workflows of other EBVs. For the virtual workshop on EBV workflows (22–24 February 2023), we will provide online templates to fill this information, but you can already use the template (in Appendix A) to prepare for the workshop.

The workshop results and outcomes will feed into a report to the European Commission which provides recommendations and guidelines for the co-design of a European Biodiversity Observation Network with observation-to-policy-support workflows.

We would like to thank you very much for your interest and contribution! We are very much looking forward to having you participating in the virtual workshop on EBV workflows on 22–24 February 2023!

This project receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101003553.

Workflow example 1: EBV 'Species abundance of butterflies'

The first example provides information about the workflow for the EBV 'Species abundance of butterflies' (Figure 6). This EBV benefits from an existing EU-wide monitoring initiative, namely the European Butterfly Monitoring Scheme (eBMS). Current aspects of this initiative are standardized sampling protocols to collect transect data of butterflies with thousands of volunteers throughout Europe, the existence of an EU-wide integration node to combine national monitoring data, and derived trend estimates of butterfly abundance with a combination of statistical models (Figure 6, left column). Emerging tools include a citizen science phone app to count butterflies more easily, a monitoring and data integration effort that is expanded to the south and east of Europe (EU Pilot Project ABLE), and the testing of an integrated modelling framework (Figure 5, middle column). Future needs include an increase in capacity building and a better integration and modelling of the data (Figure 6, right column).

