Project Report

Author-formatted document posted on 27/11/2024

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

DOI: https://doi.org/10.3897/arphapreprints.e142876

Data availability on European biodiversity, drivers and protected areas and gap analysis for European tetrapods

Laetitia Navarro, Francesca Cosentino, Virgilio Hermoso, Luigi Maiorano, Maria Paniw, Eloy Revilla, Andrea Sacchi, Luca Santini, Zulima Tablado, Wilfried Thuiller

ARPHA Preprints Author-formatted document posted on 27/11/2024. DOI: https://doi.org/10.3897/arphapreprints.e142876

NATURA CONNECT

Data on European biodiversity, drivers and protected areas

D3.1 Report and data on the biodiversity, protected areas and environmental and socioeconomic data available for the project. Including data gap analysis

Funded by the European Union

NaturaConnect receives funding under the European Union's Horizon Europe research and innovation programme under grant agreement number 101060429.

NaturaConnect receives funding under the European Union's Horizon Europe research and innovation programme under grant agreement number 101060429.

Prepared under contract from the European Commission.

Project acronym:	NaturaConnect
Project full title:	NaturaConnect - Designing a resilient and coherent Trans-European Network for Nature and People
Grant agreement number:	101060429
Start of the project:	1 July 2022
Duration:	48 months
Project coordinators:	International Institute for Applied Systems Analysis (IIASA) and Martin-Luther Universität Halle-Wittenberg (MLU) naturaconnect.eu
Scientific coordinator:	Piero Visconti, PhD, IIASA
Type:	HORIZON Innovation Actions
Call:	HORIZON-CL6-2021-BIODIV-01

Funded by the European Union

NaturaConnect receives funding under the European Union's Horizon Europe research and innovation programme under grant agreement number 101060429.

Project reference number	**101060429**
Project title	NATURACONNECT - DESIGNING A RESILIENT AND COHERENT TRANS- EUROPEAN NETWORK FOR NATURE AND PEOPLE

Deliverable title	Report and data on the biodiversity, protected areas and environmental and socioeconomic data available for the project. Including data gap analysis
Deliverable number	D3.1
Contractual date of delivery	31.07.2024
Actual date of delivery	31.10.2024
Type of deliverable	R — Document, report
Dissemination level	PU - Public
Work package number	WP3
Institution leading work package	CNRS
Task number	T3.1; T3.2
Institution leading task	CSIC (T3.1); US (T3.2)
Author(s)	Laetitia M. Navarro[1], Francesca Cosentino [2], Miguel Fernandez[3,4], Virgilio Hermoso[1], Luigi Maiorano[2], Maria Paniw[1], Eloy Revilla[1], Luca Santini [2], Andrea Sacchi [2], Zulima Tablado[1], Wilfried Thuiller[5] Author affiliations 1. Department of Conservation Biology and Global Change, Doñana Biological Station (EBD-CSIC), Américo Vespucio nº 26 –41092 Sevilla, Spain 2. Sapienza affiliation: Department of Biology and Biotechnologies ‘‘Charles Darwin’’, Sapienza University of Rome, 00185 Rome, Italy 3. German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Leipzig, Germany 4. Institute of Biology, Martin Luther University Halle Wittenberg, Halle (Saale), Germany 5. Laboratoire d'Ecologie Alpine, CNRS, Univ. Grenoble Alpes, Grenoble, France
Deliverable reviewers	Martin Jung (IIASA), Matea Osti (IIASA)
EC project officers	Christophe Coudun and Colombe Warin

Funded by the European Union

NaturaConnect receives funding under the European Union's Horizon Europe research and innovation programme under grant agreement number 101060429.

NATURA CONNECT

Deliverable description	One dataset will include all distribution data collected for terrestrial species groups (plants, vertebrates, invertebrates). An additional dataset will feature vegetation type data. We will also inventorize and harmonize trait data. A fifth dataset will include all environmental and socio-economic spatial variables, including earth observation (land cover at 100-m resolution) and climate data from Copernicus (bioclimatic indicators), soil data, land-use and other socio-economic variables (distance to cities, infrastructure, human population density, etc) from various sources (e.g. EEA, SoilGrids). The report instead will include the results of our investigation of gaps in knowledge in biodiversity data for Europe to provide measures of uncertainty, at the same time, highlighting (geographically and thematically) the areas/habitats/bioregions with monitoring and integration needs.
Keywords	Metadata; Essential Biodiversity Variables; DPSIR; data availability; data access

Funded by the European Union

NaturaConnect receives funding under the European Union's Horizon Europe research and innovation programme under grant agreement number 101060429.

Table of Contents

Funded by the European Union

NaturaConnect receives funding under the European Union's Horizon Europe research and innovation programme under grant agreement number 101060429.

Funded by the European Union

NatureConnect receives funding under the European Union's Horizon Europe research and innovation programme under grant agreement number 101060429.

NATURA CONNECT

List of Figures

Funded by the European Union
NaturaConnect receives funding under the European Union's Horizon Europe research and innovation programme under grant agreement number 101060429.

ARPHA Preprints Author-formatted document posted on 27/11/2024. DOI: https://doi.org/10.3897/arphapreprints.e142876

List of Tables

Funded by the European Union

NaturaConnect receives funding under the European Union's Horizon Europe research and innovation programme under grant agreement number 101060429.

NATURA CONNECT

ARPHA Preprints Author-formatted document posted on 27/11/2024. DOI: https://doi.org/10.3897/arphapreprints.e142876

Abbreviations

API	Application Programming Interface
CC	Creative Commons
CHELSA	Climatologies at high resolution for the earth's land surface areas
DPSIR	Drivers Pressures State Impact Response
EBV	Essential Biodiversity Variables
EEA	European Environment Agency
EESV	Essential Ecosystem Services Variables
ESA CCI	European Space Agency Climate Change Initiative
EUNIS	European nature information system
EVA	European Vegetation Archive
GBIF	Global Biodiversity Information Facility
IUCN	International Union for Conservation of Nature
LUCAS	Land Use/Cover Area frame statistical Survey
NetCDF	network Common Data Form

Funded by the European Union
NaturaConnect receives funding under the European Union's Horizon Europe research and innovation programme under grant agreement number 101060429.

ARPHA Preprints Author-formatted document posted on 27/11/2024. DOI: https://doi.org/10.3897/arphapreprints.e142876

Executive summary

This report presents an overview of data identification and documentation related to biodiversity, ecosystem services, and the associated drivers, pressures, and response mechanisms.

While not systematic nor exhaustive, our effort of data identification and documentation allowed us to describe more than 100 datasets and databases on European biodiversity (most datasets), ecosystem services, the drivers and pressures affecting them, and the mechanisms put in place to address these. These datasets represent nearly 2000 variables and metrics that can be used directly by researchers, land managers and decision-makers, for example for spatial planning in conservation or for further integration into biodiversity and ecosystem services models.

Moreover, we also evaluate the completeness of biodiversity data (occurrence, trait and biotic interactions) in Europe across four terrestrial vertebrate classes, and assess potential drivers of data completeness. Despite Europe being one of the richest continents for biodiversity data globally, there are substantial data gaps in species distribution, trait, and species interactions, particularly in Eastern Europe, and for reptiles and amphibians. Results highlight how this heterogeneity in data availability is strongly associated with socioeconomic factors.

Some of the key findings and recommendations that arise from the documentation of the datasets and gap analyses are:

Key findings:

Gaps remain in biodiversity, ecosystem services, drivers and pressures dataset availability

Within the 137 datasets identified and documented, data on freshwater systems, ecosystem functions and population abundances are overall still under-represented in large-scale biodiversity data repositories and catalogues such as the ones that we consulted to build our metadatabase. Only 4% of the datasets are dedicated to freshwater data, when 47% contain information exclusively on terrestrial systems for instance. We find that most of the metrics identified can be classified as species traits (both functional and life-history traits) although those also largely related to static data averaged across species and time, which hampers our ability to assess species response to pressure at the trait level.

Funded by the European Union

NatureConnect receives funding under the European Union's Horizon Europe research and innovation programme under grant agreement number 101060429.

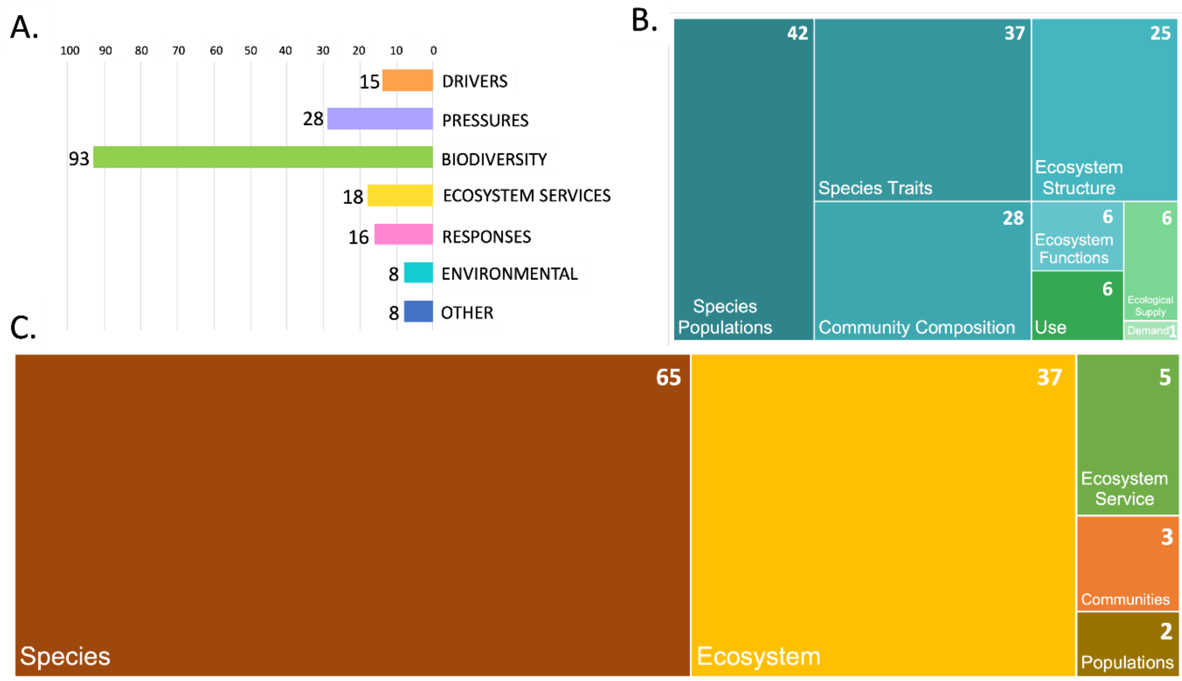


Figure 1: Summary description of the data identified and documented in the metadatabase (Sup. Table 2): (A) Number of datasets and databases described; (B) Coverage of Essential Biodiversity Variables (EBVs) and Essential Ecosystem Services Variables (EESVs); and (C) Biological entity and taxonomic resolution.

The vast majority of datasets are openly available, but vary in their formats and level of standardisation

By design, most of the datasets that we describe are openly available and easily accessible. Concretely, 88 of the 137 entries clearly identify a Creative Commons license and an additional 27 are described as Open Access, either with a sharing policy specific to the curator (e.g. EEA standard re-use policy) or without a specified license policy. Nevertheless, the datasets also vary greatly in their formats and standardisation efforts, which would impair a smooth integration into open workflows. A call for open and standardised data in biodiversity conservation should hence be made in parallel to the development of models and tools such as those that projects as NaturaConnect can develop.

Data gaps show a strong taxonomic and geographic bias

Knowledge gaps are unevenly distributed within the European continent, showing a strong taxonomic and geographic bias; Amphibian and reptile data are strongly under-sampled compared to mammals and birds considering the species distribution (Wallacean shortfall), biological traits (Raunkiæran shortfall), and trophic interactions data (Eltonian shortfall); These gaps were spatilally biased, with larger gaps located in southern and eastern Europe, strongly under-sampled compared to central and northern Europe (e.g., France, United Kingdom).

Key recommendations:

Metadata reporting should be standardised

There is a need to promote the publication of open protocols that describe in a standardised way the ontologies used and the data processing approach behind data production. Importantly, these protocols would clearly identify the inputs and outputs of models used for decision-making and research in biodiversity conservation. Adopting this approach would limit

Funded by the European Union
NaturaConnect receives funding under the European Union's Horizon Europe research and innovation programme under grant agreement number 101060429.
NATURA CONNECT

the risk of redundancy, overestimations and circularity when integrating several datasets from various sources and disciplines.

Geographical and taxonomic gaps need to be filled

Priority areas for data collection are located in southern and eastern Europe, which are strongly under-sampled compared to central and northern Europe (e.g., France, United Kingdom).

These findings enhance the understanding of data uncertainties in the biodiversity data used for modelling biodiversity and planning the TEN-N. The implementation of standardised large-scale monitoring could also improve the comprehension of biodiversity, ultimately guiding future research efforts and data collection across Europe. Addressing these issues is crucial for advancing biodiversity conservation and ecosystem service management across Europe.

Data Availability

The Supplementary Tables listed and referenced to in Part A of this deliverable are openly available (CC-BY Licence) on Zenodo (https://doi.org/10.5281/zenodo.14008205). Specifically, Sup.Table 1 is titled Typology.xlxs; Sup.Table 2 is named MetaDatabase.xlsx; and Sup. Table 3 is DetailedMetrics.xlsx.

Funded by the European Union

NaturaConnect receives funding under the European Union's Horizon Europe research and innovation programme under grant agreement number 101060429.

1. Part A - Documentation of data on biodiversity, drivers and protected areas

1.1. Introduction: Data needs for informed biodiversity conservation

Understanding and predicting the patterns of global change, their impact on biodiversity and ecosystem processes (across space and time) and the consequences for human well-being is a prominent need shared by the scientific community and policy-makers alike (Gonzalez et al., 2023; Navarro et al., 2017; Pereira & Cooper, 2006). Scientists are on a long-term quest to improve the detection of biodiversity change and its attribution to an array of interacting driving forces, from global consumption patterns to local habitat loss and degradation (e.g., Marques et al., 2019). Meanwhile, policy and decision-makers rely on science-based knowledge to support decision making in biodiversity conservation and to assess the effectiveness of the measures put in place to achieve conservation goals and targets (Perino et al., 2022). Whether those needs are addressed by biodiversity or ecosystem services models (e.g., Pereira et al., 2024) or by the use of decision support tools (e.g. conservation planning in Jung et al., 2024), they all rely on access to open and timely data that is temporally, spatially and thematically representative and that can smoothly be interpreted and integrated into data-to-knowledge workflows.

In the particular case of NaturaConnect for instance, where the main goal is to produce data, knowledge and tools to support EU Member States in developing an ecologically representative and resilient Trans-European Nature Network (TEN-N), data is needed on (1) the indirect drivers and direct pressures impacting biodiversity across Europe, (2) the distribution of species and ecosystems, their dynamics and functions, and (3) ecosystem services supply and use. Furthermore, current conservation approaches and mitigation measures also need to be identified and mapped in order to be integrated into area-based conservation planning exercises (Moreira et al., 2024, Jung et al., 2024). In this context, the aim of Task 3.1 and this deliverable was to identify and document all relevant data that could serve as input for the other tasks of the project (and beyond) either directly or after processing in biodiversity or ecosystem services models. This effort focused on data that is pan-European in scope and covers the terrestrial and freshwater realms.

1.2. A typology for data on biodiversity and pressures

The datasets of potential relevance for NaturaConnect range from data on drivers and pressure on biodiversity, to species occurrences and traits, and the mechanisms put in place to limit and/or reverse the impact of those pressures on biodiversity. To categorise these diverse datasets, we propose a common typology that builds on multiple widely used frameworks (Sup. Table 1; Figure 1). The starting point of this typology is the **DPSIR framework** originally proposed by the European Environmental Agency in the late 1990s and which is a framework integrating **D**rivers; **P**ressures; **S**tates; **I**mpacts; and **R**esponses. The DPSIR framework was designed to describe how socio-economic factors (drivers, pressures and responses) interact with the environment (states and impacts), and importantly, is meant to be policy relevant (Maxim et al., 2009; Ness et al., 2010). While the DPSIR framework is deemed too simplistic for analysis of causal relations and interactions between its components (for instance indicators related to invasive species fall within the pressure, state and impact categories), it is considered an important tool to structure (and communicate) information on

Funded by the European Union
NaturaConnect receives funding under the European Union's Horizon Europe research and innovation programme under grant agreement number 101060429.
NATURA CONNECT

socio-environmental dynamics (Maxim et al., 2009). In particular, the framework can be applied in the context of NaturaConnect, where different scenarios for drivers and pressures (related to climate and land-use change) are considered together, while area-based conservation and the restoration of connectivity can be seen as responses to the potential impact on biodiversity (species and ecosystems) and ecosystem services supply. We then adapted the DPSIR framework to the more specific needs of NaturaConnect and in order to better assess the availability of data across the components of the framework, as described below.

The **D**rivers category encompasses data on socio-economic and socio-cultural forces driving human activities that increase pressure on biodiversity, such as those related to human population size and consumption patterns.

To describe data related to **P**ressures in more details, we used the **Threats Classification Scheme** (version 3.3) of the International Union for Conservation of Nature (IUCN). This hierarchical classification covers the main direct pressures on biodiversity used by assessors of the IUCN Red List of Threatened Species to document the status of species and now ecosystems (Fig. 2), for instance “Biological resource use”, “Pollution”, and “Energy production and mining”. This classification is particularly interesting because, having been developed for conservation purposes, it sits at the intersection of the socio-ecological system, where human driven pressures such as “energy production and mining” can be more easily understood as, and broken down into, “habitat loss” and “direct mortality” and from there directly connected to the state of biodiversity. In other words, the advantage of the IUCN threats classification is that it is pertinent to biodiversity research and management.

