30/03/2026
Version 1.0

BMD
Biodiversity Meets Data

MS26 Concept to integrate use cases based on the Habitats Directive and the Birds Directive

Author(s): Namsrai Jargal (SGN), Marie Gutgesell (SGN), Peter Haase (SGN)

Co-funded by the European Union

Project funded by

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Swiss Confederation

Federal Department of Economic Affairs, Education and Research EAER
State Secretariat for Education, Research and Innovation SERI

BMD (Biodiversity Meets Data) receives funding from the European Union's Horizon Europe Research and Innovation Programme and the Swiss State Secretariat for Education, Research and Innovation (SERI) (ID No 101181294). Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union, the European Research Executive Agency (REA) or SERI. The EU, REA and SERI cannot be held responsible for them.

Contributor(s): Julian Oeser (UFZ), Taimur Khan (UFZ), Ingolf Kühn (UFZ), Alessandro Oggioni (CNR), Caterina Bergami (CNR), Quan Pan (LifeWatch), Zhiming Zhao (LifeWatch)

Prepared under contract from the European Commission
Grant agreement No. 101181294
EU Horizon Europe Research and Innovation Action

Project acronym:	**BMD**
Project full title:	**Biodiversity Meets Data**
Project duration:	01.03.2025 - 28.02.2029 (48 months)
Project coordinator:	Stichting Naturalis Biodiversity Center (Naturalis)
Call:	HORIZON-CL6-2024-BIODIV-01
Milestone title:	CONCEPT TO INTEGRATE USE CASES BASED ON THE HABITATS DIRECTIVE AND THE BIRDS DIRECTIVE
Milestone n°:	MS26
Means of verification:	Concept
Work package:	WP5
Nature of the milestone:	Concept
Contribution to deliverable n°:	D5.2
Licence of use:	Not applicable
Lead beneficiary:	Senckenberg Society of Nature Research (SGN)
Recommended citation:	Not applicable
Due date of milestone:	March 31, 2026 **(M13)**
Actual submission date:	March 30, 2026
Quality review:	Yes

Milestone status:

Version	Status	Date	Author(s)	Actions
0.1	Draft	27.02.2026	Namsrai Jargal, Marie Gutgesell, Peter Haase	Sent for review
0.2	Draft	17.03.2026		Reviewed
0.3	Draft			Finalised, with incorporation of feedback from reviewers
1.0	Final version	30.03.2026		Submitted

Table of contents

1. Summary

This milestone MS26 report presents the conceptual and technical design for integrating use cases based on the Habitats Directive (HD) and the Birds Directive (BD) as portable, customized Biodiversity Analysis Tools (BATs). The BATs represent specific, purpose-built analytical tools, tailored to defined use cases for the HD and BD and co-designed by stakeholders. Grounded in the legal frameworks of Directive 92/43/EEC (HD) and Directive 2009/147/EC (BD), these tools focus on assessing the spatial distribution of both species and habitat types, and temporal species occurrence trends using spatiotemporal data and modelling. The framework uses the Annexes of both Directives to define the habitats and species of interest and integrates species occurrence data from GBIF and realm-specific systems (OBIS and EMODnet). These data are combined with environmental predictors to generate outputs such as distribution maps and analyses of temporal trends. The BATs support biodiversity assessment and monitoring at the Natura 2000 network, Member State, and EU levels by linking legal obligations and stakeholder needs with data, workflows, and underlying computing and storage capacities in a reproducible, policy-relevant manner.

2. List of abbreviations

BAT	Biodiversity Analysis Tool
BD	Birds Directive
EEA	European Environment Agency
EMODnet	European Marine Observation and Data Network
EU	European Union
EUNIS	European Nature Information System
FCS	Favorable Conservation Status
HD	Habitats Directive
GBIF	Global Biodiversity Information Facility
NGOs	Non-Governmental Organizations
OBIS	Ocean Biodiversity Information System
SDM	Species Distribution Model

3. Conceptual framework

3.1. Purpose

The overarching goal of this milestone is to define a conceptual and technical framework for integrating use cases that support the Habitats Directive (HD) and the Birds Directive (BD) into dedicated Biodiversity Analysis Tools (BATs).

These tools will facilitate coherent, reproducible and harmonised assessments of the spatial distribution of listed species and habitat types, temporal trends in species occurrence, and projected changes in species distribution under climate change scenarios across European terrestrial, freshwater, and marine ecosystems. Each BAT is tailored to a specific ecosystem and analytical theme to enable these core analyses.