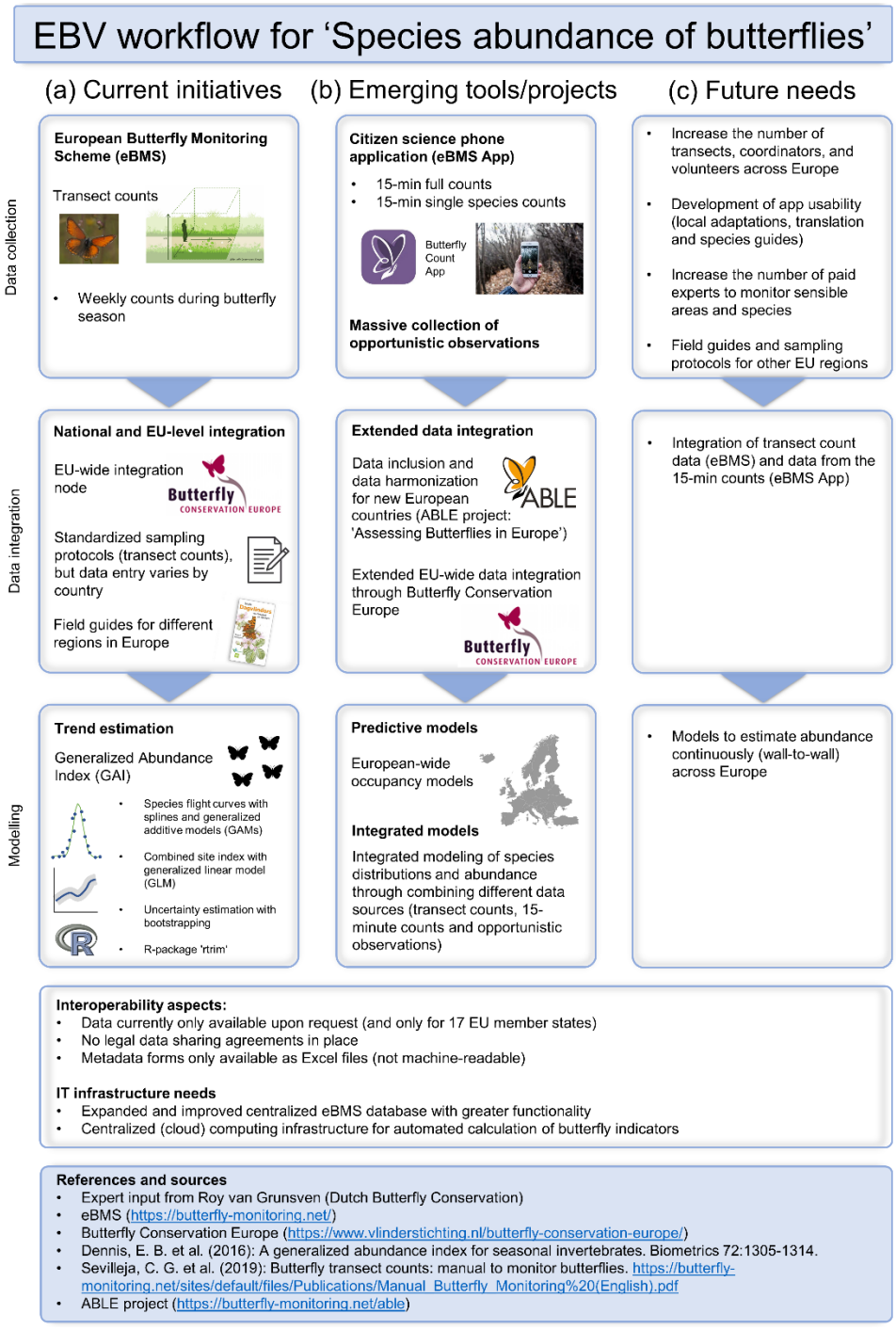

Figure 6: Example workflow for the EBV 'Species abundance of butterflies'. Further details are provided in Appendix B.

ARPHA Preprints *Author-formatted document posted on 14/02/2023*. DOI: https://doi.org/10.3897/arphapreprints.e101949

Workflow example 2: EBV 'Species abundances of selected terrestrial mammals'

The second example provides information about the workflow for the EBV 'Species abundances of selected terrestrial mammals' (Figure 7). This EBV illustrates how novel technologies (i.e. camera trapping and AI species identification) can complement traditional data collection and monitoring from diverse sources. Current aspects of mammal monitoring center around the European Mammal Atlas (EMMA2) which aggregates observations (records) at 50km x 50km grid cell resolution across Europe (Figure 7, left column). Emerging tools are currently developed through the EU-funded ENETWILD project and the European Observatory of Wildlife (EOW), e.g. by developing models for data integration, habitat suitability and abundance based on hunting statistics, roadkill information, density data, occurrence points (presence/absence), radio-tagging, and camera trapping (Figure 7, middle column). In addition, portals and AI models for wildlife images are emerging. Future needs include the expansion and coordination of a wildlife monitoring network across Europe, improved data integration and sharing, and improvements in modelling (Figure 6, right column).