State and **I**mpacts were merged into a single category of “Essential Variables” which, by describing trends, cover both state and change. Here we considered the frameworks of **Essential Biodiversity Variables** (EBVs) and **Essential Ecosystem Services Variables** (EESVs) for biodiversity and ecosystem services respectively. The framework for the EBVs has been developed to identify the “minimum set of biological state variables that are needed to detect biodiversity change” (Navarro et al., 2017). EBVs are particularly useful to facilitate the description, mobilisation and integration of biodiversity data and to monitor biodiversity change across species, ecosystems and realms, thus benefiting both the scientific community and decision-makers (Geijzendorffer et al., 2016; Pereira et al., 2013). The essential variables for biodiversity are organised into six EBV classes (Sup. Table 1), with 3 classes of variables focusing on the species level, that is Genetic Composition, Species Populations and Species Traits; and 3 classes focusing on the ecosystem level, that is Community Composition, Ecosystem Structures and Ecosystem Functions (Fernández et al., 2020). More detailed classifications of the variables, as well as their production workflows, have been described for Species Populations (Jetz et al., 2019; Kissling, Ahumada, et al., 2018), Species Traits (Kissling, Walls, et al., 2018) and Genetic Composition (Hoban et al., 2022). The Essential Variables framework for ecosystem services accounts for their human dimension, and describes the ecosystem services flow from nature to societies (Balvanera et al., 2022) by organising the data into six classes (Sup. Table 1). The first two classes cover the creation and supply of the services, that is the Ecological Supply (i.e. the capacity of ecosystems to provide the services) and Anthropogenic Contributions (i.e. human contributions to the supply of services). Two other classes encompass the actual appropriation of the services by humans, that is Demand and Use. Lastly, two classes cover data that describe the values that services bring to society, both Instrumental Values and Relational Values (Díaz et al., 2018; IPBES, 2022).

RESPONSES in the context of NaturaConnect relates to data on the mechanisms put in place to mitigate the impacts of drivers and pressures on biodiversity and ecosystem services. This is the case for instance of data relating to the implementation of the Birds and Habitats

Funded by the European Union
NaturaConnect receives funding under the European Union's Horizon Europe research and innovation programme under grant agreement number 101060429.

Directives (and their reporting), to the distribution of protected areas, and ecosystem restoration measures. Specifically, we divided the responses category into "area-based conservation", "species listing" and "habitats listing". Lastly, we considered two additional data categories beyond the DPSIR framework: ENVIRONMENTAL data that encompass for instance climatic variables; and OTHERS, which regroups static data that does not belong to any part of the frameworks discussed above such as the spatial data for the NUTS classification, or the map of water bodies. Note that the different components of the frameworks do not need to be considered in silo (Fig. 1). For instance, datasets documenting "habitat change" and "biological resources use" within the Pressures typology can also inform on the "use" of Ecosystem Services. Likewise, data on drivers of biodiversity change are conceptually tightly connected to the "demand" of ecosystem services.

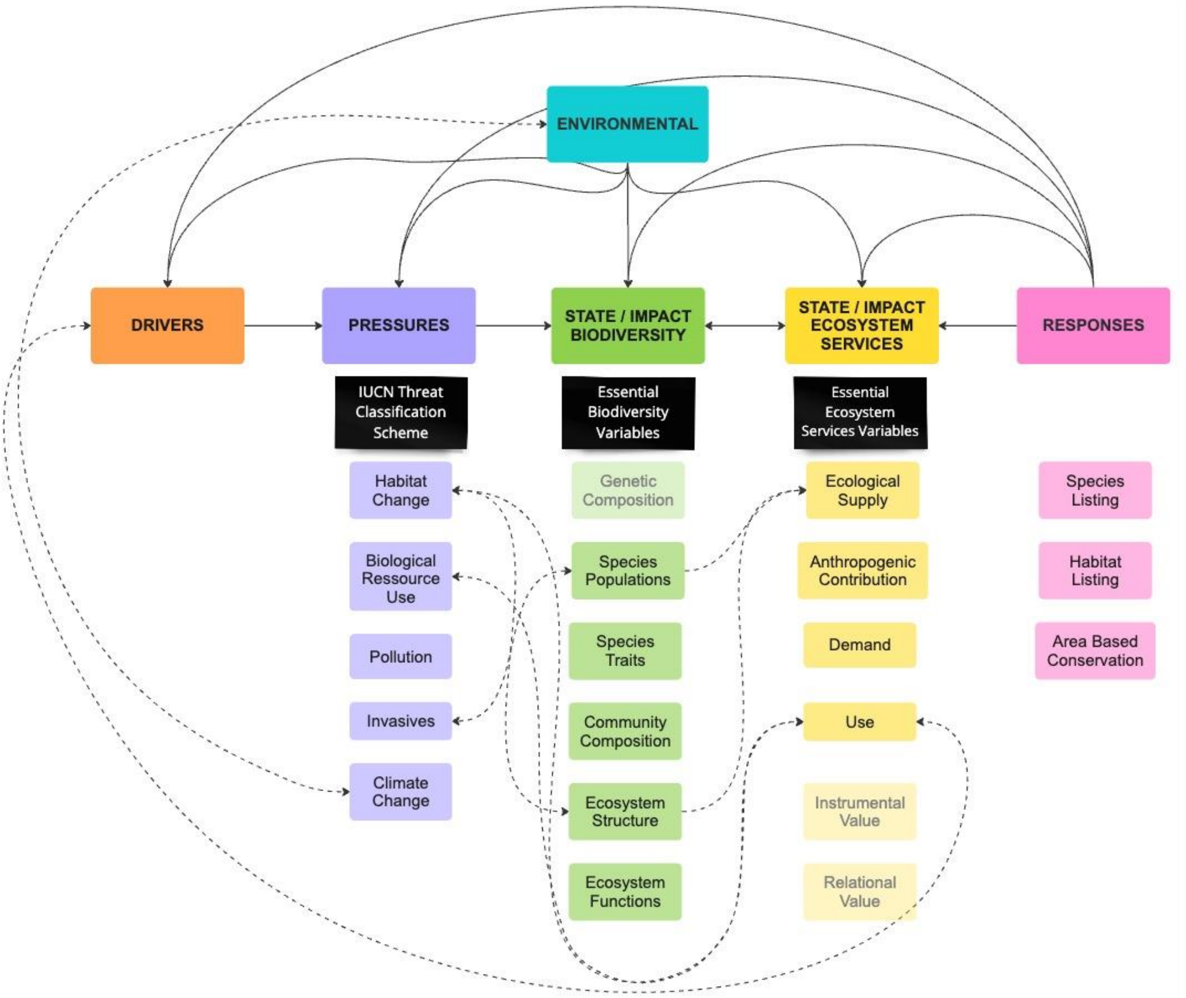


Figure 2: Schematic representation of the typology used to organise the different datasets documented in the deliverable. The interactions between the components of the typology are represented with solid lines while dashed lines are used to represent equivalences (e.g., data corresponding to habitat change and biological resource use within the "pressure" category can also correspond to use within the "ecosystem services" category). Components of the framework represented with transparent boxes were not considered in the data identification and documentation effort (see text for more details).

In the specific context of NaturaConnect, and after consultation with the members of the project consortium, some elements of the full typology were considered of lesser relevance and hence ignored. This is the case for instance with "Geological events" (e.g., volcanoes and earthquakes) as a pressure. We also considered that data related to Genetic Composition would not be used in the project and hence go beyond the scope of the data identification and

Funded by the European Union
NaturaConnect receives funding under the European Union's Horizon Europe research and innovation programme under grant agreement number 101060429.

documentation effort presented here. For other data types, the development, both conceptual and technical, is still in its infancy and they were ignored for lack of feasibility (e.g. data on relational values of ecosystems).

1.3. Methods

1.3.1. The metadatabase: description of fields and coverage

In order to document the datasets and databases identified as relevant, that is containing data on biodiversity, ecosystem services, drivers, pressures and/or responses (Fig. 2), we built a metadatabase (Supplementary Table 2). This metadatabase documents for each dataset or database:

General information on each entry, that is its name, the corresponding component of the data typology (Fig. 2, Supplementary Table 1), for instance if the data concerns biodiversity, or pressures on biodiversity. This section also documents the type of information or metrics contained in the entry and their unit as well as the realm (Terrestrial or Freshwater) covered by the data. In many cases, an entry will contain data on more than one variable or product, in which case we labelled it as “multiple” in the general information and list all individual metrics and their unites in a separate table (Supplementary Table 3).

Biological information: if the entry relates to data on biodiversity or ecosystem services, this section is used to inform about the biological entity and taxonomic resolution of the data (e.g. species), the coverage of the biological entity (e.g. amphibians), and the coverage of Essential Variables (EBV or EESV – e.g., species traits).

Non-biological information: for entries that provide data on drivers, pressures or responses, we document the entity (e.g., type of pressure) and the coverage or scope of the entity.

Temporal information: we describe the temporal extent of each entry and their temporal resolution for those that are repeated measurements in time.

Spatial information: This section of the metadatabase documents, for the entries that are spatially explicit, which is the spatial scope (e.g., global, national), the spatial extent (e.g. EU28, Spain), and the spatial resolution of the data.

Method: for each entry, we document whether the data is modelled, interpreted or raw, as well as the dependencies with other datasets. Specifically, we identify if the data is also shared or used in another dataset (either documented in the metadatabase or not).

Accessibility: this last part of the metadatabase documents the links to (and references of) the data, and, when appropriate, the scientific publication accompanying them. We also keep track of the curator and contact person as well as the last update of the entry. This section is also used to document the data format (e.g., NetCDF, csv), licensing and whether the data can be accessed via an Application Programming Interface (API) or another tool.

The metrics of each entry are listed in a separate table (Supplementary Table 3) and are mapped to the data typology presented above, and when appropriate to the corresponding Essential Biodiversity Variable or Essential Ecosystem Service Variable. This table documents the name of the metric, or field, as given in the source material, its type (e.g., number, categorical, characters) and when appropriate, its unit. When the information is provided in the source material, we also add a definition of the metric as well as the different options in the case of categorical data.

Funded by the European Union

NaturaConnect receives funding under the European Union's Horizon Europe research and innovation programme under grant agreement number 101060429.

NATURA CONNECT

1.3.2. Identification of datasets and databases through consultation with consortium and open databases search

The entries of the metadatabase were identified through three main approaches. First, a list of online catalogues and repositories was produced and scoped for relevant datasets or databases (Table 1). We consulted for instance the data catalogues of the European Environment Agency, and the COPERNICUS Land Monitoring Service, which provided a substantial number of datasets on biodiversity at the ecosystem level. This first scoping exercise was meant to focus the identification of datasets on those that are made widely available and known and understood by data users, including non-academic users beyond the consortium. This also allowed us to focus on data availability at the global and regional scale since documenting national and sub-national datasets, while clearly valuable, would go beyond the initial scope of the metadatabase. Second, a survey was sent out to all NaturaConnect consortium members in the third quarter of 2022 to identify both their needs and uses of data across the data typology. Lastly, the research team punctuallyadded scientific publications of large-scale datasets, although it is important to highlight that this is not resulting from a systematic survey effort of the literature.

Table 1: Catalogues and repositories used to identify, document and access the datasets and databases described in the metadatabase.

Name	Link
European Environment Agency Datahub	https://www.eea.europa.eu/
European Environment Agency EIONET Central Data Repository	https://cdr.eionet.europa.eu/
Copernicus Land Monitoring Service	https://land.copernicus.eu/en
Essential Biodiversity Variables (EBV) data portal of the Group on Earth Observations Biodiversity Observation Network	https://portal.geobon.org/home
OpenTraits Network Catalogue	https://opentraits.org/datasets.html
Open Environmental Data Cube Europe	https://stac.ecodatacube.eu/
NASA's Earth Data	https://www.earthdata.nasa.gov/
NASA's SEDAC (Socioeconomic data and application center)	https://sedac.ciesin.columbia.edu/
Euro-Lex (access to European Union Law)	https://eur-lex.europa.eu/homepage.html
JRC - ESDAC (European Soil Data Center)	https://esdac.jrc.ec.europa.eu/resource-type/datasets
Database of European Vegetation, Habitats and Flora	https://floraveg.eu/
ESA (European Space Agency) Climate Office	https://climate.esa.int/en/data/#/dashboard

1.4. Current availability of data on biodiversity, drivers and protected areas in Europe

1.4.1. Overall description of the metadatabase

As of September 2024, we identified and documented 137 databases or datasets in the metadatabase (Supplementary Table 2; Fig. 3A.). The majority of the entries in the metadatabase concern biodiversity data (N=93; 68%), followed by data on pressures (N=29; 21%). Note that a single entry in the metadatabase can cover more than one element of the typology (Fig. 2): some databases provide for instance data on both biodiversity and the pressures on it, including its use, that is ecosystem services (e.g., IUCN Red List of

Funded by the European Union
NaturaConnect receives funding under the European Union's Horizon Europe research and innovation programme under grant agreement number 101060429.

Threatened Species; annual European Land use and Land cover dataset). About a third of those datasets (N=43) provide data that is repeated in time with frequencies ranging from days (e.g., 10 days for the Fraction of green vegetation cover) to decades (e.g., 10 years for the global wetland loss reconstruction). In contrast, the majority (N=94) of entries are of spatially explicit data, with 53 datasets providing global data and 49 focusing on continental Europe (19 of which are specifically for the Member States of the European Union). While we do list 4 national datasets that were identified at some point in the process of populating the metadatabase, we acknowledge that this is in no way representative of the availability of national data, which goes beyond the scope of the effort documented here. The spatial resolution of the datasets ranged from point or plot data (N=16), to meters (N=23) and in some cases country level data (N=5). Nearly half of the entries cover both the terrestrial and freshwater realms (N=65), although an equivalent number of datasets are exclusively terrestrial, and an additional 6 datasets cover exclusively freshwater systems.

There is variability in the number of data layers and/or metrics provided in each dataset. For instance, the HomeRange dataset provides a single type of information, that is estimates of home range sizes for mammals while the LUCAS TOPSOIL dataset provides 12 different layers of data on soil properties. Thus, for each dataset, we extracted the list of variables, their type (e.g. number; categorical) and unit and when available their definition. Those variables were then mapped to the data typology, to the lowest level possible (Supplementary Table 3; Fig. 3B). For biodiversity data for instance, we mapped the variables to the Essential Biodiversity Variables classes and variables themselves. In total, the datasets and databases documented contain data on 1,849 variables. The vast majority (>80%) provide information and data about biodiversity (N=1537) although it is important to note that this figure is to some extent artificially inflated by data on species traits (936 variables in 37 datasets).

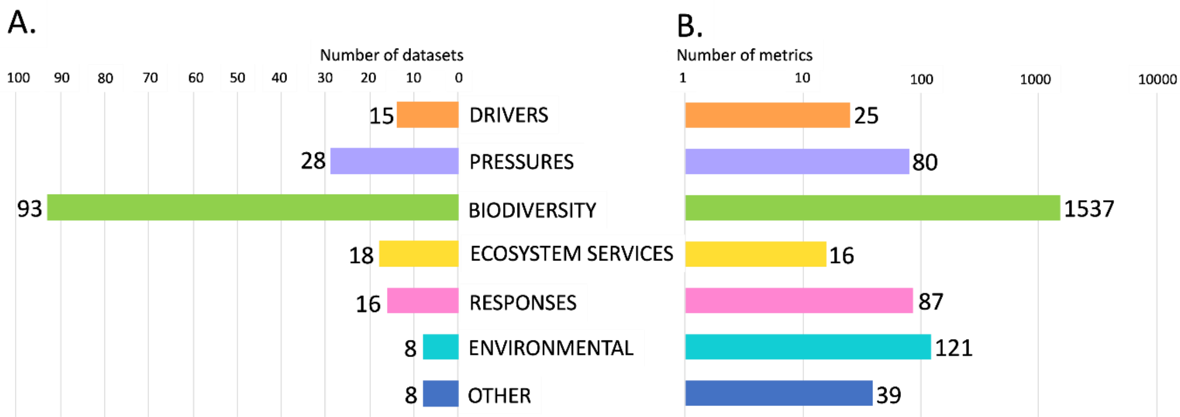


Figure 3. Number of databases and datasets (A.) and variables (B.) documented in the metadatabase per element of the data typology (note the use of a logarithmic scale for the variables). All datasets and databases are listed in Supplementary Table 2 and the corresponding variables in Supplementary Table 3. The typology is available in Supplementary Table 1.

Keeping in mind that some datasets can provide both raw and interpreted, or modeled data (for instance the reporting on the Birds and Habitats Directives), we found that 23% (N=32) of the entries documented in the metadatabase give access to raw data, another 33% (N=46) contained data that are already interpreted, and 40% of the datasets (N=55) provide modelled data. Although the information is not always easily accessible, we were also able to identify dependencies between the entries of the metadatabase, that is datasets that either contain previously produced data, or that use other datasets in their production workflows. For instance, the Elton Traits 1.0 database (Wilman et al., 2014), containing trait data for birds and mammals, is widely used and incorporated in at least 5 other datasets that we identified and documented (Supplementary Table 2). The ESA CCI land cover product is also widely

Funded by the European Union
NaturaConnect receives funding under the European Union's Horizon Europe research and innovation programme under grant agreement number 101060429.

ARPHA Preprints Author-formatted document posted on 27/11/2024. DOI: https://doi.org/10.3897/arphapreprints.e142876

incorporated into workflows to produce data layers at the ecosystem level (including ecosystem services), or on pressures on ecosystems (e.g., relative magnitude of fragmentation). In total, we were able to document dependencies within the metadatabase for 47 entries and additional external dependencies for 35 entries (that is, identified use of datasets that are not documented in this metadatabase).