The BATs will be co-designed with stakeholders, including Natura 2000 site managers, national environmental agencies, and NGOs. This collaborative process ensures the BATs are practical, user-friendly, and directly support biodiversity assessment, monitoring, and reporting mandated under the HD and BD.

3.2. Scope

- Realms: Terrestrial, freshwater, and marine ecosystems
- Spatial scales: Regional (incl. Natura 2000 sites), national, EU-wide assessments
- Biological levels: Species, habitat types

3.3. Key reference links

Links to EU directives:
HD: https://environment.ec.europa.eu/topics/nature-and-biodiversity/habitats-directive_en
BD: https://environment.ec.europa.eu/topics/nature-and-biodiversity/birds-directive_en

Links to relevant EU habitat and species lists to the HD:
Annex I habitat types: https://eunis.eea.europa.eu/references/2324/habitats
Annex II species (animal and plant species of community interest whose conservation requires the designation of special areas of conservation): https://eunis.eea.europa.eu/references/2325/species
Annex IV species (animal and plant species of community interest in need of strict protection): https://eunis.eea.europa.eu/references/2326/species

Link to EU bird list under the BD
Annex I species (bird species that are endangered, vulnerable to habitat changes, rare because of small populations or restricted distribution, and require particular attention due to the specific nature of their habitat)**:** http://data.europa.eu/eli/dir/2009/147/oj

We focus only on Annex I species of the BD, as it includes nearly 200 threatened and rare bird species of highest conservation priority. This ensures policy relevance and aligns with stakeholder needs, while keeping the BATs focused and practical.

3.4. Introduction

The HD and BD are the cornerstone legislative instruments for biodiversity monitoring and conservation in the European Union. Their primary aim is to maintain favourable and restore unfavourable conservation status for Europe's most valuable and threatened habitats and species through the establishment of the Natura 2000 network of protected sites. The state of these habitats and species, and the pressures they face, have profound consequences:

Environmental status

- Habitats of community interest (Annex I, HD) and wild bird species provide essential ecosystem structure and functions.
- Their decline directly reduces ecosystem resilience, genetic diversity, and the provision of essential services such as pollination, water purification, and climate regulation (IPBES, 2019).
- Habitat loss and degradation, largely driven by land-use change, pollution and natural resource use and exploitation, are leading drivers of biodiversity loss in Europe (IPBES, 2019; EEA, 2020; Jaureguiberry et al. 2022).

Socio-economic impacts

- Healthy, biodiverse ecosystems underpin key economic sectors such as agriculture, forestry, fisheries, and tourism (IPBES, 2019).
- The degradation of these natural assets imposes long-term costs related to lost services, reduced productivity, and the need for restoration (European Commission, 2021).
- The Natura 2000 network itself supports local economies and livelihoods through sustainable use (Rayment et al. 2011; Gantioler et al. 2014).

Given this context, effective monitoring, assessment, and management are legal requirements for Natura 2000 sites and Member States. These requirements must be translated into practical tools that address the diverse needs of stakeholders–from site managers requiring timely data for conservation actions, to national agencies responsible for reporting, and policymakers seeking evidence for assessments. Reliable data on species distribution and occurrence trends over time are critical for assessing the conservation status, measuring the effectiveness of conservation actions, and prioritizing these actions. A virtual integration of species occurrence data, environmental predictors, and modeling frameworks can facilitate evidence-based decision-making, support standardized reporting, and can enhance the implementation of both Directives by putting actionable information directly into the hands of those who need it.

3.5. Regulatory background

The HD (Directive 92/43/EEC) and the BD (Directive 2009/147/EC) establish a comprehensive framework for species and habitat protection. They mandate a six-yearly reporting cycle by Member States on the status of all listed habitats and species (Article 17 of the HD, Article 12 of the BD (EEA, 2020)), as well as on the conservation measures undertaken within the Natura 2000 network.

The Commission Implementing Decision (EU) 2011/484/EU and subsequent updates specify the detailed structure, formats, and indicators for this reporting. Key reported parameters include:

- **Range**: The distribution area of the habitat or species.
- **Population**: For species, parameters such as population size, density, and structure.
- **Habitat for species**: The area and quality of habitat available for the species.
- **Future prospects**: An assessment of the likelihood of maintaining FCS in the future.
- **Pressures and threats**: The main anthropogenic and natural factors impacting the habitat or species.
- **Conservation status**: The overall assessment (Favourable, Unfavourable-Inadequate, Unfavourable-Bad, or Unknown).