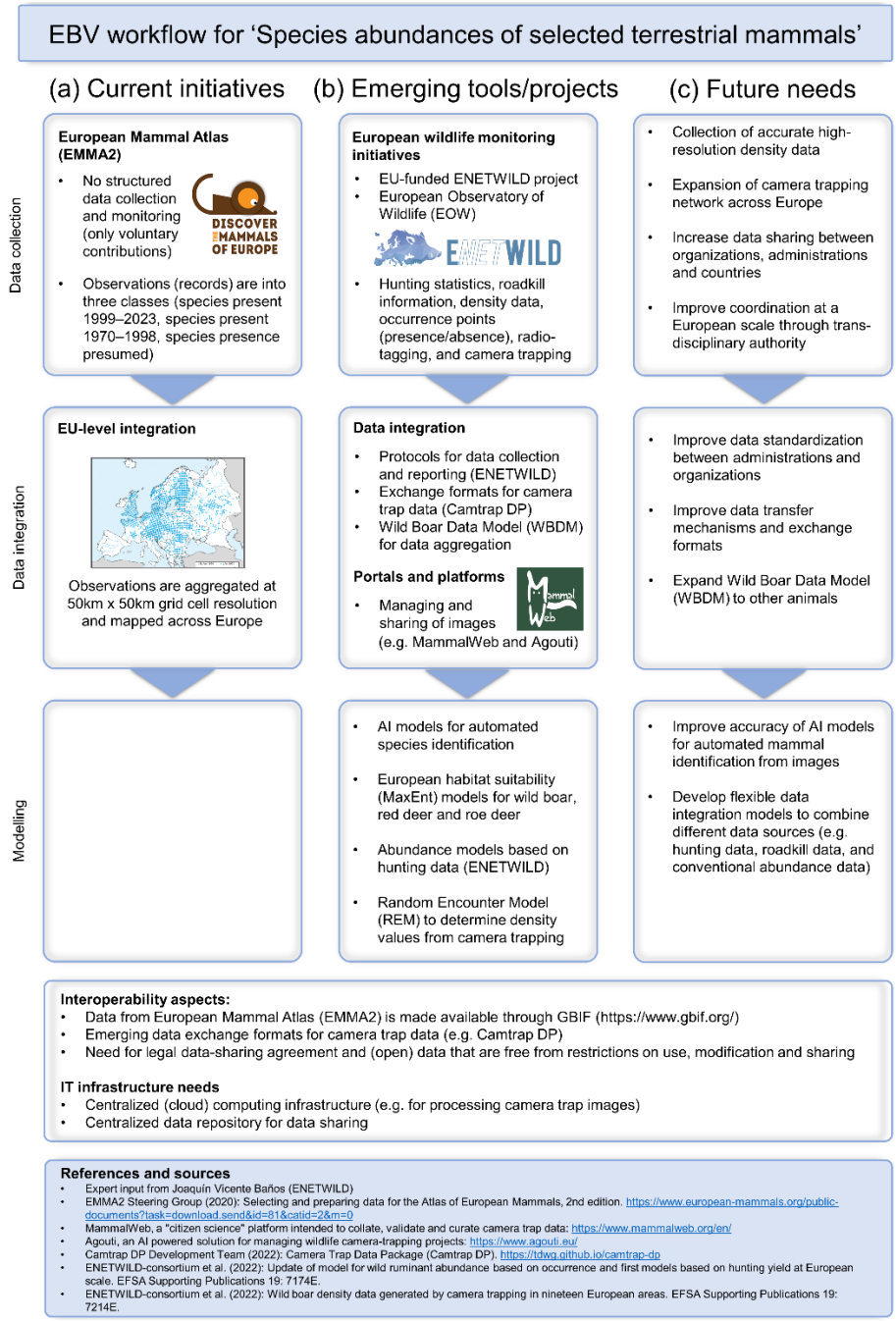

Figure 7: Example workflow for the EBV 'Species abundances of selected terrestrial mammals'. Further details are provided in Appendix C.

This project receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101003553.

Workflow example 3: EBV 'Ecosystem distribution of terrestrial EUNIS habitats'

The third example provides information about the workflow for the EBV 'Ecosystem distribution of terrestrial EUNIS habitats' (Figure 8). This EBV is derived from a combination of satellite remote sensing and ground truth data. Current aspects are the massive Earth Observation data collection through the Copernicus program of the European Space Agency (ESA) and in-situ measurements of European soils and vegetation plots (Figure 8, left column). Emerging tools are hyperspectral satellite missions to provide geochemical, biochemical and biophysical variables (EnMAP) and the mapping of European habitats using predictive habitat models and machine learning (Figure 7, middle column). Future needs include centralized data integration services for satellite, vegetation and soil data at the European level (Figure 8, right column).