Tabular data represent nearly half (40%) of the entries, particularly in the form of csv and xls files, while for spatial data, GeoTIFF and NetCDF are the most commonly used formats, with 27% and 8% of the entries respectively. Furthermore, 22 entries in the metadatabase are accessible via an Application Programming Interface, or API, with more describing that these applications are either planned or in progress. The development of dedicated R packages also accompanied the release of several datasets such as “HomeRange” (Broekman et al., 2022), “RCompadre” (Jones et al., 2021) and “RRedList” (Gearty and Chamberlain 2022) to name a few. In terms of accessibility itself, our efforts were by design directed towards publicly available databases and datasets (see Table 1). Concretely, 60% of the documented entries (N=83) made clear the use of the Creative Commons licensing, either CC0 (the most open) or CC-BY, that is with attribution. An additional 23% (N=32) of the data can be labelled as “Open Access”, for instance those following the standard re-use policy of the European Environmental Agency, with some requesting a prior log-in and/or request for data use. Four of the datasets documented in the metadatabase are either closed or have restrictions on reposting and use: the IUCN Red List of Threatened Species (IUCN, 2022), GlobalAnts (Parr et al., 2017), the European Vegetation Archive (EVA, Chytrý et al., 2016), as well as the catalogue of National Red Lists of Threatened Species hosted by ZSL (https://www.nationalredlist.org/).

In the following sections, we discuss more in detail the availability of data across the different elements of the typology.

1.4.2. Data on drivers, pressures and environmental data

We identify and describe 51 databases and datasets containing information that can be used to attribute or model changes in biodiversity and ecosystem services: 15 datasets with data on drivers, 28 on pressures and 8 environmental datasets. Available data labelled as drivers relate mostly to human population and density, consumption and footprint as well as urbanisation and linear infrastructures. All these data are spatially explicit, most of them with a resolution of at least 1 km, with the exception of the dataset of the impact of consumption provided by the EEA at the country level. Most data on drivers are also available for several points in time, with some reconstructions going back to the mid-1970s. Of the 28 entries that correspond to pressures on biodiversity, nearly half (N=12) provide data on habitat change such as wetland loss, soil sealing or more generally Land-Use Land Cover (LULC) change. These entries often overlap with data on ecosystem structure discussed in the following section. Similarly, to data on drivers, data on pressure are largely spatially explicit but most (N=17) do not provide data that is repeated in time. Some datasets, such as the global, European or National IUCN Red Lists of Threatened Species are key for obtaining data on the threats affecting each species individually, although not in a spatially or temporally explicit way. For instance, we could extract that species X is affected by threat Y somewhere within its distribution range in country Z but the dataset will not provide precise location(s). Lastly, we also identified a further 8 datasets that we classify as “environmental data” such as the CHELSA bioclimatic dataset (Brun et al., 2022), the European dataset on sub-canopy microclimate temperature (Haesen et al., 2021), or the LUCAS Topsoils dataset with data on soil chemical properties (https://esdac.jrc.ec.europa.eu/content/lucas-2018-topsoil-data). While relatively small in number, in total, these environmental datasets represent over 120 variables (Figure 2B; Sup. Table 3). Some, such as CHELSA-BIOCLIM+ and ForestClim,

Funded by the European Union

NaturaConnect receives funding under the European Union's Horizon Europe research and innovation programme under grant agreement number 101060429.

provide bioclimatic environmental data that are tailored for biodiversity modelling (Brun et al., 2022).

1.4.3. Data on biodiversity and ecosystem services

We identified and documented a total of 105 entries providing data on biodiversity (N=93) and/or ecosystem services (N=17). Most of the biodiversity-related datasets documented in the metadatabase provide data at the species or ecosystem level (Fig. 4A) with very few datasets in comparison providing data on populations or communities. Precisely describing the taxonomic scope of all the identified datasets goes beyond the scope of the purpose of the metadatabase. Nevertheless, based on the information that was extracted from the metadata of the entries, we observe that plants, birds and mammals are consistently more represented within the datasets (Fig. 4B). We identify 6 datasets that are specifically dedicated to some insect groups, that is bumblebees (Polce et al., 2018; Rasmont and Iserbyt 2010-2014), ants (Parr et al., 2017), butterflies (Shirey et al., 2022) and coleoptera (Jeppsson et al., 2014), but that is not to say that insect data is not present in other datasets with ampler scopes. In particular, we identify 13 datasets that we consider “broad” in taxonomic scope (Supplementary Table 2; Fig. 4C), for instance GlobTherm (Bennett et al., 2018), which provides data on thermal tolerance of algae, plants, fungi and animals, or the PREDICTS dataset (Hudson et al., 2017), which encompasses data on plants, animals, fungi and protozoa. Datasets providing information at the ecosystem level usually make use of common classifications of habitats (e.g., EUNIS, IUCN Habitats) or Land-use and Land Cover (e.g., CORINE; HNV farmland). When datasets cover a single ecosystem type, forest is by far the most represented within the metadatabase (N=9), although the Copernicus Land Service also makes available datasets on grasslands and wetlands. In terms of ecosystem services, 2 datasets provide data on soil protection, 1 on carbon mitigation; 1 on water supply and 1 on wood supply. When considering the EBV and EESV frameworks to classify the entries of the metadatabase (Fig. 4C), we find that most datasets that we documented provide data on Species Populations (N=42) and Species Traits (N=37), followed by Community Composition (N=28) and Ecosystem Structure (N=25).

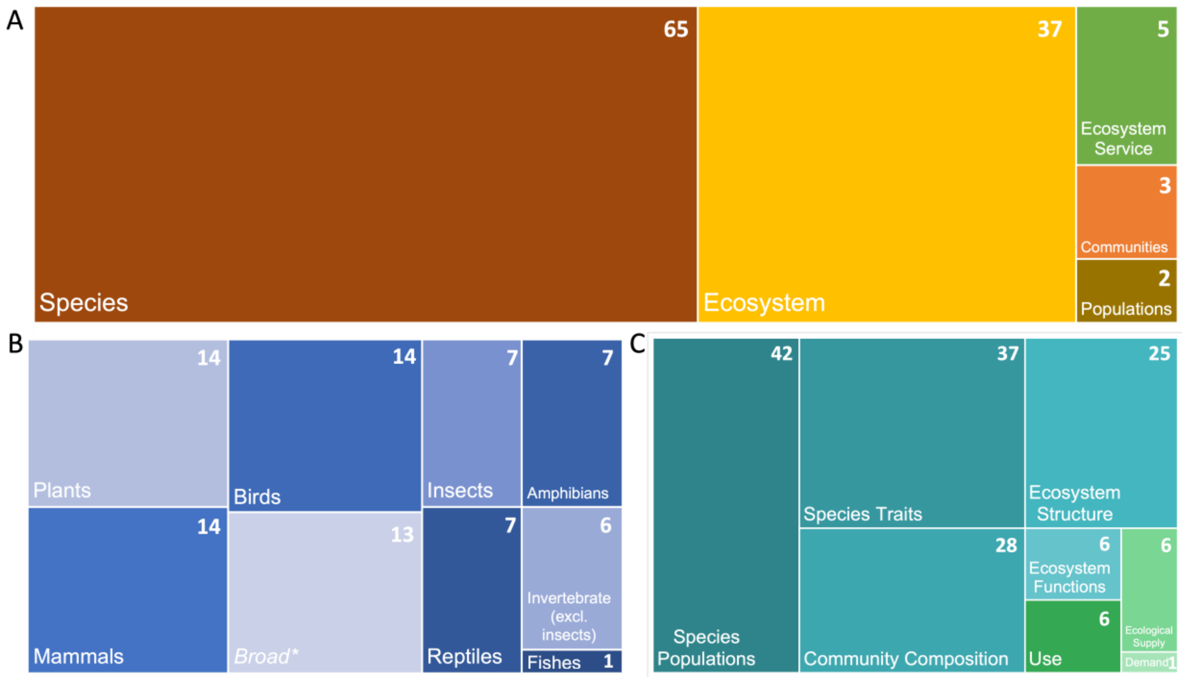


Figure 4. Thematic scope of the entries of the metadatabase. A: Biological entity and taxonomic resolution; B: Coverage of the biological entity. (the “Broad” category is selected for entries with broad taxonomic scope, bringing together data on algae, fungi and animal species for instance); C: Coverage of classes of Essential*

Funded by the European Union
NaturaConnect receives funding under the European Union's Horizon Europe research and innovation programme under grant agreement number 101060429.
NATURA CONNECT

ARPHA Preprints Author-formatted document posted on 27/11/2024. DOI: https://doi.org/10.3897/arphapreprints.e142876

Variables, either for Biodiversity (EBVs – in turquoise tones) or Ecosystem Services (EESVs – in green tones). Note that a single entry in the metadatabase can cover multiple biological entities, taxonomic groups and/or essential variables (see Supplementary Table 2 for the full list of entries in the metadatabase).

Looking more specifically at the data within each entry of the metadatabase (Supplementary Table 3), we found that 1,536 correspond to metrics that can be classified as biodiversity or ecosystem services. We then mapped the 1,390 identified biodiversity metrics to the different components of the Essential Biodiversity Variables framework (Fig. 5), that is to the different EBVs within each of the classes discussed above (see also Fig. 4C and Supplementary Table 1). By doing so, we were able to identify 261 Species Populations data, although this includes a majority of data on species' habitat affinity (N=137) which is more of an ancillary data to species distributions, as well as data on demography (N=34), which is not consistently classified as Species Population EBVs. We could also map 936 Species Traits data to specific EBVs *sensu* (Kissling, Walls, et al., 2018), particularly data on the morphology (N=463) and reproduction (N=206) of species. More than 200 data types could also be mapped to Community Composition EBVs. In contrast, data on ecosystems, whether their structure, or functions, were considerably less represented in the metadatabase, with 91 and 16 different metrics identified for Ecosystem Structure and Ecosystem Functions EBVs respectively. After early consultation with the NaturaConnect consortium, we did not specifically search for data at the Genetic Composition level, which is why they are not included in this analysis. Nevertheless, two of the datasets documented in the metadatabase, EuroBaTrait and the World Spider Traits database, do provide data on genetic diversity and/or differentiation among other EBVs (Supplementary Tables 2 and 3). In most cases we were able to classify whether the data was numerical (65% of the biodiversity data), categorical (21%) or Boolean (14%). In the case of species interaction data, nearly half of the available information, that is 47%, was categorical. Note that the Essential Variables framework for Ecosystem Services (EESVs) does not go beyond classes in its current version (Balvanera et al., 2022) which is why this mapping could only be achieved for biodiversity data.

Funded by the European Union

NaturaConnect receives funding under the European Union's Horizon Europe research and innovation programme under grant agreement number 101060429.

NATURA CONNECT

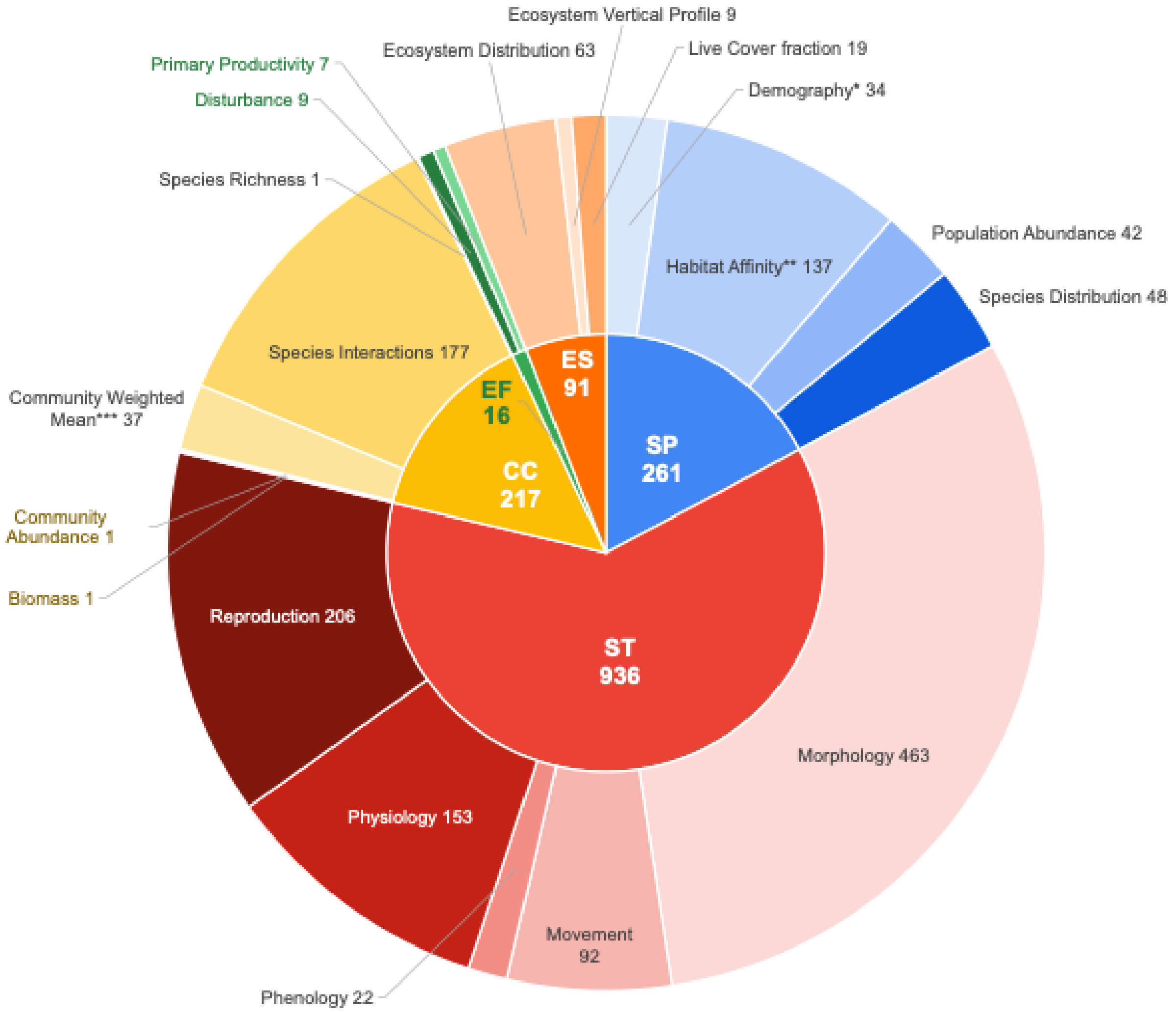


*Figure 5. Repartition of variables within the documented datasets, per EBV class. SP: Species Populations; ST: Species Traits; CC: Community Composition; EF: Ecosystem Functions; ES: Ecosystem Structure. * Demography is not consistently classified within the EBV framework but was here listed under Species Populations; ** Habitat Affinity is not a variable per se but is most relevant within the Species Populations EBV; *** Community Weighted Means are here classified within the Community Composition EBV class.*

1.4.4. Response data: conservation status and actions

Response data identified and documented in the metadatabase can be of three types (Supplementary Table 1): spatial data on area-based conservation, species listing, and ecosystem listing. Specifically, listing corresponds to lists of species or ecosystems established for conservation purposes. For instance, for species, we considered the Red List of Threatened Species (IUCN, 2022), which provides information on the conservation status of species, population trends, as well as the main pressures affecting the species. We also identified the list of invasive alien species of Union concern (EU Regulation No 1143/2014). Response data, which often overlaps with data classified as biodiversity in the metadatabase (e.g., population abundance and ecosystem distributions), can for instance be used to assess the effectiveness of conservation measures and/or progress towards achieving conservation targets such as those of the EU's Biodiversity Strategy for 2030. In total, we documented 16 responses datasets (or databases) encompassing 87 metrics (Supplementary Tables 2 and 3). Apart from the reporting on the Birds and Habitats Directives, most response data was not repeated in time. Area based conservation data include for instance data layers on Nationally Designated Protected Areas (https://www.eea.europa.eu/data-and-maps/data/nationally-designated-areas-national-cdda-17) and measures to improve river connectivity such as fish

Funded by the European Union
NaturaConnect receives funding under the European Union's Horizon Europe research and innovation programme under grant agreement number 101060429.
NATURA CONNECT

passes (Belletti et al., 2020). Species listing data encompass the following datasets: the IUCN Red List of Threatened Species at the global, European and national scales (IUCN, 2022; EEA, 2019), the annexes I to III of the Birds Directive and annexes II, IV and V of the Habitats Directives; as well as the European list of invasive alien species of Union concern (Supplementary Table 2). Listings at the ecosystem level include for instance the EUNIS habitats list for Europe (Chytrý et al., 2020), the dataset on High Nature Value Farmland (EEA, 2017) and the European Red List of Habitats (EEA, 2022).

1.5. Potential uses and future directions

While not systematic by design, our effort of identification and documentation of datasets relevant for NaturaConnect (and beyond) show a wealth of openly available data, from drivers of global change to biodiversity and ecosystem services, and mechanisms implemented to address those changes. The resulting metadatabase (Supplementary Table 2) can serve various purposes by being used as a catalogue of data that could be integrated into workflows of models and tools used in biodiversity research and conservation. For instance, Species Distribution Models could integrate data identified on species occurrences, habitat affinity, movement data, ecosystem distributions, and bioclimatic variables, while the latter two data types could be integrated with various life-history and functional trait data within population dynamics models. Ecosystem Services models could also make use of the identification and integration of data on species traits, on ecosystem functions and on species and ecosystems distributions, coupled with, for instance, data on human density. For more applied uses, the datasets and variables identified in the metadatabase can be integrated to assess the Favorable Reference Values for species (e.g., by crossing occurrence data with trends in population sizes, or density, and traits across species listings) (Jung et al. 2024). This can further be integrated with ecosystem services mapping and distributions of protected areas for spatial planning assessments. Spatially explicit data on species and ecosystems could also be integrated with species trait data to map the structural and functional connectivity across Europe (Moreira et al., 2024).