Effective implementation requires translating these legal obligations into actionable analytical use cases that address core questions:

A. Where are listed habitats currently located?
B. What are current distributions of listed species?
C. How are species occurrences changing over time?, and
D. What are the future changes in species distributions under different climate scenarios?

This MS26 milestone focuses on the data and modeling components that directly support the assessment of distribution and the temporal trends. The BATs will leverage species occurrence data from sources like GBIF and OBIS to generate the analytical outputs needed to inform these reporting parameters and support conservation management. By design, the tools are being co-developed with end-users, including Natura 2000 site managers, national environmental agencies, and NGOs, to ensure the resulting tools are practically aligned with stakeholder monitoring needs and flexible in their usage.

3.6. Potential integrated use cases across realms

- **Habitat and species information**
  Generate standardized lists and reference data for habitat types (Annex I, HD) and species listed under the HD (Annex II, IV) and BD (Annex I). Provide authoritative identifiers (e.g., EUNIS codes for habitats, scientific names for species) to ensure compatibility with regulatory reporting. These outputs form the foundational taxonomic and habitat layer for all subsequent spatial and temporal analyses required under Article 17 (HD) and Article 12 (BD).
- **Distribution and status of habitat types**

Use available data sources on the distribution of EUNIS habitat types for visualization and summary statistics on the presence and distribution of habitat types (Annex I, HD). If possible, use remote sensing satellite time series data to infer information on the status and potential changes occurring in habitat types.

- **Species presence**
  Compile and assess current occurrence records for species from integrated data streams. BAT workflows combine taxonomic lists from the Directives' Annexes with occurrence data from GBIF and OBIS to produce up-to-date presence records. These records are the essential input for generating outputs such as occurrence maps and aggregated summaries at Natura 2000 sites and/or Member States, and for spatial and temporal modelling.
- **Distribution of species**
  Model the current and future potential distributions of species listed under the HD and BD, as well as the spatial extent of HD Annex I habitat types. Species distribution models across realms and species are applied for this purpose and future projections will be under climate change scenarios (e.g., SSP pathways) for future periods (mid- or late-century). These outputs support the assessment of species range and habitat suitability (key parameters for determining conservation status under the directives) and help identify potential range shifts and conservation priorities under future conditions.
- **Temporal trends of species**
  Analyze occurrence-based temporal trends for species within key spatial units, including Natura 2000 sites and/or Member States. Trend assessments can be derived from time-series occurrence data, corrected for sampling bias where possible, or through more advanced modeling (e.g., occupancy models). These analyses provide essential information on population dynamics, which is a core parameter for assessing conservation status under Article 17 (HD) and Article 12 (BD).

3.7. Implementation

The potential BATs will be implemented using the R programming language, which provides extensive libraries for biodiversity informatics, spatial analysis, and species distribution modeling. User-facing components will be delivered through interactive dashboards, offering an intuitive graphical user interface (GUI).

The design of the BAT will strictly follow the FAIR principles (Findable, Accessible, Interoperable, Reusable) to ensure long-term scientific and policy relevance:

- Rich and standardized metadata: All datasets and workflows will include detailed provenance documentation, specifying data source, date, version, and spatial/temporal resolution.
- Data governance and quality assessments: Built-in checks will ensure data consistency and reliability. Primary alignment will be with data standards and formats from official reporting under the HD (Article 17) and BD (Article 12).
- Reproducible workflows: Version-controlled scripts and containerized environments (e.g., via Docker) will enable fully repeatable and auditable analyses, supporting trustworthiness for

regulatory reporting. A workflow engine will centralize workflow runs and act as a compute backend to the BATs.

Additionally, BATs, based on different use cases linked to the HD and BD, will be scalable, from individual Natura 2000 sites to EU-wide analyses, and interoperable with existing biodiversity data infrastructures such as GBIF and OBIS. This ensures that outputs can be directly used for both regulatory compliance and stakeholder decision support.

3.8. Data resources

Authoritative lists

- **Annexes I, II, IV of the HD:** The definitive lists for habitat types and species of Community interest (see section "Key reference links").
- **Annex I of the BD:** The list of bird species that are threatened and require special conservation measures (see section "Key reference links").

Species occurrence data

- **GBIF:** Primary source for occurrence data for species across terrestrial, freshwater, and marine realms.
- **OBIS** and **EMODnet Biology:** Sources for occurrence data for marine species.