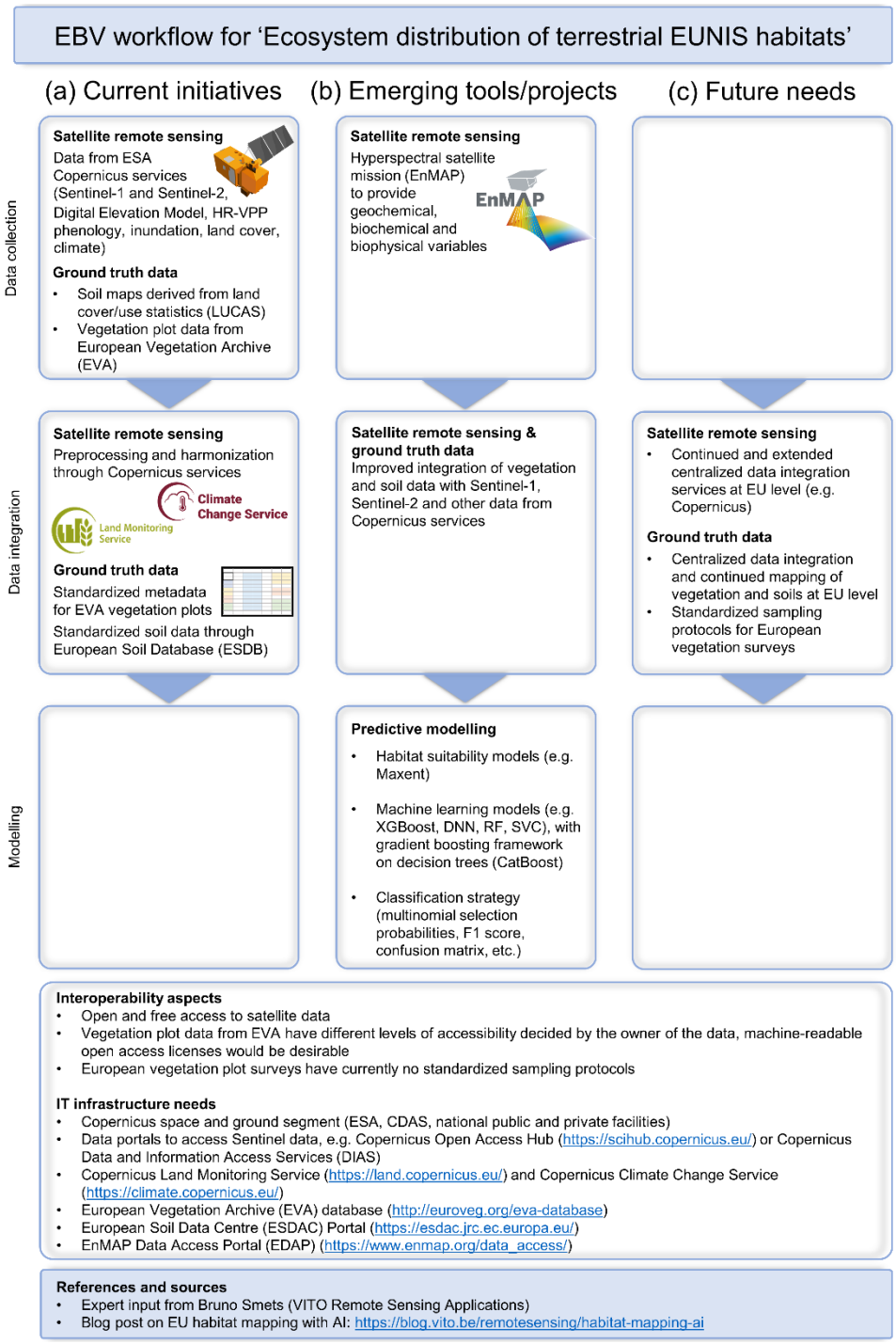

Figure 8: Example workflow for the EBV 'Ecosystem distribution of terrestrial EUNIS habitats'. Further details are provided in Appendix D.

This project receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101003553.

ARPHA Preprints Author-formatted document posted on 14/02/2023. DOI: https://doi.org/10.3897/arphapreprints.e101949

Appendix A: Template for describing EBV workflows

Workflow description for EBV (PROVIDE THE EBV NAME)			
	Current initiatives	**Emerging tools and projects**	**Future needs**
Data collection and sampling Data collection method Sampling design (EU-wide monitoring) Type of raw data Novel monitoring methods Capacity building			
Data integration Standardization & harmonization Preprocessing Protocols & metadata forms Way of data aggregation Integration nodes (national or EU) Automated data streams			
Modelling Types of models Predictors Estimation & uncertainty Software			

Interoperability aspects (e.g. access to and sharing of primary data, metadata standards, open access licenses, APIs, machine readability): •
IT infrastructure needs (e.g. data portals, use of European Research Infrastructures, data storage, central repositories, scalable computing, cloud services): •
References and sources (e.g. name and institution of expert who provided information for this template, literature, online sources, web pages of EU project): •