Nevertheless, while most of the datasets are findable and accessible, they are often shared with different file formats, that require prior download and often need an additional step of standardisation (for instance taxonomic for species data) which act as bottlenecks to the smooth integration of data within open and repeatable workflows. A more generalised use of data repositories and portals that use a common metadata standard, such as the EBV portal, would help streamline workflows and provide easier access to the data for end users. Yet, an additional hurdle comes from the difficulty to easily identify dependencies within datasets and hence an increased risk for redundancy, overestimations and circularity. These limitations could be addressed by sharing the protocols implemented in the process of producing a given dataset, that is to list all input data used and to document the different underlying steps of data preparation, integration, and processing.

Our approach of data identification and documentation also presents its own limitations. The scoping for relevant entries in the metadatabase is comprehensive within the framework of the NaturaConnect project but cannot be considered as spatially or thematically exhaustive. Some data was ignored intentionally such as national atlases or data on genetic composition. In this regard, the metadatabase should be understood as a description of the data that is available, but is not, in its current form, adequate for a gap analysis. For this, we would need a more in-depth and comprehensive review of biodiversity-related data needs and a more systematic review of data availability across the elements of the adapted DPSIR framework (Fig. 2).

Funded by the European Union

NaturaConnect receives funding under the European Union's Horizon Europe research and innovation programme under grant agreement number 101060429.

2. Part B – Gap analysis of knowledge on biodiversity data for Europe

2.1. Introduction: Biodiversity shortfalls in Europe, a baseline to build the TEN-N

A comprehensive knowledge of biodiversity is an essential element for developing an ecologically representative and resilient Trans-European Nature Network (TEN-N), ensuring effective protection of ecosystems across the continent (European Environment Agency, 2023). At the global level, Europe shows a good coverage of biodiversity data, with many biodiversity initiatives and millions of records obtained through citizen science, scientific databases, and publications (Garcia-Rosello et al., 2023; Oliver et al., 2021). Despite Europe being one of the continents with the most comprehensive biodiversity data, a number of biases in biodiversity knowledge exist (Hortal et al., 2015; see also part A of this report). For example, western Europe is usually characterised by data-rich areas, while the eastern part of the continent is often under-sampled due to accessibility and socio-economic factors (Meyer et al., 2015; Moersberger et al., 2024). Similarly, while some taxa have been extensively studied (e.g., birds and large mammals), others have received little attention (e.g., reptiles and amphibians). Even among mammals and birds, most data belong to charismatic, umbrella species, or species relevant for wildlife management and conservation (Vergara-Asenjo et al., 2023).

In the last two decades, shortfalls in biodiversity knowledge have received attention to identify possible biases in our understanding and conservation planning, and priorities for data collection (Pollock et al., 2020). Hortal et al. (2015) identified seven biodiversity shortfalls (Table 2):

Table 2. Definitions (and original references) for the seven main shortfalls of biodiversity knowledge (table edited from Hortal et al., 2015).

Shortfall	Aspect of biodiversity	Definition
Linnean	Species	Most of the species on Earth have not been described and catalogued (Brown & Lomolino 1998); this concept can be extended to extinct species
Wallacean	Geographic distribution	Knowledge about the geographic distribution of most species is incomplete; it is inadequate at all scales most of the time (Lomolino 2004)
Prestonian	Populations	Data on species abundance and population dynamics in space and time are often scarce (Cardoso et al. 2011)
Darwinian	Evolution	Lack of knowledge about the tree of life and the evolution of species and their traits (Diniz-Filho et al. 2013)
Raunkiæran	Functional traits and ecological functions	Lack of knowledge about species' traits and their ecological functions
Hutchinsonian	Abiotic tolerances	Lack of knowledge about the responses and tolerances of species to abiotic conditions (i.e., their scenopoetic niche; redefined from Cardoso et al. 2011)
Eltonian	Ecological interactions	Lack of knowledge on species' interactions and these interactions' effects on individual survival and fitness

The Linnean shortfall refers to the lack of species described and classified in relation to the total number of existing species (Brown & Lomolino, 1998); the Wallacean shortfall relates to the gap in the spatial distribution of species (Lomolino, 2004); the Prestonian shortfall focuses

Funded by the European Union
NaturaConnect receives funding under the European Union's Horizon Europe research and innovation programme under grant agreement number 101060429.
NATURA CONNECT

on the knowledge of population dynamics and their abundances (Cardoso et al., 2011); the Darwinian shortfall represents the ignorance for evolution processes of species and their relationship (Diniz-Filho et al., 2013); the Raunkiæran shortfall analyses the lack of knowledge for the functional traits and ecological functions (Hortal et al., 2015); the Hutchinsonian shortfall is the deficiency of notions regarding species limits for abiotic factors (Cardoso et al., 2011); and the Eltonian shortfall shows the scarcity of data for species ecological interactions (Hortal et al., 2015). These shortfalls are likely driven by both accessibility and socio-economic factors whose investigation is crucial to highlight priority areas for inventory completeness addressing the knowledge gap on biodiversity data.

The objective of this part of the research was to investigate the existing gaps of knowledge in biodiversity data for Europe highlighting (geographically and thematically) the areas that would most benefit from targeted data collection. Below we present an analysis aimed at identifying knowledge gaps on species distribution (Wallacean shortfall), trait (Raunkieran shortfall), and trophic interaction data (Eltonian shortfall) of European tetrapods. We further explore the socio-economic, ecological and geographic drivers of the Wallacean shortfall. These knowledge gap analyses can be used as a baseline to identify priority areas where the cost-benefit of additional data collection would be highest at the EU level.

2.2. Materials and methods

2.2.1. Distribution data and key socio-economic drivers

For all our analyses, we considered a study area covering the entire geographical extent of Europe, going from Spain and Portugal in the west, to the Urals in the east; we included Turkey (geographically part of Asia) to cover the entire northern Mediterranean shoreline.

For the Wallacean shortfall, we downloaded all available occurrences for terrestrial vertebrates (mammals, regularly breeding birds, reptiles, and amphibians) occurring in the study area from GBIF (https://www.gbif.org/) and iNaturalist (https://www.inaturalist.org/) in December 2023 (Table 1 in Annex I). Although most of the iNaturalist data are included in the GBIF database, the data upload is not always up to date. Therefore, we decided to use both databases to include all available occurrences at the current time. We combined the two databases harmonising the taxonomy according to the GBIF nomenclature (e.g., occurrences of *Alytes almogavarii obstetricans* downloaded from iNaturalist we re-assigned to *Alytes almogavarii* as specified in GBIF), and we removed all introduced and pet species, and all erroneous data referring to what is available from the IUCN Red List (IUCN, 2024) or class-specific sources (e.g., Birds of the World and Reptile Database; Cornell Lab of Ornithology, 2019; Uetz et al., 2021) when the IUCN Red List was incorrect (e.g., *Xenopus laevis* native in Italy according to IUCN). We filtered out all occurrences with unranked taxon (e.g., only genus) and data with unknown or unverified basis of record (occurrence status) or other data issues (e.g., presumed swapped coordinates). We also removed all occurrences without coordinates or with an uncertainty higher than 10,000 meters, and all duplicated records considering the species name, coordinates, year, and the basis of records. Furthermore, we excluded all fossil records and data collected before 1970 to provide a more reliable description of the current knowledge gap on European tetrapods' distribution.

To evaluate the knowledge gap on European tetrapods' distribution, we used the occurrence data to calculate the inventory completeness within spatial units of approximately 10 km^2. Although several national and European checklists already exist for terrestrial tetrapods (e.g., Speybroeck et al., 2020), we did not assume that all known species are already collected in the study area. Therefore, we decided to estimate statistically the saturation of data collection within each pixel of our study area through a species accumulation curve (SAC). For this

Funded by the European Union
NaturaConnect receives funding under the European Union's Horizon Europe research and innovation programme under grant agreement number 101060429.

purpose, for each class, we calculated a species accumulation curve (SAC) in which the cumulative number of species is plotted against the cumulative number of database records which are used as a proxy of sampling effort (Colwell et al., 2012). The degree of completeness of each spatial unit is thus given by the slope of the relationship between the number of species and the number of records. We calculated the SAC using the exact estimator (Ugland et al., 2003) available in the R package 'KnowBR' (Lobo et al., 2018) performing all analyses at the level of bioregions (European Environment Agency, 2016). We then merged the results of each biogeographic region analysis to map the percentage of completeness at the continental scale for each tetrapod class.

To explore the drivers of the knowledge gap in species distribution, we modelled the completeness estimate with accessibility factors (e.g., road density, distance to universities and museums), land cover variables, and socio-economic factors (e.g., GDP per capita, human population density; Table 2 in Annex III) considering the metadatabase built in the previous section as well as other online sources.

Accessibility factors are expected to be important drivers of the knowledge gap in species distribution, with remote areas (e.g., tropical areas) typically resulting in being under-sampled (García-Roselló et al., 2023). Indeed, the road density as the distance to the main cities, museums, or protected areas may affect the degree of sampling effort in certain areas. In our analysis, we considered four layers of road density (highways, primary roads, secondary roads, tertiary roads) calculated as the sum of linear kilometres within each 10 km^2 cell from the vector dataset GRIP (Global Roads Inventory Dataset; Meijer et al., 2018). We included two layers of distance to protected areas (WDPA; UNEP-WCMC & IUCN, 2024) and to museums, universities, and colleges (OSM data; OpenStreetMap, 2017). As accessibility factors, we also included three layers of travel time to the nearest cities with a different number of people, from small to large (small: cities with > 10,000 people; medium: cities with > 100,000 people; large: cities with > 1,000,000 people; Nelson et al., 2019).

Different land cover types could also lead to differential sampling efforts due to research interest as well as different conditions (e.g., visibility in open versus closed areas). We included three land cover variables (forest, wetland, cropland) as the percentage of the land cover class within each cell (Global CCI-LC 2022; ESA, 2017).

Since socio-economic factors may be correlated with inventory completeness (e.g., financial and institutional resources; Meyer et al., 2015), we considered seven socio-economic variables in our analysis. We downloaded the human population density (number of people) dataset from the SEDAC database (https://sedac.ciesin.columbia.edu; Gao, 2020) at 30 arc-seconds resolution. As a measure of the country's achievements, we obtained the human development index (HDI) from the United Nations Development Programme (UNDP, 2024). From the World Bank database (The World Bank Group, 2024), we obtained the following indexes: GDP per capita, N° scientific papers, N° researchers, educational attainment (at least Bachelor's or equivalent) in population 25+ (%), Research and development expenditure (% of GDP) (Table 2 in Annex I). All variables were resampled to 10 km^2 resolution. To exclude collinearity among predictors, we excluded all variables with a Pearson correlation ≥0.7 (Figure 1 in Annex III).

To identify the main drivers of the geographic pattern of the Wallacean shortfall, for each tetrapod class we ran a random forest model using all the uncorrelated predictors illustrated above ('randomForest' R package; 1,000 trees; Breiman, 2001). To avoid spatial autocorrelation, for each class, we ran a model with 500 subsamples of 1,000 records each.

Funded by the European Union
NaturaConnect receives funding under the European Union's Horizon Europe research and innovation programme under grant agreement number 101060429.

We measured the mean variable importance among the 500 iterations by using the percentage increase in mean squared error following the permutation of each predictor variable.

2.2.2. Functional traits data

For the Raunkiæran shortfall, following the same taxonomy used for the Wallacean shortfall (GBIF nomenclature), we considered all species occurring in the study area for which the IUCN geographic range polygon was available. For each species, we collected data on morphological, reproduction, and life-history traits from peer-reviewed published databases (Table 3; R package 'traitdata') and species' range polygons from the IUCN Red List (IUCN, 2024). We harmonised taxonomically both range polygons and trait databases with our species list, and we merged several trait databases focusing on four biological traits (body size, litter/clutch size, sexual maturity age, trophic guild). For IUCN range polygons, we considered only polygons with origin 'native' or 'reintroduced', with presence 'extant', 'possibly extant' or 'possibly extinct', and with seasonality 'resident' or 'breeding season'.

To obtain a map of trait data completeness, we stacked the species' range polygons, assigning to the polygon the number of traits available for the species divided by the total number of species within 10 km^2 cells.

Funded by the European Union

NaturaConnect receives funding under the European Union's Horizon Europe research and innovation programme under grant agreement number 101060429.

Table 3: Trait databases considered for the Raunkiæran shortfall analysis (available at: https://doi.org/10.5281/zenodo.14008205).

Database	Taxonomic coverage	Reference
AmnioteDB	Amphibians, birds, mammals, reptiles	Myhrvold, N. P., Baldridge, E., Chan, B., Sivam, D., Freeman, D. L., & Ernest, S. M. (2015). An amniote life-history database to perform comparative analyses with birds, mammals, and reptiles: Ecological Archives E096-269. Ecology, 96(11), 3109-3109. https://esapubs.org/archive/ecol/E096/269/
Amphibian traits database	Amphibians	Huang, N., Sun, X., Song, Y., Yuan, Z., & Zhou, W. (2023). Amphibian traits database: A global database on morphological traits of amphibians. Global Ecology and Biogeography, 32(5), 633-641.
AmphiBIO	Amphibians	Oliveira, B. F., São-Pedro, V. A., Santos-Barrera, G., Penone, C., & Costa, G. C. (2017). AmphiBIO, a global database for amphibian ecological traits. Scientific data, 4(1), 1-7.
AnAge	Amphibians, birds, mammals, reptiles	De Magalhaes, J. P., & Costa, A. J. (2009). A database of vertebrate longevity records and their relation to other life-history traits. Journal of evolutionary biology, 22(8), 1770-1774. http://genomics.senescence.info/species/
AvianBodySize	Birds	Lislevand, T., Figuerola, J., & Székely, T. (2007). Avian body sizes in relation to fecundity, mating system, display behaviour, and resource sharing: Ecological archives E088-096. Ecology, 88(6), 1605-1605.
AVONET	Birds	Tobias, J. A., Sheard, C., Pigot, A. L., Devenish, A. J., Yang, J., Sayol, F., ... & Schleuning, M. (2022). AVONET: morphological, ecological and geographical data for all birds. Ecology Letters, 25(3), 581-597.
Bird_behav	Birds	Tobias, J. A., & Pigot, A. L. (2019). Integrating behaviour and ecology into global biodiversity conservation strategies. Philosophical Transactions of the Royal Society B, 374(1781), 20190012.
COMBINE	Mammals	Soria, C. D., Pacifici, M., Di Marco, M., Stephen, S. M., & Rondinini, C. (2021). COMBINE: a coalesced mammal database of intrinsic and extrinsic traits.
EAmphDB	Amphibians	Trochet, A., Moulherat, S., Calvez, O., Stevens, V. M., Clobert, J., & Schmeller, D. S. (2014). A database of life-history traits of European amphibians. Biodiversity Data Journal, (2).
EltonTraits	Mammals, birds	Wilman, H., Belmaker, J., Simpson, J., de la Rosa, C., Rivadeneira, M. M., & Jetz, W. (2014). EltonTraits 1.0: Species-level foraging attributes of the world's birds and mammals: Ecological Archives E095-178. Ecology, 95(7), 2027-2027.
Eubirds	Birds	Storchová, L., & Hořák, D. (2018). Life-history characteristics of European birds. Global Ecology and Biogeography, 27(4), 400-406.
EuroBaTrait	Bats	Froidevaux, J. S., Toshkova, N., Barbaro, L., Benítez-López, A., Kerbiriou, C., Le Viol, I., ... & Razgour, O. (2023). A species-level trait dataset of bats in Europe and beyond. Scientific data, 10(1), 253.
LHT_ERep	Reptiles	Grimm, A., Ramírez, A. M. P., Moulherat, S., Reynaud, J., & Henle, K. (2014). Life-history trait database of European reptile species. Nature Conservation, 9, 45-67.
LizardTraits	Reptiles	Meiri, S. (2018). Traits of lizards of the world: Variation around a successful evolutionary design. Global ecology and biogeography, 27(10), 1168-1172.
panTHERIA	Mammals	Jones, K. E., Bielby, J., Cardillo, M., Fritz, S. A., O'Dell, J., Orme, C. D. L., ... & Purvis, A. (2009). PanTHERIA: a species-level database of life history, ecology, and geography of extant and recently extinct mammals. Ecological Archives E090-184. Ecology, 90(9), 2648-2648. https://tinyurl.com/5czp2t33

Funded by the European Union
NaturaConnect receives funding under the European Union's Horizon Europe research and innovation programme under grant agreement number 101060429.
NATURA CONNECT

2.2.3. Trophic interaction data

Interaction data are often insufficient or biased given the significant efforts to detect direct or indirect species interactions, together with issues linked to collecting methods (different observers or recording methods), or variability of trophic interactions among different populations (of the same species). For this reason, we used a metaweb of potential trophic interactions to map the Eltonian completeness of European tetrapods (Maiorano et al., 2020; Morales-Castilla et al., 2015).

We considered the same study area and list of species used in the previous analyses (2.2.1 and 2.2.2), and we collected the metaweb data from TETRA-EU 1.0 (Maiorano et al., 2020) and the species' geographic range polygon from the IUCN Red List (IUCN, 2024). We harmonised the taxonomy of all data sources, and we filtered species' range polygon as in the previous analysis.

To obtain a map of trophic interaction data completeness, we stacked the species' range polygon with trophic data divided by the total number of species within 10 km^2 cells.

2.3. Results and discussion: the state of biodiversity knowledge in Europe

2.3.1. Wallacean shortfall: patterns and drivers

Combining GBIF and iNaturalist databases after data cleaning, we obtained a total of 169,556,617 occurrences for 1,150 species of European tetrapods. In particular, we obtained 1,390,917 occurrences for 119 species of amphibians, 409,512 occurrences for 234 species of reptiles, 5,820,392 occurrences for 249 species of mammals, and 161,935,796 for 548 species of breeding birds. The data collected already showed a strong taxonomic bias, with breeding birds with a much higher number of occurrences compared to the other tetrapod classes. This uneven availability of distribution data among tetrapods may be due to both the species detectability, with diurnal and large species more easily detected and recognised (e.g., wild boar) compared to nocturnal and small species (e.g., bats), and species charisma, as usually large mammals and birds are more attractive to ecotourism and citizen-science programs (e.g., birdwatching; Mammola et al., 2023).

Although the European continent showed good coverage of biodiversity data from a global point of view (e.g., García-Roselló et al., 2023; Meyer et al., 2015), our analysis showed some differences at the continental level. The completeness analysis highlighted substantial gaps of knowledge, with strong taxonomic but also geographic bias (Fig. 6).

Funded by the European Union
NaturaConnect receives funding under the European Union's Horizon Europe research and innovation programme under grant agreement number 101060429.

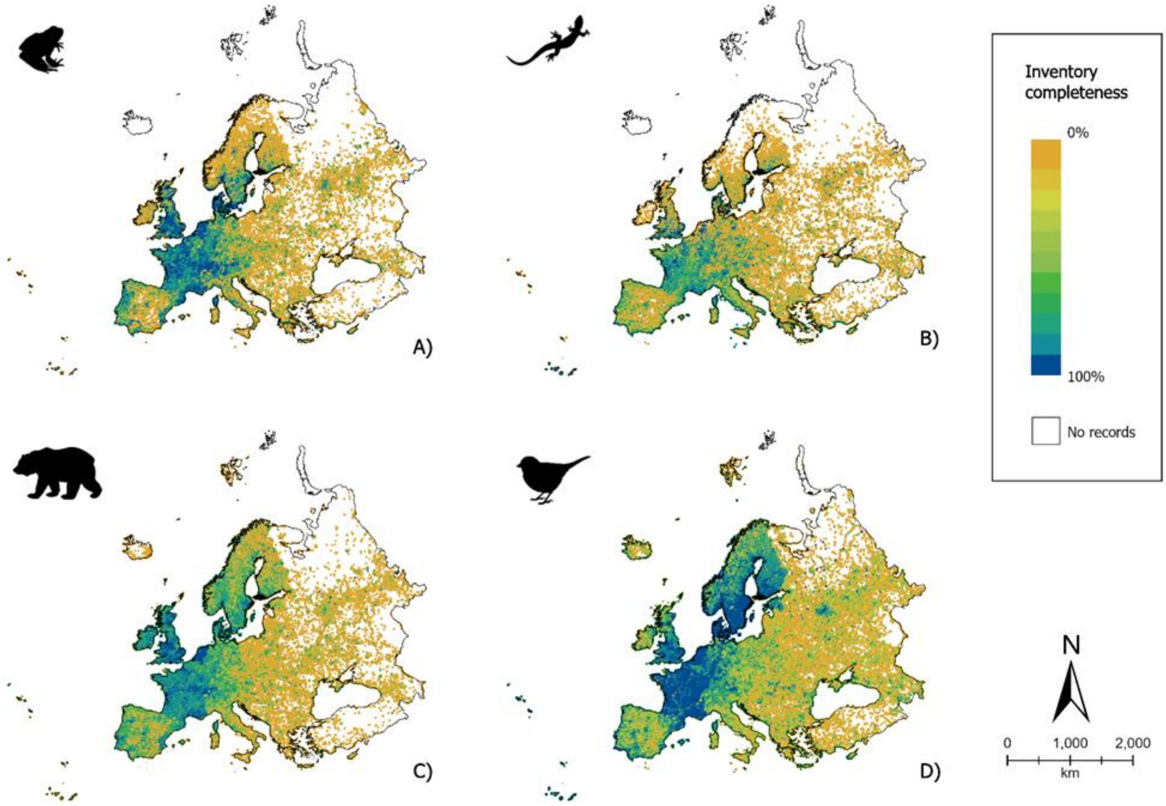


Figure 6. Map of inventory completeness (Wallacean shortfall) of amphibians (A), reptiles (B), mammals (C), breeding birds (D). Colour scale represents the percentage of completeness within 10 km^2 cells.

For all tetrapods, results show a clear longitudinal gradient of biodiversity inventory completeness from east to west Europe. In particular, the inventory completeness for both amphibians and reptiles showed values close to or equal to zero from the Balkans and Poland, to the eastern border of the study area, with some exceptions around Moscow, Helsinki, and Carpathians (Fig. 6 A and B). The knowledge gap of mammal distribution showed a similar pattern but with better coverage of Scandinavia and Ireland (Fig. 6 C). Similarly, the inventory completeness of birds showed a longitudinal gradient but with higher values in Scandinavia as well as in Eastern Europe (≤40%; Fig. 6 D). These discrepancies in inventory completeness may be driven by socio-economic and historical factors, with countries that have historically invested extensively in naturalistic research (e.g., France, UK) and others where naturalistic research efforts have been more recent (e.g., Turkey, Kazakhstan; Zizka et al., 2020). At the same time, the degree of geographical accessibility may affect the feasibility of the sampling, as confirmed by our analysis of the drivers of inventory completeness (Geldmann et al., 2016).

For all tetrapods, the most important drivers of inventory completeness were socio-economic variables, with values of increase in Mean Squared Error (%) ranging from 24.0% (in reptiles) to 46.15% (in birds) for the GDP per capita (Fig. 7).

Funded by the European Union
NaturaConnect receives funding under the European Union's Horizon Europe research and innovation programme under grant agreement number 101060429.

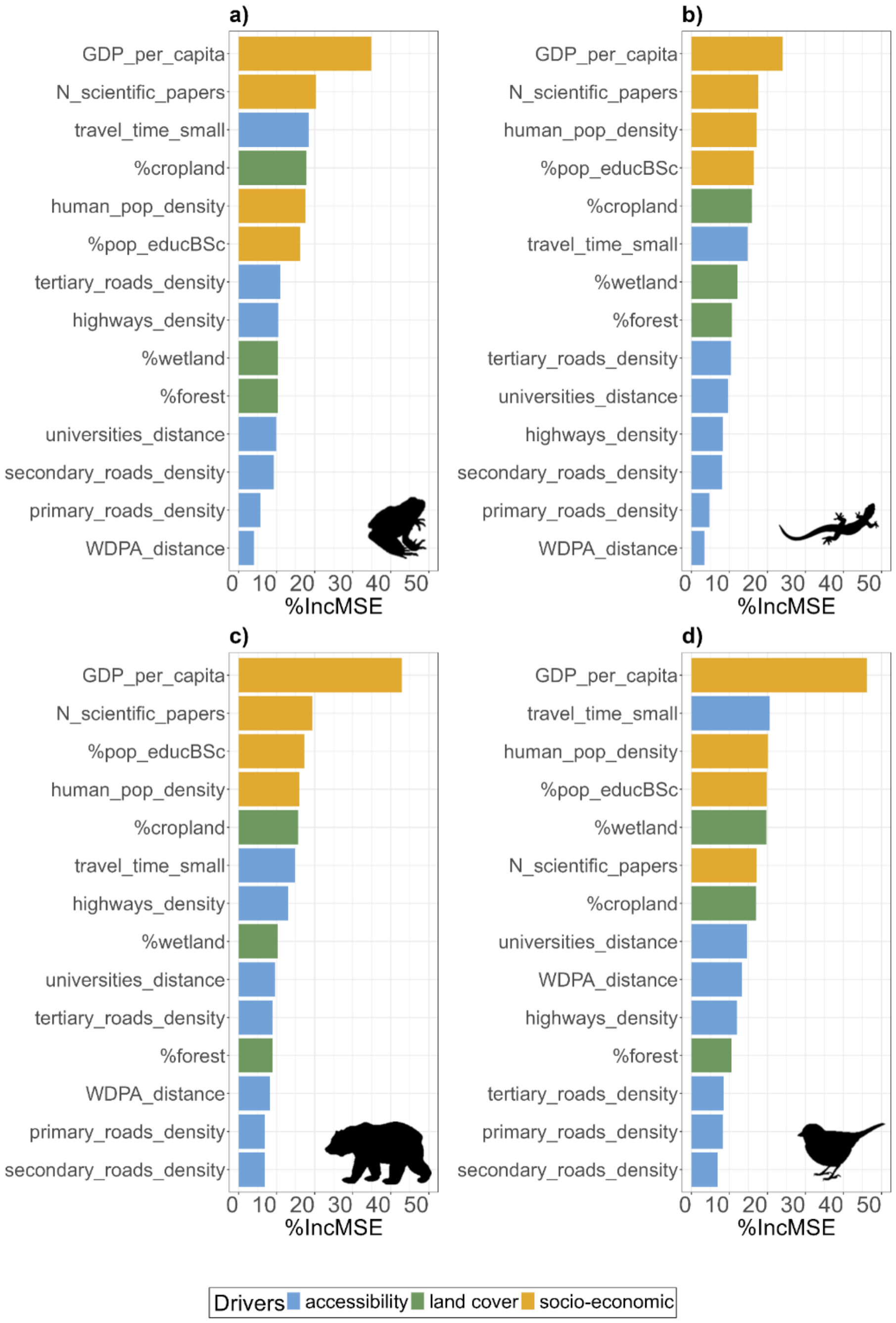


Figure 7: Variable importance showing increase in Mean Squared Error (%IncMSE) when each variable is permuted in the random forest model. a) Amphibians; b) Reptiles; c) Mammals; d) Breeding birds. Abbreviated names in y-axis: WDPA_distance = distance to protected areas, %pop_educBSc = percentage of population ages 25 and over that attained or completed Bachelor's or equivalent. See Table 2 in Annex III for more information on variables.

In particular, the percentage of data completeness showed an increase at increasing values of GDP per capita for all tetrapod classes, highlighting the importance of the country's wealth as a starting point to invest in basic research as data collection for biodiversity conservation (Fig. 8).

Funded by the European Union
NaturaConnect receives funding under the European Union's Horizon Europe research and innovation programme under grant agreement number 101060429.

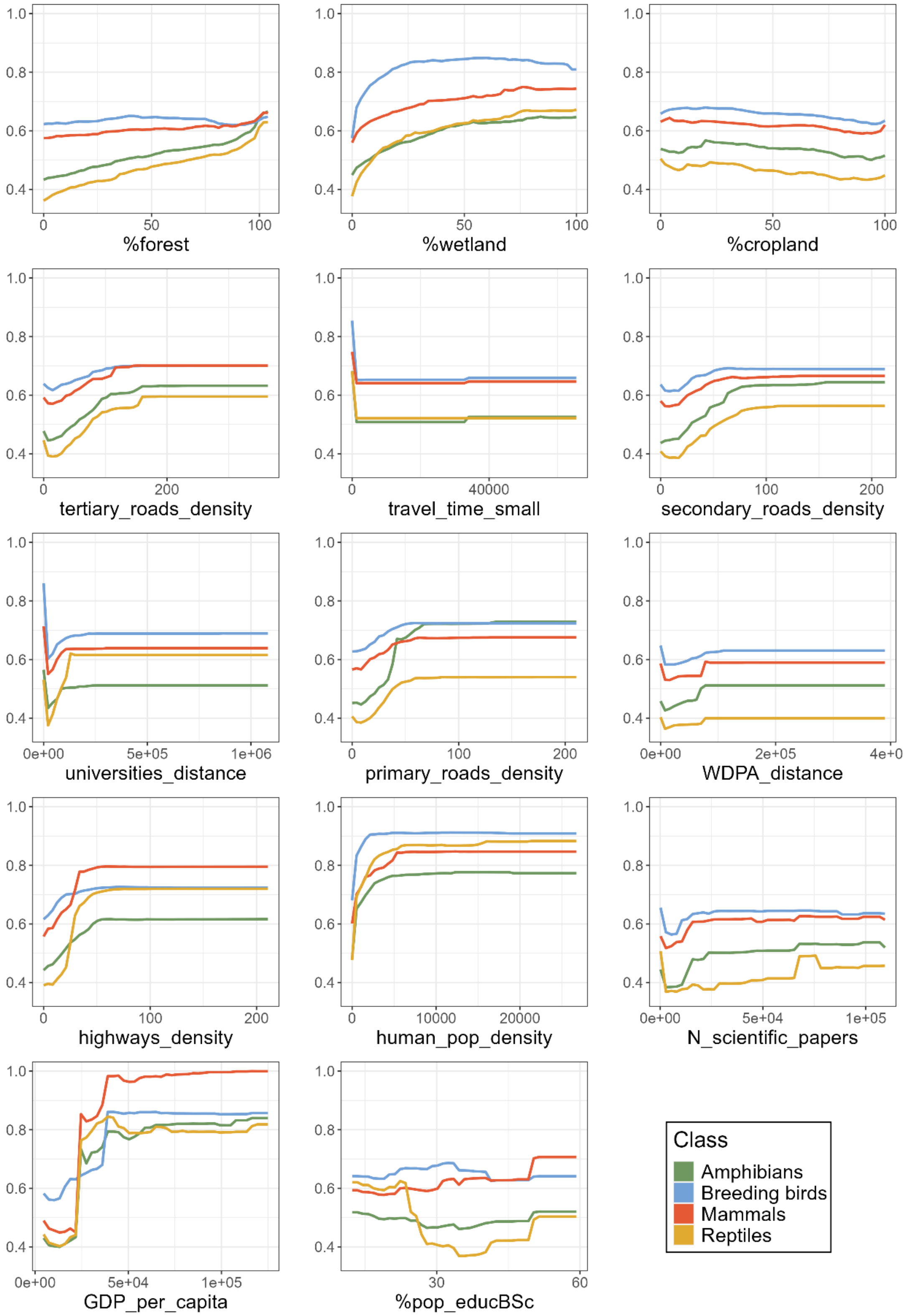


Figure 8: Partial dependence plot of all variables from each random forest model of European tetrapods. In the y-axis, values of completeness were normalized to make the four classes comparable. Abbreviated names in the x-axis: WDPA_distance = distance to protected areas, %pop_educBSc = percentage of population ages 25 and over that attained or completed Bachelor's or equivalent. See Table 2 in Annex III for more information on variables.

Along the same line, the data completeness showed an increase at increasing values of the number of papers published within the country immediately reaching a plateau at low values, indicating that the produced research in the form of 'scientific publication' is not always a good

Funded by the European Union
NaturaConnect receives funding under the European Union's Horizon Europe research and innovation programme under grant agreement number 101060429.

proxy of the biodiversity knowledge of a country. In addition, the ‘%pop_educBSc’ showed almost no effect for all tetrapods except for reptiles for which it showed a decrease of completeness at increasing percentages of educated people (at least BSc). The latter result may be driven by the absence of reptile species within countries with a high number of Bachelor's Degree graduates (e.g., Scandinavian countries).

On the contrary, for all tetrapods, the data completeness showed an increase at increasing values of human population density getting up to a plateau at low values (2,500-5,000 people; Fig. 8). Although the species observations may be facilitated in more natural areas, it is also true that the probability of inventory completeness is higher in the proximity of cities (Fig. 6) as more people may be available for sampling (Bowler et al., 2022; Hugo & Van Rensburg, 2008).

With some differences between the four classes, other important drivers were represented by accessibility factors (travel time to ‘small’ cities) as well as land cover variables (Fig. 7). The percentage of completeness showed a decrease with increasing values of travel time to the nearest (‘small’) cities, immediately reaching a plateau after which the completeness remained constant. The remoteness of a natural place may reduce the possibility of being reached by researchers, naturalists, and even tourists, thus negatively influencing the species observation (Narváez-Gómez et al., 2021). The importance of accessibility factors in data completeness was observed in previous studies from narrow to global scale, in particular highlighting the important role of roads in simplifying access to natural ecosystems such as wetlands (Girardello et al., 2018; Meyer et al., 2015). Indeed, the latter showed an increase of completeness at increasing coverage of wetlands for all tetrapods, although it was particularly important for birds’ completeness. Besides that, wetlands are important drinking areas for most species among tetrapods, these areas are often used as birdwatching sites facilitating high levels of inventory completeness of birds (Prenda et al., 2024).

2.3.2. Raunkiæran shortfall

For the Raunkiæran shortfall, we considered a total of 1,188 species (120 amphibians, 251 reptiles, 292 mammals, 525 birds). Trait data coverage was unbalanced both among traits and taxa (Fig. 9).

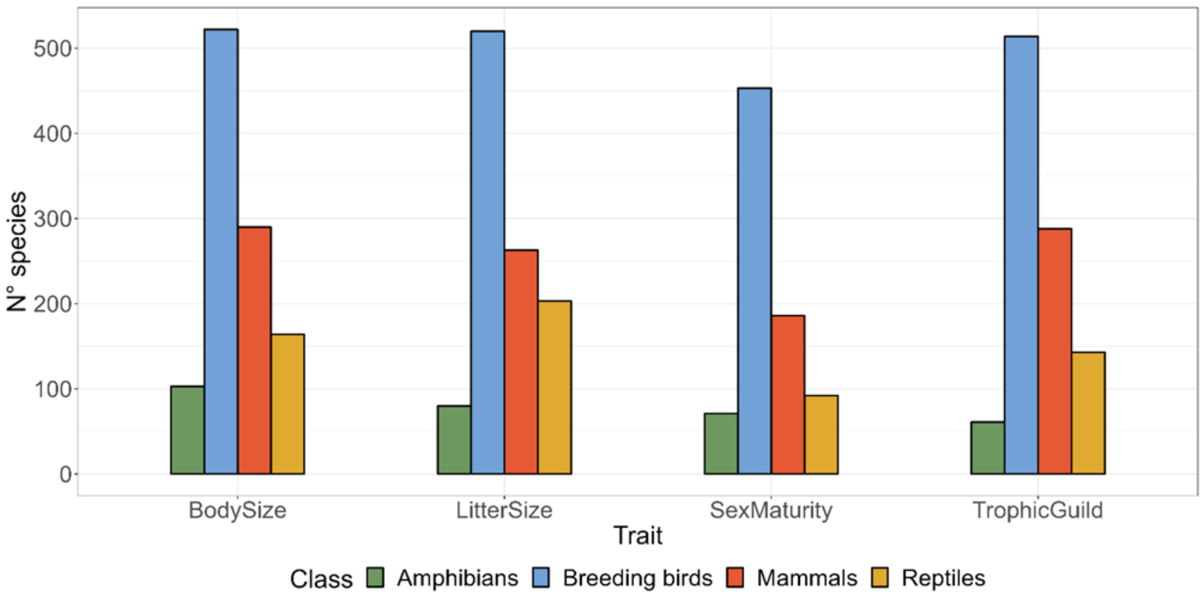


Figure 9: Trait data coverage of the European tetrapods according to four biological traits. The y-axis represents the number of species for which the trait was available.