ARPHA Preprints Author-formatted document posted on 14/02/2023. DOI: https://doi.org/10.3897/arphapreprints.e101949

Appendix B: Details for the EBV workflow 'Species abundance of butterflies'

Workflow description for EBV (Species abundance of butterflies)			
	Current initiatives	**Emerging tools and projects**	**Future needs**
Data collection and sampling Data collection method Sampling design (EU-wide monitoring) Type of raw data Novel monitoring methods Capacity building	- European Butterfly Monitoring Scheme (eBMS) - Transect counts (weekly during the butterfly season (depending on the regions)	- Citizen science phone application (eBMS App) for 15-min full counts and 15-min single species counts - Massive collection of opportunistic observations	- Increase the number of transects, coordinators, and volunteers across Europe - Further development of app usability, including new local adaptations (translation and species guides) - Increase the number of paid experts to monitor sensible areas and species - Field guides and sampling protocols for other regions in Europe
Data integration Standardization & harmonization Preprocessing Protocols & metadata forms Way of data aggregation Integration nodes (national or EU) Automated data streams	- EU-wide integration node (Butterfly Conservation Europe) - Standardized sampling protocols for transect counts - Field guides for different regions in Europe - Different methods of data entry depending on the county	- Data inclusion and data harmonization for new European countries (ABLE project: 'Assessing Butterflies in Europe') - Extended EU-wide data integration through Butterfly Conservation Europe	- Integration of transect count data (eBMS) and data from the 15-min counts (eBMS App)
Modelling Types of models Predictors Estimation & uncertainty Software	- Species flight curves with splines and generalized additive models (GAMs) - Trend estimation with Generalized Abundance Index (GAI) - Combined site index with a generalized linear model (GLM) - Uncertainty estimation with bootstrapping - R-package 'rtrim'	- European-wide occupancy models - Integrated modeling of species distributions and abundance through combining different data sources (transect counts, 15-minute counts and opportunistic observations)	- Models to estimate abundance continuously (wall-to-wall) across Europe

Interoperability aspects (e.g. access to and sharing of primary data, metadata standards, open access licenses, APIs, machine readability):

ARPHA Preprints Author-formatted document posted on 14/02/2023. DOI: https://doi.org/10.3897/arphapreprints.e101949

• Data currently only available upon request (and only for 17 EU member states) • No legal data sharing agreements in place • Metadata forms only available as Excel files (not machine-readable)
IT infrastructure needs (e.g. data portals, use of European Research Infrastructures, data storage, central repositories, scalable computing, cloud services): • Expanded and improved centralized eBMS database with greater functionality • Centralized (cloud) computing infrastructure for automated calculation of butterfly indicators
References and sources (e.g. name and institution of expert who provided information for this template, literature, online sources, web pages of EU project): • Expert input from Roy van Grunsven (Dutch Butterfly Conservation) • eBMS (https://butterfly-monitoring.net/) • Butterfly Conservation Europe (https://www.vlinderstichting.nl/butterfly-conservation-europe/) • Dennis, E. B. et al. (2016): A generalized abundance index for seasonal invertebrates. Biometrics 72:1305-1314. • Sevilleja, C. G. et al. (2019): Butterfly transect counts: manual to monitor butterflies. https://butterfly-monitoring.net/sites/default/files/Publications/Manual_Butterfly_Monitoring%20(English).pdf • ABLE project (https://butterfly-monitoring.net/able)

This project receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101003553.