Funded by the European Union
NaturaConnect receives funding under the European Union's Horizon Europe research and innovation programme under grant agreement number 101060429.

Among all classes, birds showed the highest trait data coverage, with 86% of birds with all selected traits available, while reptiles showed the lowest coverage with only 29% of species with all trait data available. Among the selected traits, body size was the most available trait for all classes except for reptiles for which the clutch size showed a higher coverage compared to the body size (81% and 65% of the species). On the contrary, sexual maturity age was the least available trait for all classes excluding amphibians for which the trophic guild showed a lower coverage (59% and 51% of the species). The inequality in the trait coverage reflects the taxonomic bias founded in the Wallacean shortfall. In fact, the main attention is focused on charismatic species such as large mammals or birds, for which a large number of trait databases are available if compared to reptiles and amphibians (Trimble & van Aarde, 2010; Troudet et al., 2017).

The trait data completeness maps confirmed a strong bias among taxa, with birds and mammals almost completely sampled across the study area and according to the selected traits, while amphibians and reptiles showed less availability of trait data (Fig. 10).

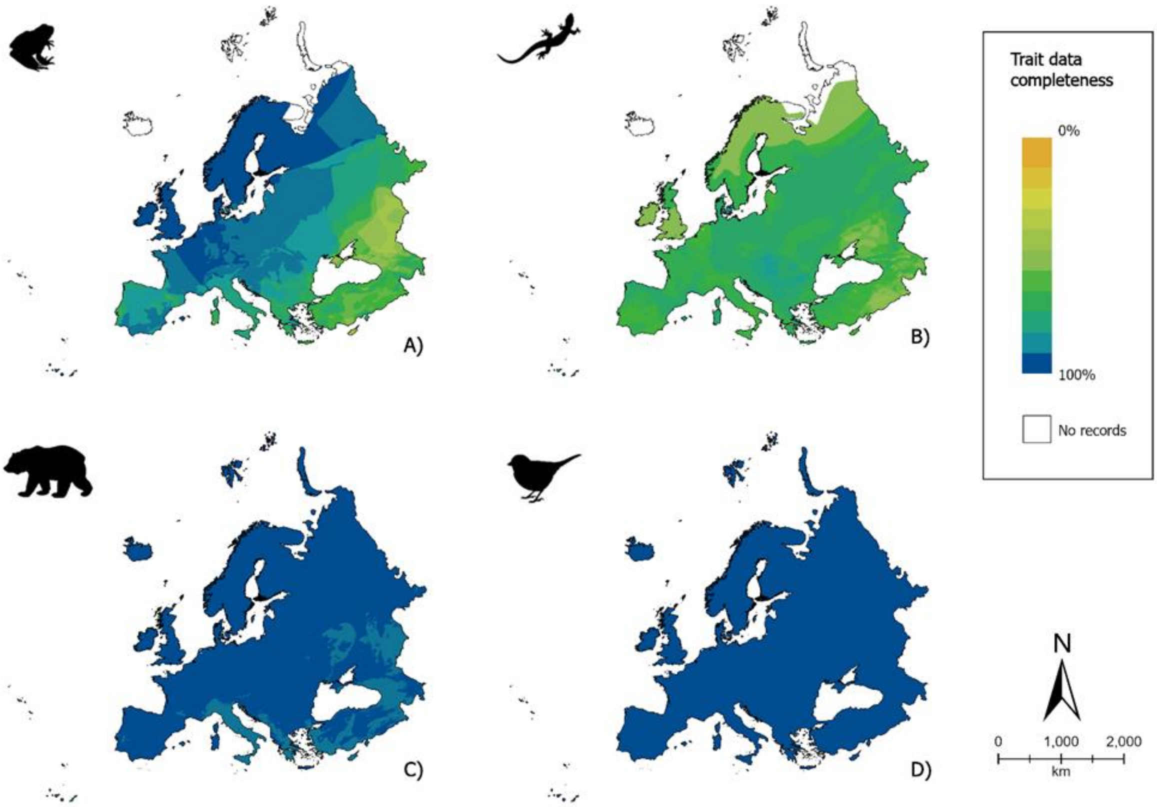


Figure 10: Map of trait data completeness (Raunkiæran shortfall) within 10 km^2 cells showing the completeness considering four biological traits (body size, litter/clutch size, sexual maturity age, trophic guild) for amphibians (A), reptiles (B), mammals (C), and breeding birds (D). Zero values (orange) indicate no available traits for species occurring within the 10km^2 cells, while 100% values indicate that all four selected traits are available for species occurring within the 10km^2 cells according to the IUCN range polygons.

In particular, the areas with the main gap of knowledge on trait data of amphibians were located almost in southern Europe (Cyprus, Greece, Turkey, Georgia, Azerbaijan), while reptiles were under-sampled in both southern (south of Iberian Peninsula and Italy, Turkey, Georgia) and northern Europe (Irland, United Kingdom, Scandinavian, north of Russia; Figure

Funded by the European Union
NaturaConnect receives funding under the European Union's Horizon Europe research and innovation programme under grant agreement number 101060429.

9 A and B), confirming that reptiles are the most understudied taxon among tetrapods (Etard et al., 2020). The lack of knowledge in the eastern part of the study area is linked to several factors, both historical (national conflicts), socio-economic (social inequality), and biogeographic (high level of endemism), that make these remote areas difficult to investigate.

2.3.3. Eltonian shortfall

For the Eltonian shortfall, we considered a total of 1,188 species (120 amphibians, 251 reptiles, 292 mammals, 525 birds). Metaweb data were available for 79% of amphibians, 78% of reptiles, 91% of mammals, and 93% of birds, confirming the previous taxonomic bias.

The trophic interaction data completeness maps showed a comparable pattern to trait data completeness, with birds and mammals better sampled than amphibians and reptiles (Fig. 11).

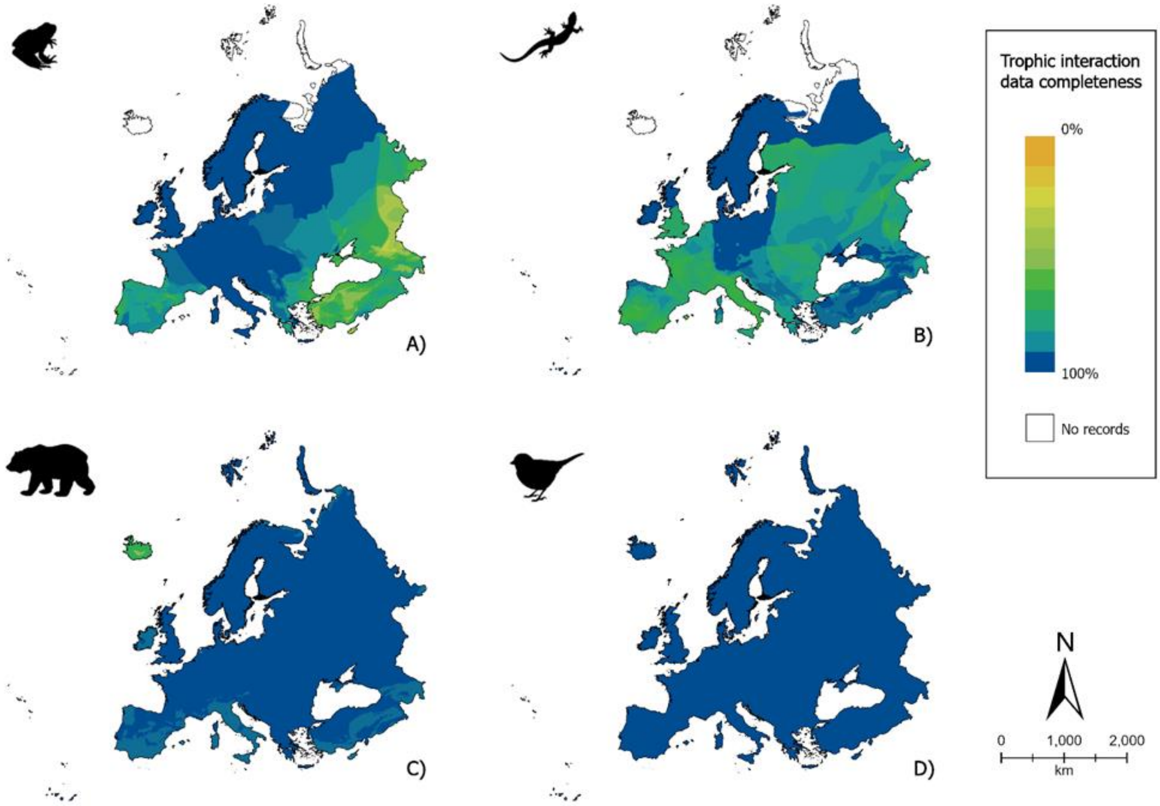


Figure 11: Map of trophic interaction data completeness (Eltonian shortfall) showing the completeness of potential trophic metaweb within 10 km^2 cells. A) amphibians, B) reptiles, C) mammals, D) breeding birds.

The Eltonian of amphibians showed major gaps in the Iberian Peninsula and eastern Europe, with lower data availability in Turkey and Azerbaijan. Besides the above factors, this gap could be also given by recent taxonomic splits (e.g., *Bufo spinosus*, *Triturus karelinii* group; Arntzen et al., 2013; Wielstra et al., 2013) often resulting in a complete lack of ecological knowledge for these species. For reptiles, the areas with a higher gap of knowledge on trophic interactions were located in western and eastern Europe, from the Iberian Peninsula to Italy, and from Hungary to the eastern border of the study area.

Funded by the European Union
NaturaConnect receives funding under the European Union's Horizon Europe research and innovation programme under grant agreement number 101060429.
NATURA CONNECT

2.4. Priorities for data collection and future directions

Biodiversity is severely threatened by human activities, therefore increasing attention is given to the conservation and protection of biodiversity with the aid of systematic conservation planning approaches (WP7). However, our knowledge of biodiversity is limited and biased, with shortfalls encompassing species systematics, distribution, population dynamics, evolution, abiotic tolerances, ecological functions, and species interactions. In this context, sampling strategies and conservation programs are often focused on specific taxa or geographic regions, overlooking other important aspects of biodiversity (e.g., evolutionary distinctiveness) and/or geographic areas that are potentially unique biodiversity hotspots but still unexplored.

Our findings highlight a strong taxonomic bias among tetrapod species, with amphibians and reptiles strongly under-sampled compared to mammals and birds considering all analysed shortfalls. These results call attention to the necessity of investing more effort in European herpetofauna research.

Although during the last decades, several trait databases have been published, including specifically on herpetofauna (e.g., Grimm et al., 2014; Huang et al., 2023; Trochet et al., 2014), the heterogeneity among trait sampling and ontologies undermine the utility of these data for the scientific community. Therefore, a standard protocol for data collection and publication is highly needed to ensure better harmonisation of data (e.g., species nomenclature, trait ontology; Schneider et al., 2019).

From a spatial point of view, based on the gap analysis for tetrapods, priority areas for data collection are located in southern and eastern Europe, which are strongly under-sampled compared to central and northern Europe (e.g., France, United Kingdom). Besides the lack of data, these areas may be important biodiversity hotspots due to the high levels of endemicity (e.g., Greece, Anatolia), therefore major research investment is highly recommended in these areas. Moreover, standardised large-scale monitoring could enhance our understanding of biodiversity and therefore inform systematic conservation planning in order to develop valuable conservation practices at the level of the European continent.

Funded by the European Union

NaturaConnect receives funding under the European Union's Horizon Europe research and innovation programme under grant agreement number 101060429.

ARPHA Preprints Author-formatted document posted on 27/11/2024. DOI: https://doi.org/10.3897/arphapreprints.e142876

3. Conclusion

Despite Europe being a data rich continent, as demonstrated by the extent of the metadatabase compiled in this project, there are still geographical and taxonomic gaps in the data, as well as lack of a standardised protocol to report datasets and accessibility. Enhancing biodiversity conservation and ecosystem service management across Europe would, therefore, benefit from addressing some of the issues highlighted in this report:

Standardisation of metadata reporting: There is a need to promote the publication of open protocols that describe in a standardised way the inputs and outputs of models used for decision-making and research in biodiversity conservation and that would limit the risk for redundancy, overestimations and circularity when integrating several datasets from various sources and disciplines.

Filling geographical and taxonomic gaps: Priority areas for data collection are located in southern and eastern Europe, which are strongly under-sampled compared to central and northern Europe (e.g., France, United Kingdom).

The implementation of a standardised large-scale monitoring can improve our comprehension of biodiversity, through guiding systematic conservation planning to establish effective practices across Europe.

Funded by the European Union

NatureConnect receives funding under the European Union's Horizon Europe research and innovation programme under grant agreement number 101060429.

NATURA CONNECT

4. References

Arntzen, J. W., Recuero, E., Canestrelli, D., & Martínez-Solano, I. (2013). How complex is the Bufo bufo species group? Molecular Phylogenetics and Evolution, 69(3), 1203–1208. https://doi.org/10.1016/j.ympev.2013.07.012

Balvanera, P., Brauman, K. A., Cord, A. F., Drakou, E. G., Geijzendorffer, I. R., Karp, D. S., Martín-López, B., Mwampamba, T. H., & Schröter, M. (2022). Essential ecosystem service variables for monitoring progress towards sustainability. Current Opinion in Environmental Sustainability, 54, 101152. https://doi.org/10.1016/j.cosust.2022.101152

Belletti, B., Garcia de Leaniz, C., Jones, J., Bizzi, S., Börger, L., Segura, G., Castelletti, A., van de Bund, W., Aarestrup, K., Barry, J., Belka, K., Berkhuysen, A., Birnie-Gauvin, K., Bussettini, M., Carolli, M., Consuegra, S., Dopico, E., Feierfeil, T., Fernández, S., … Zalewski, M. (2020). More than one million barriers fragment Europe's rivers. Nature, 588(7838), 436–441. https://doi.org/10.1038/s41586-020-3005-2

Bennett, J. M., Calosi, P., Clusella-Trullas, S., Martínez, B., Sunday, J., Algar, A. C., Araújo, M. B., Hawkins, B. A., Keith, S., Kühn, I., Rahbek, C., Rodríguez, L., Singer, A., Villalobos, F., Ángel Olalla-Tárraga, M., & Morales-Castilla, I. (2018). GlobTherm, a global database on thermal tolerances for aquatic and terrestrial organisms. Scientific Data, 5(1), 180022

Bowler, D. E., Callaghan, C. T., Bhandari, N., Henle, K., Benjamin Barth, M., Koppitz, C., Klenke, R., Winter, M., Jansen, F., Bruelheide, H., & Bonn, A. (2022). Temporal trends in the spatial bias of species occurrence records. Ecography, 2022(8), 1–13. https://doi.org/10.1111/ecog.06219

Breiman, L. (2001). Random Forests. Machine Learning, 45(1), 5–32. https://doi.org/10.1023/A:1010933404324

Broekman, M.J.E., Hoeks, S., Freriks, R., Langendoen, M.M., Runge, K.M., Savenco, E., ter Harmsel, R., Huijbregts, M.A.J. & Tucker, M.A. (2022). HomeRange: A global database of mammalian home ranges. Global Ecology and Biogeography. https://doi.org/10.1111/geb.13625.

Brown, J. H., & Lomolino, M. V. (1998). Biogeography. Sinuer Associates Publishers.

Brun, P., Zimmermann, N. E., Hari, C., Pellissier, L., & Karger, D. N. (2022). Global climate-related predictors at kilometer resolution for the past and future. Earth System Science Data, 14(12), 5573–5603. https://doi.org/10.5194/essd-14-5573-2022

Cardoso, P., Erwin, T. L., Borges, P. A. V., & New, T. R. (2011). The seven impediments in invertebrate conservation and how to overcome them. Biological Conservation, 144(11), 2647–2655. https://doi.org/10.1016/j.biocon.2011.07.024

Chytrý, M., Hennekens, S. M., Jiménez-Alfaro, B., Knollová, I., Dengler, J., Jansen, F., Landucci, F., Schaminée, J. H. J., Aćić, S., Agrillo, E., Ambarlı, D., Angelini, P., Apostolova, I., Attorre, F., Berg, C., Bergmeier, E., Biurrun, I., Botta-Dukát, Z., Brisse, H., … Yamalov, S. (2016). European Vegetation Archive (EVA): An integrated database of European vegetation plots. Applied Vegetation Science, 19(1), 173–180.

Chytrý, M., Tichý, L., Hennekens, S. M., Knollová, I., Janssen, J. A. M., Rodwell, J. S., Peterka, T., Marcenò, C., Landucci, F., Danihelka, J., Hájek, M., Dengler, J., Novák, P., Zukal, D., Jiménez-Alfaro, B., Mucina, L., Abdulhak, S., Aćić, S., Agrillo, E., … Schaminée, J. H. J. (2020). EUNIS Habitat Classification: Expert system, characteristic species combinations and distribution maps of European habitats. Applied Vegetation Science, 23(4), 648–675.