ARPHA Preprints Author-formatted document posted on 14/02/2023. DOI: https://doi.org/10.3897/arphapreprints.e101949

Appendix C: Details for the EBV workflow 'Species abundances of selected terrestrial mammals'

Workflow description for EBV 'Species abundances of selected terrestrial mammals'			
	Current initiatives	**Emerging tools and projects**	**Future needs**
Data collection and sampling Data collection method Sampling design (EU-wide monitoring) Type of raw data Novel monitoring methods Capacity building	- European Mammal Atlas (EMMA2) - No structured data collection and monitoring (only voluntary contributions) - Observations (records) are into three classes (species present 1999–2023, species present 1970–1998, species presence presumed)	- European wildlife monitoring initiatives through EU-funded ENETWILD project and European Observatory of Wildlife (EOW) - Hunting statistics, roadkill information, density data, occurrence points (presence/absence), radio-tagging, and camera trapping - Guidance to estimate population density with different methods	- Collection of accurate high-resolution density data - Expansion of camera trapping network across Europe - Increase data sharing between organizations, administrations and countries - Improve coordination at a European scale through trans-disciplinary authority
Data integration Standardization & harmonization Preprocessing Protocols & metadata forms Way of data aggregation Integration nodes (national or EU) Automated data streams	- Observations are aggregated at 50km x 50km grid cell resolution and mapped across Europe	- Harmonization protocols for data collection and reporting (ENETWILD) - Portals and platforms for managing or sharing wildlife images (e.g. MammalWeb and Agouti) - Data exchange formats for camera trap data (e.g. Camtrap DP) - Wild Boar Data Model (WBDM) to aggregate distribution and abundance data on wild boar and other ungulates	- Improve data standardization between administrations and organizations - Improve data transfer mechanisms and exchange formats - Expand Wild Boar Data Model (WBDM) to other animals
Modelling Types of models Predictors Estimation & uncertainty Software		- AI models for automated mammal identification from images (e.g. Agouti, Conservation AI) - European habitat suitability (MaxEnt) models for wild boar, red deer and roe deer, with topography, bioclimatic data, and land cover as predictors - Abundance model based on hunting data for wild boar, red deer, and roe	- Improve accuracy of AI models for automated mammal identification from images - Develop flexible data integration models to combine different data sources (e.g. hunting data, roadkill data, and conventional abundance data)

ARPHA Preprints Author-formatted document posted on 14/02/2023. DOI: https://doi.org/10.3897/arphapreprints.e101949

		deer, with topography, bioclimatic data, land cover, and human footprint as predictors (ENETWILD) - Random Encounter Model (REM) to determine density values from camera trapping	

Interoperability aspects (e.g. access to and sharing of primary data, metadata standards, open access licenses, APIs, machine readability): • Data from European Mammal Atlas (EMMA2) is made available through GBIF (https://www.gbif.org/), using the Darwin Core (DwC) metadata standard • Emerging data exchange formats for camera trap data (e.g. Camtrap DP) • Need for legal data-sharing agreements • Need for open data or data that are free from restrictions on use, modification and sharing
IT infrastructure needs (e.g. data portals, use of European Research Infrastructures, data storage, central repositories, scalable computing, cloud services): • Centralized (cloud) computing infrastructure (e.g. for processing camera trap images) • Centralized data repository for data sharing
References and sources (e.g. name and institution of expert who provided information for this template, literature, online sources, web pages of EU project): • Expert input from Joaquín Vicente Baños (ENETWILD) • EMMA2 Steering Group (2020): Selecting and preparing data for the Atlas of European Mammals, 2nd edition. https://www.european-mammals.org/public-documents?task=download.send&id=81&catid=2&m=0 • MammalWeb, a "citizen science" platform intended to collate, validate and curate camera trap data: https://www.mammalweb.org/en/ • Agouti, an AI powered solution for managing wildlife camera-trapping projects: https://www.agouti.eu/ • Camtrap DP Development Team (2022): Camera Trap Data Package (Camtrap DP). https://tdwg.github.io/camtrap-dp • ENETWILD-consortium et al. (2022): Update of model for wild ruminant abundance based on occurrence and first models based on hunting yield at European scale. EFSA Supporting Publications 19: 7174E. • ENETWILD-consortium et al. (2022): Wild boar density data generated by camera trapping in nineteen European areas. EFSA Supporting Publications 19: 7214E.

This project receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101003553.

ARPHA Preprints Author-formatted document posted on 14/02/2023. DOI: https://doi.org/10.3897/arphapreprints.e101949

Appendix D: Details for the EBV workflow 'Ecosystem distribution of terrestrial EUNIS habitats'

Workflow description for EBV 'Ecosystem distribution of terrestrial EUNIS habitats'			
	Current initiatives	**Emerging tools and projects**	**Future needs**
Data collection and sampling Data collection method Sampling design (EU-wide monitoring) Type of raw data Novel monitoring methods Capacity building	**Satellite remote sensing (ESA)** - Sentinel-1 and Sentinel-2 missions from Copernicus - Digital Elevation Model from Copernicus (e.g. 10m DEM) - High Resolution Vegetation Phenology and Productivity (HR-VPP) - Inundation from Copernicus - Land cover from Copernicus - Climate (annual precipitation, temperature, snow) from Copernicus **Ground truth data** - Soil maps (bulk density, clay fraction, sand fraction, pH, cation exchange capacity, course fractions, organic carbon), derived from land cover/use statistics (LUCAS) - In-situ vegetation plot data from European Vegetation Archive (EVA)	**Satellite remote sensing** - Hyperspectral satellite mission to provide geochemical, biochemical and biophysical variables, through Environmental Mapping and Analysis Program (EnMAP)	
Data integration Standardization & harmonization Preprocessing Protocols & metadata forms Way of data aggregation Integration nodes (national or EU) Automated data streams	**Satellite remote sensing** - Preprocessing and harmonization through Copernicus services (1 to 3 years of contiguous input data) **Ground truth data** - Standardized metadata for vegetation plots from European Vegetation Archive (EVA)	**Satellite remote sensing & ground truth data** - Improved integration of vegetation and soil data with Sentinel-1, Sentinel-2 and other data from Copernicus services	**Satellite remote sensing** - Continued and extended centralized data integration services at EU level (e.g. Copernicus) **Ground truth data** - Centralized data integration and continued mapping of vegetation and soils at EU level

ARPHA Preprints Author-formatted document posted on 14/02/2023. DOI: https://doi.org/10.3897/arphapreprints.e101949

	- Standardized soil data through European Soil Database (ESDB)		- Standardized sampling protocols for European vegetation surveys
Modelling Types of models Predictors Estimation & uncertainty Software		- Habitat suitability models (e.g. Maxent) - Machine learning models (e.g. XGBoost, DNN, RF, SVC), with gradient boosting framework on decision trees (CatBoost) - Classification strategy (multinomial selection probabilities, F1 score, confusion matrix, etc.)	

Interoperability aspects (e.g. access to and sharing of primary data, metadata standards, open access licenses, APIs, machine readability):

- Open and free access to satellite data
- Vegetation plot data from EVA have different levels of accessibility decided by the owner of the data, machine-readable open access licenses would be desirable
- European vegetation plot surveys have currently no standardized sampling protocols

IT infrastructure needs (e.g. data portals, use of European Research Infrastructures, data storage, central repositories, scalable computing, cloud services):

- Copernicus space and ground segment (ESA, CDAS, national public and private facilities)
- Data portals to access Sentinel data, e.g. through Copernicus Open Access Hub (https://scihub.copernicus.eu/) or through Copernicus Data and Information Access Services (DIAS) (several platforms, e.g. ONDA, sobloo, CREODIAS, mundi WEB SERVICES)
- Copernicus Land Monitoring Service (https://land.copernicus.eu/) and Copernicus Climate Change Service (https://climate.copernicus.eu/)
- European Vegetation Archive (EVA) database (http://euroveg.org/eva-database)
- European Soil Database (ESDB) and European Soil Data Centre (ESDAC) Portal (https://esdac.jrc.ec.europa.eu/)
- EnMAP Data Access Portal (EDAP) (https://www.enmap.org/data_access/)

References and sources (e.g. name and institution of expert who provided information for this template, literature, online sources, web pages of EU project):

- Expert input from Bruno Smets (VITO Remote Sensing Applications)
- Blog post on habitat mapping with AI: https://blog.vito.be/remotesensing/habitat-mapping-ai

This project receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101003553.