Colwell, R. K., Chao, A., Gotelli, N. J., Lin, S. Y., Mao, C. X., Chazdon, R. L., & Longino, J. T. (2012). Models and estimators linking individual-based and sample-based rarefaction, extrapolation and comparison of assemblages. Journal of Plant Ecology, 5(1), 3–21. https://doi.org/10.1093/jpe/rtr044

Cornell Lab of Ornithology. (2019). All About Birds. Cornell Lab of Ornithology, Ithaca, New York. Accessed on 2024. https://www.allaboutbirds.org

Díaz, S., Pascual, U., Stenseke, M., Martín-López, B., Watson, R. T., Molnár, Z., Hill, R., Chan, K. M. A., Baste, I. A., Brauman, K. A., Polasky, S., Church, A., Lonsdale, M., Larigauderie, A., Leadley, P. W., Oudenhoven, A. P. E. van, Plaat, F. van der, Schröter, M., Lavorel, S., … Shirayama, Y. (2018). Assessing nature's contributions to people. Science, 359(6373), 270–272. https://doi.org/10.1126/science.aap8826

Diniz-Filho, J. A. F., Loyola, R. D., Raia, P., Mooers, A. O., & Bini, L. M. (2013). Darwinian shortfalls in biodiversity conservation. Trends in Ecology and Evolution, 28(12), 689–695. https://doi.org/10.1016/j.tree.2013.09.003

European Environment Agency (2019). European Red Lists of species. https://sdi.eea.europa.eu/catalogue/srv/api/records/202f3c2e-54a9-4ff4-a1da-ed7ca524f634

Funded by the European Union

NaturaConnect receives funding under the European Union's Horizon Europe research and innovation programme under grant agreement number 101060429.

European Environment Agency (2017). High Nature Value (HNV) farmland 2012 (100 m) accounting version, Nov. 2017.

European Environment Agency (2022). European Red List of Habitats - enhanced by EEA. https://sdi.eea.europa.eu/catalogue/srv/api/records/de2276d8-e295-4cd7-89c9-88812065db87

ESA. (2017). Land Cover CCI Product User Guide Version 2. Tech. Rep. maps.elie.ucl.ac.be/CCI/viewer/download/ESACCI-LC-Ph2-PUGv2_2.0.pdf

Etard, A., Morrill, S., & Newbold, T. (2020). Global gaps in trait data for terrestrial vertebrates. Global Ecology and Biogeography, 29(12), 2143–2158. https://doi.org/10.1111/geb.13184

European Environment Agency. (2016). Biogeographical regions, Europe 2016, version 1. https://sdi.eea.europa.eu/catalogue/srv/api/records/c6d27566-e699-4d58-a132-bbe3fe01491b

European Environment Agency. (2023). Building a coherent Trans-European Nature Network Key messages. 1–17.

Fernández, N., Ferrier, S., Navarro, L. M., & Pereira, H. M. (2020). Essential Biodiversity Variables: Integrating In-Situ Observations and Remote Sensing Through Modeling. In J. Cavender-Bares, J. A. Gamon, & P. A. Townsend (Eds.), Remote Sensing of Plant Biodiversity (pp. 485–501). Springer International Publishing. https://doi.org/10.1007/978-3-030-33157-3_18

Gao, J. (2020). Global 1-km Downscaled Population Base Year and Projection Grids Based on the Shared Socioeconomic Pathways, Revision 01. Palisades, New York: NASA Socioeconomic Data and Applications Center (SEDAC). https://doi.org/10.7927/q7z9-9r69. Accessed 2024.

García-Roselló, E., González-Dacosta, J., & Lobo, J. M. (2023). The biased distribution of existing information on biodiversity hinders its use in conservation, and we need an integrative approach to act urgently. Biological Conservation, 283(April). https://doi.org/10.1016/j.biocon.2023.110118

Garcia-Rosello, E., Gonzalez-Dacosta, J., Guisande, C., & Lobo, J. M. (2023). <scp>GBIF</scp> falls short of providing a representative picture of the global distribution of insects. Systematic Entomology, 48(4), 489–497. https://doi.org/10.1111/syen.12589

Gearty W, Chamberlain S (2022). _rredlist: 'IUCN' Red List Client_. R package version 0.7.1,

<https://CRAN.R-project.org/package=rredlist>

Geijzendorffer, I. R., Regan, E. C., Pereira, H. M., Brotons, L., Brummitt, N., Gavish, Y., Haase, P., Martin, C. S., Mihoub, J.-B., Secades, C., Schmeller, D. S., Stoll, S., Wetzel, F. T., & Walters, M. (2016). Bridging the gap between biodiversity data and policy reporting needs: An Essential Biodiversity Variables perspective. Journal of Applied Ecology, 53(5), 1341–1350. https://doi.org/10.1111/1365-2664.12417

Geldmann, J., Heilmann-Clausen, J., Holm, T. E., Levinsky, I., Markussen, B., Olsen, K., Rahbek, C., & Tøttrup, A. P. (2016). What determines spatial bias in citizen science? Exploring four recording schemes with different proficiency requirements. Diversity and Distributions, 22(11), 1139–1149. https://doi.org/10.1111/ddi.12477

Girardello, M., Martellos, S., Pardo, A., & Bertolino, S. (2018). Gaps in biodiversity occurrence information may hamper the achievement of international biodiversity targets: Insights from a cross-Taxon analysis. Environmental Conservation, 45(4), 370–377. https://doi.org/10.1017/S0376892918000115

Gonzalez, A., Vihervaara, P., Balvanera, P., Bates, A. E., Bayraktarov, E., Bellingham, P. J., Bruder, A., Campbell, J., Catchen, M. D., Cavender-Bares, J., Chase, J., Coops, N., Costello, M. J., Dornelas, M., Dubois, G., Duffy, E. J., Eggermont, H., Fernandez, N., Ferrier, S., … Wright, E. (2023). A global biodiversity observing system to unite monitoring and guide action. Nature Ecology & Evolution, 1–5. https://doi.org/10.1038/s41559-023-02171-0

Grimm, A., Ramírez, A. M. P., Moulherat, S., Reynaud, J., & Henle, K. (2014). Life-history trait database of European reptile species. Nature Conservation, 9, 45–67. https://doi.org/10.3897/natureconservation.9.8908

Haesen, S., Lembrechts, J. J., De Frenne, P., Lenoir, J., Aalto, J., Ashcroft, M. B., Kopecký, M., Luoto, M., Maclean, I., Nijs, I., Niittynen, P., van den Hoogen, J., Arriga, N., Brůna, J., Buchmann, N., Čiliak, M., Collalti, A., De Lombaerde, E., Descombes, P., … Van Meerbeek, K. (2021). ForestTemp – Sub-canopy microclimate temperatures of European forests. Global Change Biology, 27(23), 6307–6319.

Hoban, S., Archer, F. I., Bertola, L. D., Bragg, J. G., Breed, M. F., Bruford, M. W., Coleman, M. A., Ekblom, R., Funk, W. C., Grueber, C. E., Hand, B. K., Jaffé, R., Jensen, E., Johnson, J. S., Kershaw, F., Liggins, L., MacDonald, A. J., Mergeay, J., Miller, J. M., … Hunter, M. E. (2022). Global genetic diversity status and trends: Towards a suite of Essential Biodiversity Variables (EBVs) for genetic composition. Biological Reviews, 97(4), 1511–1538. https://doi.org/10.1111/brv.12852

Funded by the European Union

NaturaConnect receives funding under the European Union's Horizon Europe research and innovation programme under grant agreement number 101060429.

Hortal, J., De Bello, F., Diniz-Filho, J. A. F., Lewinsohn, T. M., Lobo, J. M., & Ladle, R. J. (2015). Seven Shortfalls that Beset Large-Scale Knowledge of Biodiversity. Annual Review of Ecology, Evolution, and Systematics, 46, 523–549. https://doi.org/10.1146/annurev-ecolsys-112414-054400

Huang, N., Sun, X., Song, Y., Yuan, Z., & Zhou, W. (2023). Amphibian traits database: A global database on morphological traits of amphibians. Global Ecology and Biogeography, 32(5), 633–641. https://doi.org/10.1111/geb.13656

Hudson, L. N., Newbold, T., Contu, S., Hill, S. L. L., Lysenko, I., Palma, A. D., Phillips, H. R. P., Alhusseini, T. I., Bedford, F. E., Bennett, D. J., Booth, H., Burton, V. J., Chng, C. W. T., Choimes, A., Correia, D. L. P., Day, J., Echeverría-Londoño, S., Emerson, S. R., Gao, D., … Purvis, A. (2017). The database of the PREDICTS (Projecting Responses of Ecological Diversity In Changing Terrestrial Systems) project. Ecology and Evolution, 7(1), 145–188.

Hugo, S., & Van Rensburg, B. J. (2008). The maintenance of a positive spatial correlation between South African bird species richness and human population density. Global Ecology and Biogeography, 17(5), 611–621. https://doi.org/10.1111/j.1466-8238.2008.00391.x

IPBES. (2022). Summary for policymakers of the methodological assessment of the diverse values and valuation of nature of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services. (U. Pascual, P. Balvanera, M. Christie, B. Baptiste, D. González-Jiménez, C.B. Anderson, S. Athayde, R. Chaplin-Kramer, S. Jacobs, E. Kelemen, R. Kumar, E. Lazos, A. Martin, T.H. Mwampamba, B. Nakangu, P. O'Farrell, C.M. Raymond, S.M. Subramanian, M. Termansen, M. Van Noordwijk, A. Vatn). IPBES Secretariat. https://doi.org/10.5281/zenodo.6522392

IUCN. 2022. The IUCN Red List of Threatened Species. Version 2022-1. https://www.iucnredlist.org. Accessed on 2022.10.18.

IUCN. (2024). The IUCN Red List of Threatened Species. Version 2024-1. Accessed on 2024. https://www.iucnredlist.org

Jeppsson, T., and P. Forslund. 2014. "Species' Traits Explain Differences in Red List Status and Long-Term Population Trends in Longhorn Beetles." Animal Conservation 17(4):332–41.

Jetz, W., McGeoch, M. A., Guralnick, R., Ferrier, S., Beck, J., Costello, M. J., Fernandez, M., Geller, G. N., Keil, P., Merow, C., Meyer, C., Muller-Karger, F. E., Pereira, H. M., Regan, E. C., Schmeller, D. S., & Turak, E. (2019). Essential biodiversity variables for mapping and monitoring species populations. Nature Ecology & Evolution, 3(4), Article 4. https://doi.org/10.1038/s41559-019-0826-1

Jones, O. R., Barks, P., Stott, I., James, T. D., Levin, S., Petry, W. K., Capdevila, P., Che-Castaldo, J., Jackson, J., Römer, G., Schuette, C., Thomas, C. C., & Salguero-Gómez, R. (2022). Rcompadre and Rage—Two R packages to facilitate the use of the COMPADRE and COMADRE databases and calculation of life-history traits from matrix population models. Methods in Ecology and Evolution, 13(4), 770–781. https://doi.org/10.1111/2041-210X.13792

Jung, M., Alagador, D., Chapman, M., Hermoso, V., Kujala, H., O'Connor, L., Schinegger, R., Verburg, P. H., & Visconti, P. (2024). An assessment of the state of conservation planning in Europe. Philosophical Transactions of the Royal Society B: Biological Sciences, 379(1902), 20230015. https://doi.org/10.1098/rstb.2023.0015

Jung, M., Davoli, M., Cavalcante, T., O'Connor, L., D'Alessio, A., Cimatti, M., Pacifici, M., Beher, J., Visconti, P., Marco, M. D., & Kujala, H. (2024). Report on network problem formulations, targets and preferences, including guidance and data on targets and optimal TEN-N design criteria. ARPHA Preprints, 5, e138574. https://doi.org/10.3897/arphapreprints.e138574

Kissling, W. D., Ahumada, J. A., Bowser, A., Fernandez, M., Fernández, N., García, E. A., Guralnick, R. P., Isaac, N. J., Kelling, S., & Los, W. (2018). Building essential biodiversity variables (EBV s) of species distribution and abundance at a global scale. Biological Reviews, 93(1), 600–625.

Kissling, W. D., Walls, R., Bowser, A., Jones, M. O., Kattge, J., Agosti, D., Amengual, J., Basset, A., van Bodegom, P. M., Cornelissen, J. H. C., Denny, E. G., Deudero, S., Egloff, W., Elmendorf, S. C., Alonso García, E., Jones, K. D., Jones, O. R., Lavorel, S., Lear, D., … Guralnick, R. P. (2018). Towards global data products of Essential Biodiversity Variables on species traits. Nature Ecology & Evolution, 2(10), Article 10. https://doi.org/10.1038/s41559-018-0667-3

Lobo, J. M., Hortal, J., Yela, J. L., Millán, A., Sánchez-Fernández, D., García-Roselló, E., González-Dacosta, J., Heine, J., González-Vilas, L., & Guisande, C. (2018). KnowBR: An application to map the geographical variation of survey effort and identify well-surveyed areas from biodiversity databases. Ecological Indicators, 91(March), 241–248. https://doi.org/10.1016/j.ecolind.2018.03.077

Lomolino, M. V. (2004). Conservation biogeography. Frontiers of Biogeography: New Directions in the Geography of Nature, 293.

Funded by the European Union
NaturaConnect receives funding under the European Union's Horizon Europe research and innovation programme under grant agreement number 101060429.

Maiorano, L., Montemaggiori, A., Ficetola, G. F., O'Connor, L., & Thuiller, W. (2020). TETRA-EU 1.0: A species-level trophic metaweb of European tetrapods. Global Ecology and Biogeography, 29(9), 1452–1457. https://doi.org/10.1111/geb.13138

Mammola, S., Adamo, M., Antić, D., Calevo, J., Cancellario, T., Cardoso, P., Chamberlain, D., Chialva, M., Durucan, F., Fontaneto, D., Goncalves, D., Martínez, A., Santini, L., Rubio-Lopez, I., Sousa, R., Villegas-Rios, D., Verdes, A., & Correia, R. A. (2023). Drivers of species knowledge across the tree of life. ELife, 12, 1–23. https://doi.org/10.7554/eLife.88251

Marques, A., Martins, I. S., Kastner, T., Plutzar, C., Theurl, M. C., Eisenmenger, N., Huijbregts, M. A. J., Wood, R., Stadler, K., Bruckner, M., Canelas, J., Hilbers, J. P., Tukker, A., Erb, K., & Pereira, H. M. (2019). Increasing impacts of land use on biodiversity and carbon sequestration driven by population and economic growth. Nature Ecology & Evolution, 1. https://doi.org/10.1038/s41559-019-0824-3

Maxim, L., Spangenberg, J. H., & O'Connor, M. (2009). An analysis of risks for biodiversity under the DPSIR framework. Ecological Economics, 69(1), 12–23. https://doi.org/10.1016/j.ecolecon.2009.03.017

Meijer, J. R., Huijbregts, M. A. J., Schotten, K. C. G. J., & Schipper, A. M. (2018). Global patterns of current and future road infrastructure. Environmental Research Letters, 13(6), 64006. https://doi.org/10.1088/1748-9326/aabd42

Meyer, C., Kreft, H., Guralnick, R., & Jetz, W. (2015). Global priorities for an effective information basis of biodiversity distributions. Nature Communications, 6. https://doi.org/10.1038/ncomms9221

Moersberger, H., Valdez, J., Martin, J. G. C., Junker, J., Georgieva, I., Bauer, S., Beja, P., Breeze, T. D., Fernandez, M., Fernández, N., Brotons, L., Jandt, U., Bruelheide, H., Kissling, W. D., Langer, C., Liquete, C., Lumbierres, M., Solheim, A. L., Maes, J., … Bonn, A. (2024). Biodiversity monitoring in Europe: User and policy needs. Conservation Letters, June, 1–12. https://doi.org/10.1111/conl.13038

Morales-Castilla, I., Matias, M. G., Gravel, D., & Araújo, M. B. (2015). Inferring biotic interactions from proxies. Trends in Ecology and Evolution, 30(6), 347–356. https://doi.org/10.1016/j.tree.2015.03.014

Moreira, F., Dias, F. S., Dertien, J., Hasse, A. C., Borda-de-Água, L., Carvalho, S., Porto, M., Cosentino, F., Maiorano, L., Sacchi, A., Santini, L., Borgwardt, F., Gruber, G., Poulsen, N., Schinegger, R., Seliger, C., & Fernández, N. (2024). Guidelines for connectivity conservation and planning in Europe. ARPHA Preprints, 5, e129021. https://doi.org/10.3897/arphapreprints.e129021

Narváez-Gómez, J. P., Guedes, T. B., & Lohmann, L. G. (2021). Recovering the drivers of sampling bias in Bignonieae (Bignoniaceae) and identifying priority areas for new survey efforts. Biodiversity and Conservation, 30(8–9), 2319–2339. https://doi.org/10.1007/s10531-021-02195-7

Navarro, L. M., Fernández, N., Guerra, C., Guralnick, R., Kissling, W. D., Londoño, M. C., Muller-Karger, F., Turak, E., Balvanera, P., Costello, M. J., Delavaud, A., El Serafy, G., Ferrier, S., Geijzendorffer, I., Geller, G. N., Jetz, W., Kim, E.-S., Kim, H., Martin, C. S., … Pereira, H. M. (2017). Monitoring biodiversity change through effective global coordination. Current Opinion in Environmental Sustainability, 29, 158–169. https://doi.org/10.1016/j.cosust.2018.02.005

Nelson, A., Weiss, D. J., van Etten, J., Cattaneo, A., McMenomy, T. S., & Koo, J. (2019). A suite of global accessibility indicators. Scientific Data, 6(1), 1–9. https://doi.org/10.1038/s41597-019-0265-5

Ness, B., Anderberg, S., & Olsson, L. (2010). Structuring problems in sustainability science: The multi-level DPSIR framework. Geoforum, 41(3), 479–488. https://doi.org/10.1016/j.geoforum.2009.12.005

Oliver, R. Y., Meyer, C., Ranipeta, A., Winner, K., & Jetz, W. (2021). Global and national trends, gaps, and opportunities in documenting and monitoring species distributions. PLoS Biology, 19(8), 1–14. https://doi.org/10.1371/journal.pbio.3001336

OpenStreetMap. (2017). Planet dump retrieved from https://planet.osm.org.

Parr, C. L., Dunn, R. R., Sanders, N. J., Weiser, M. D., Photakis, M., Bishop, T. R., Fitzpatrick, M. C., Arnan, X., Baccaro, F., Brandão, C. R. F., Chick, L., Donoso, D. A., Fayle, T. M., Gómez, C., Grossman, B., Munyai, T. C., Pacheco, R., Retana, J., Robinson, A., … Gibb, H. (2017). GlobalAnts: A new database on the geography of ant traits (Hymenoptera: Formicidae). Insect Conservation and Diversity, 10(1), 5–20.

Pereira, H. M., & Cooper, D. H. (2006). Towards the global monitoring of biodiversity change. Trends in Ecology & Evolution, 21(3), 123–129.

Pereira, H. M., Ferrier, S., Walters, M., Geller, G. N., Jongman, R. H. G., Scholes, R. J., Bruford, M. W., Brummitt, N., Butchart, S. H. M., Cardoso, A. C., Coops, N. C., Dulloo, E., Faith, D. P., Freyhof, J., Gregory, R. D., Heip, C., Höft, R., Hurtt, G., Jetz, W., … Wegmann, M. (2013). Essential Biodiversity Variables. Science, 339(6117), 277–278. https://doi.org/10.1126/science.1229931

Pereira, H. M., Martins, I. S., Rosa, I. M. D., Kim, H., Leadley, P., Popp, A., van Vuuren, D. P., Hurtt, G., Quoss, L., Arneth, A., Baisero, D., Bakkenes, M., Chaplin-Kramer, R., Chini, L., Di Marco, M., Ferrier, S., Fujimori, S.,

Funded by the European Union
NaturaConnect receives funding under the European Union's Horizon Europe research and innovation programme under grant agreement number 101060429.

Guerra, C. A., Harfoot, M., … Alkemade, R. (2024). Global trends and scenarios for terrestrial biodiversity and ecosystem services from 1900 to 2050. Science, 384(6694), 458–465. https://doi.org/10.1126/science.adn3441

Perino, A., Pereira, H. M., Felipe-Lucia, M., Kim, H., Kühl, H. S., Marselle, M. R., Meya, J. N., Meyer, C., Navarro, L. M., van Klink, R., Albert, G., Barratt, C. D., Bruelheide, H., Cao, Y., Chamoin, A., Darbi, M., Dornelas, M., Eisenhauer, N., Essl, F., … Bonn, A. (2022). Biodiversity post-2020: Closing the gap between global targets and national-level implementation. Conservation Letters, 15(2), e12848. https://doi.org/10.1111/conl.12848

Polce, C., Maes, J., Rotllan-Puig, X., Michez, D., Castro, L., Cederberg, B., Dvorak, L., Fitzpatrick, Ú., Francis, F., Neumayer, J., Manino, A., Paukkunen, J., Pawlikowski, T., Roberts, S., Straka, J., & Rasmont, P. (2018). Distribution of bumblebees across Europe. One Ecosystem, 3, e28143.

Pollock, L. J., O'Connor, L. M. J., Mokany, K., Rosauer, D. F., Talluto, M. V., & Thuiller, W. (2020). Protecting Biodiversity (in All Its Complexity): New Models and Methods. Trends in Ecology and Evolution, 35(12), 1119–1128. https://doi.org/10.1016/j.tree.2020.08.015

Prenda, J., Domínguez-Olmedo, J. L., López-Lozano, E., Fernández de Villarán, R., & Negro, J. J. (2024). Assessing citizen science data quality for bird monitoring in the Iberian Peninsula. Scientific Reports, 14(1), 1–17. https://doi.org/10.1038/s41598-024-70827-3

Rasmont P. & Iserbyt I. 2010-2014. Atlas of the European Bees: genus Bombus. 3d Edition. STEP Project, Atlas Hymenoptera, Mons, Gembloux. http://www.atlashymenoptera.net/page.aspx?ID=169

Shirey, V., Larsen, E., Doherty, A., Kim, C. A., Al-Sulaiman, F. T., Hinolan, J. D., Itliong, M. G. A., Naive, M. A. K., Ku, M., Belitz, M., Jeschke, G., Barve, V., Lamas, G., Kawahara, A. Y., Guralnick, R., Pierce, N. E., Lohman, D. J., & Ries, L. (2022). LepTraits 1.0 A globally comprehensive dataset of butterfly traits. Scientific Data, 9(1), Article 1.

Schneider, F. D., Fichtmueller, D., Gossner, M. M., Güntsch, A., Jochum, M., König-Ries, B., Le Provost, G., Manning, P., Ostrowski, A., Penone, C., & Simons, N. K. (2019). Towards an ecological trait-data standard. Methods in Ecology and Evolution, 10(12), 2006–2019. https://doi.org/10.1111/2041-210X.13288

Speybroeck, J., Beukema, W., Dufresnes, C., Fritz, U., Jablonski, D., Lymberakis, P., ... & Crochet, P. A. (2020). Species list of the European herpetofauna–2020 update by the Taxonomic Committee of the Societas Europaea Herpetologica. Amphibia-Reptilia, 41(2), 139-189. http://doi.org/10.1163/15685381-bja10010

The World Bank Group. (2024). The World Bank, World Development Indicators. http://data.worldbank.org

Trimble, M. J., & van Aarde, R. J. (2010). Inequidad de especies en el estudio científico. Conservation Biology, 24(3), 886–890. https://doi.org/10.1111/j.1523-1739.2010.01453.x

Trochet, A., Moulherat, S., Calvez, O., Stevens, V. M., Clobert, J., & Schmeller, D. S. (2014). A database of life-history traits of European amphibians. Biodiversity Data Journal, 2(1). https://doi.org/10.3897/BDJ.2.e4123

Troudet, J., Grandcolas, P., Blin, A., Vignes-Lebbe, R., & Legendre, F. (2017). Taxonomic bias in biodiversity data and societal preferences. Scientific Reports, 7(1), 1–14. https://doi.org/10.1038/s41598-017-09084-6

Uetz, P., Koo, M. S., Aguilar, R., Brings, E., Catenazzi, A., Chang, A. T., Chaitanya, R., Freed, P., Hammermann, M., Jiří Hošek, G., Lambert, M., Sergi, Z., Spencer, C. L., Summers, K., Tarvin, R., Vredenburg, V. T., & Wake, D. B. (2021). A Quarter Century of Reptile and Amphibian Databases. Herpetological Review, 52(2), 246–255. http://www.reptile-database.org

Ugland, K. I., Gray, J. S., & Ellingsen, K. E. (2003). The species-accumulation curve and estimation of species richness. Journal of Animal Ecology, 72(5), 888–897. https://doi.org/10.1046/j.1365-2656.2003.00748.x

UNDP. (2024). REPORT 2023 / 2024 Reimagining cooperation in a polarized world.

UNEP-WCMC, & IUCN. (2024). Protected Planet: The World Database on Protected Areas (WDPA) [Online], Cambridge, UK: UNEP-WCMC and IUCN. Accessed on 2024. https://www.protectedplanet.net

Vergara-Asenjo, G., Alfaro, F. M., & Pizarro-Araya, J. (2023). Linnean and Wallacean shortfalls in the knowledge of arthropod species in Chile: Challenges and implications for regional conservation. Biological Conservation, 281(November 2022), 110027. https://doi.org/10.1016/j.biocon.2023.110027

Wielstra, B., Litvinchuk, S. N., Naumov, B., Tzankov, N., & Arntzen, J. W. (2013). A revised taxonomy of crested newts in the Triturus karelinii group (Amphibia: Caudata: Salamandridae), with the description of a new species. Zootaxa, 3682(3), 441–453. https://doi.org/10.11646/zootaxa.3682.3.5

Wilman, H., Belmaker, J., Simpson, J., de la Rosa, C., Rivadeneira, M. M., & Jetz, W. (2014). EltonTraits 1.0: Species-level foraging attributes of the world's birds and mammals. Ecology, 95(7), 2027–2027. https://doi.org/10.1890/13-1917.1

Funded by the European Union

NaturaConnect receives funding under the European Union's Horizon Europe research and innovation programme under grant agreement number 101060429.

Zizka, A., Rydén, O., Edler, D., Klein, J., Aronsson, H., Perrigo, A., Silvestro, D., Jagers, S. C., Lindberg, S. I., & Antonelli, A. (2020). Exploring the Impact of Political Regimes on Biodiversity. SSRN Electronic Journal, March. https://doi.org/10.2139/ssrn.3561430

Funded by the European Union
NaturaConnect receives funding under the European Union's Horizon Europe research and innovation programme under grant agreement number 101060429.

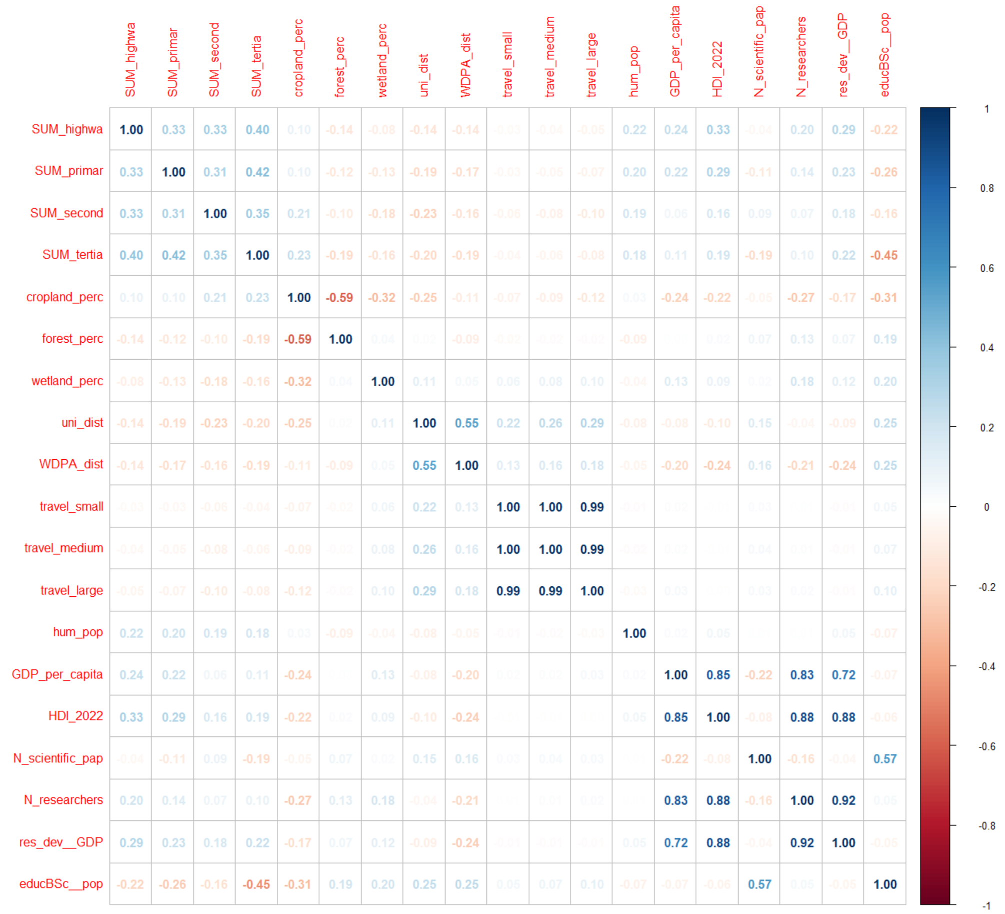
5. Annexes

5.1. Annex I: Metadatabase of data on biodiversity, drivers and protected areas

Metadatabase available at: https://doi.org/10.5281/zenodo.14008205

5.2. Annex II: Data sources and variables used in the gap analysis

Figure 1: Pearson correlation matrix showing correlation coefficients between variables within the study area.



Funded by the European Union

NaturaConnect receives funding under the European Union's Horizon Europe research and innovation programme under grant agreement number 101060429.

Table A1: List of citations of GBIF downloaded data.

https://doi.org/10.15468/dl.2t9veg

https://doi.org/10.15468/dl.pevunv

https://doi.org/10.15468/dl.a4fwnr

https://doi.org/10.15468/dl.uw5n7r

https://doi.org/10.15468/dl.3yjw3u

https://doi.org/10.15468/dl.9dreuk

https://doi.org/10.15468/dl.9dreuk

https://doi.org/10.15468/dl.7qz3ft

https://doi.org/10.15468/dl.gq7jne

https://doi.org/10.15468/dl.f7jty3

https://doi.org/10.15468/dl.3wvxxn

https://doi.org/10.15468/dl.araeg9

https://doi.org/10.15468/dl.ktayy9

https://doi.org/10.15468/dl.ffwdye

https://doi.org/10.15468/dl.mn7x9z

https://doi.org/10.15468/dl.ny8ffu

https://doi.org/10.15468/dl.hyd3wg

https://doi.org/10.15468/dl.38s3nw

https://doi.org/10.15468/dl.r9mxhg

https://doi.org/10.15468/dl.ngz44v

https://doi.org/10.15468/dl.sr4vxm

https://doi.org/10.15468/dl.f7bkb5

https://doi.org/10.15468/dl.b85ty4

https://doi.org/10.15468/dl.rx9b69

https://doi.org/10.15468/dl.7fkfgt

https://doi.org/10.15468/dl.3wuu3y

https://doi.org/10.15468/dl.rfpdzx

https://doi.org/10.15468/dl.rzas3j

https://doi.org/10.15468/dl.846pke

https://doi.org/10.15468/dl.qrvupg

https://doi.org/10.15468/dl.7k7gj6

https://doi.org/10.15468/dl.ks3q98

Funded by the European Union

NaturaConnect receives funding under the European Union's Horizon Europe research and innovation programme under grant agreement number 101060429.

*Table A2: Variables selected for the drivers of knowledge gap analysis. Variables with * were excluded from the analysis (Pearson correlation ≥0.7).*

Variable name	Description	Unit	Resolution	Data source
%forest	Percentage of forest within 10x10km cell	%	300 m	Global CCI-LC 2022
%wetland	Percentage of wetland within 10x10km cell	%	300 m	Global CCI-LC 2022
%cropland	Percentage of cropland within 10x10km cell	%	300 m	Global CCI-LC 2022
travel_time_small	Travel time to the nearest cities with > 10.000 people	Minutes per meters	30 arc-sec	Nelson et al. 2019
***travel_time_medium**	Travel time to the nearest cities with > 100.000 people	Minutes per meters	30 arc-sec	Nelson et al. 2019
***travel_time_large**	Travel time to the nearest cities with > 1.000.000 people	Minutes per meters	30 arc-sec	Nelson et al. 2019
highways_road_density	Sum of linear km of highways within each 10x10km cell	km	-	GRIP
primary_road_density	Sum of linear km of primary road within each 10x10km cell	km	-	GRIP
secondary_road_density	Sum of linear km of secondary road within each 10x10km cell	km	-	GRIP
tertiary_road_density	Sum of linear km of tertiary road within each 10x10km cell	km	-	GRIP
universities_distance	Euclidean distance from universities, museums, and colleges	km	-	OSM
WDPA_distance	Euclidean distance from protected areas	km	-	WDPA

Funded by the European Union
NaturaConnect receives funding under the European Union's Horizon Europe research and innovation programme under grant agreement number 101060429.

human_pop_density	Human population density	N° of people per cell	30 arc-sec	SEDAC
N_scientific_papers	Number of scientific and engineering articles published	N° of papers	-	World Bank Open Data
***research_expenditure**	Gross domestic expenditures on research and development	% of GDP	-	World Bank Open Data
GDP_per_capita	Gross domestic product divided by midyear population	U.S. dollars	-	World Bank Open Data
***N_researchers**	Number of researchers engaged in Research & Development	N° of researcher/million of people	-	World Bank Open Data
%pop_educBSc	The percentage of population ages 25 and over that attained or completed Bachelor's or equivalent	%	-	World Bank Open Data
***HDI**	Human development index	-	-	UNDP

Funded by the European Union

NaturaConnect receives funding under the European Union's Horizon Europe research and innovation programme under grant agreement number 101060429.

More information about the project:

NaturaConnect has 22 partner institutions: International Institute for Applied System Analysis (project lead; Austria); German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig (project co-lead; Germany); Associacao Biopolis (Portugal); BirdLife Europe (Netherlands); Birdlife International (United Kingdom); Centre National De La Recherche Scientifique (France); Doñana Research Station - Agencia Estatal Consejo Superior De Ivestigaciones Cientificas (Spain); Europarc Federation (Germany); Finnish Environment Institute (Finland); Humboldt-University of Berlin (Germany); Institute for European Environmental Policy (Belgium); Netherlands Environmental Assessment Agency (Netherlands); Rewilding Europe (Netherlands); University of Evora (Portugal); University of Helsinki (Finland); University of Natural Resources and Life Sciences, Vienna (Austria); University of Rome La Sapienza (Italy); University of Warsaw (Poland); Vrie University of Amsterdam (Netherlands); WWF Central and Eastern Europe (Austria); WWF Romania and WWF Hungary.

NaturaConnect aims to design and develop a blueprint for a truly coherent **Trans-European Nature Network** (TEN-N) of conserved areas that protect at least 30% of land in the European Union, with at least one third of it under strict protection. Our project unites universities and research institutes, government bodies and non-governmental organizations, working together with key stakeholders to create targeted knowledge and tools, and build the capacity needed to support European Union Member States in realizing an ecologically representative, resilient and well-connected network of conserved areas across Europe.

https://naturaconnect.eu/

Funded by the European Union

NaturaConnect receives funding under the European Union's Horizon Europe research and innovation programme under grant agreement number 101060429.

Funded by the European Union
NaturaConnect receives funding under the European Union's Horizon Europe research and innovation programme under grant agreement number 101060429.