Environmental predictor data

- CHELSA Bioclimatic variables
- CORINE Land cover/use data
- Bio-ORACLE
- Earth observation data
- Other realms/taxon-specific environmental factors

4. Design of a Biodiversity Analysis Tools (BATs)

4.1. BATs design workflow

Figure 1 illustrates the conceptual framework used for defining the potential BATs, focused on use cases linked to the HD and the BD (Oeser, 2025). The design process begins by defining monitoring needs based on input from relevant stakeholders and reporting requirements of both directives. Stakeholder input will be gathered through co-design workshops, for example, in the form of user stories, and will inform BATs design, user interfaces, and data presentation to ensure the BATs are practical, user-friendly, and aligned with stakeholder needs. Stakeholder inputs will be incorporated iteratively throughout the design and prototyping phases to refine outputs and confirm usability. As the BATs are developed, initial prototypes will be shared with relevant stakeholders to gather feedback on design, utility, and usability. This iterative process, which is facilitated by WP1 through online workshops, questionnaires, and in-person meetings, ensures continuous stakeholder engagement and refinement throughout the project. To design specific BATs, we need to define four conceptual key components:

- Task: Short description of the specific monitoring task/question the BATs address. These tasks are derived from monitoring needs defined based on stakeholder requirements (e.g., based on user stories) and directives. [WP1] [WP5]
- Data: Description of relevant required and available datasets. [WP2] [WP5]
- Methods: Description of data analysis and modeling methods. [WP5]
- Outputs: Description of outputs (required/provided) and their formatting. [WP1] [WP5]

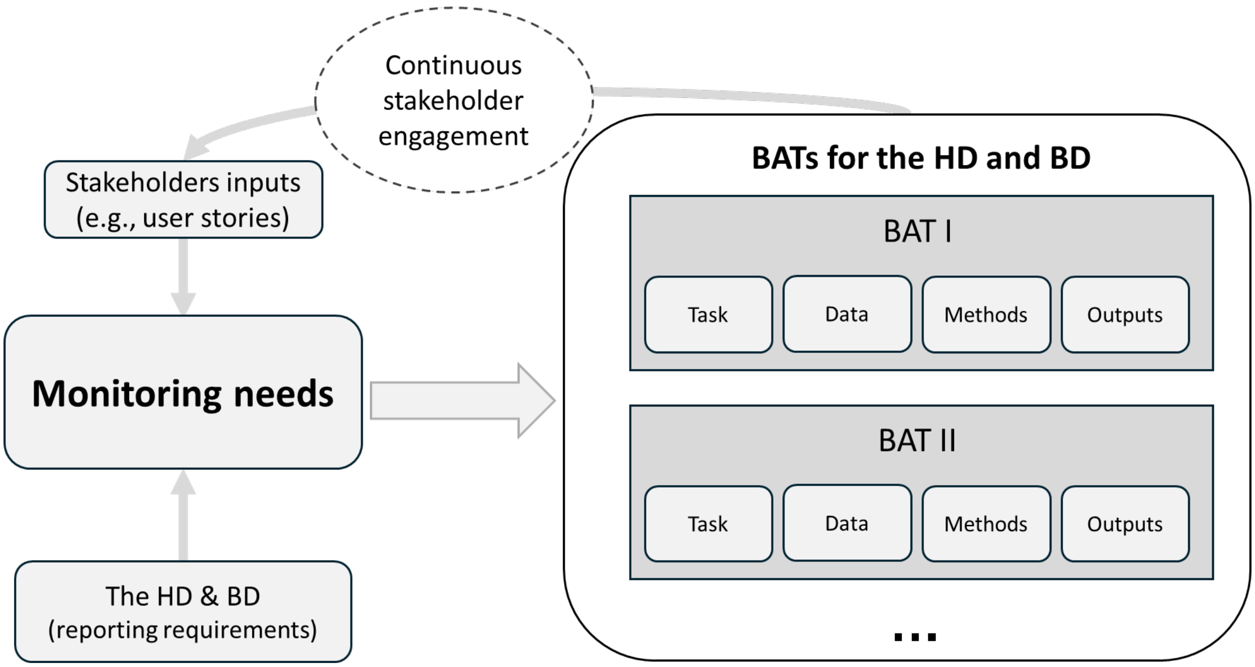

Figure 1: Conceptual representation of the BATs design workflow.

The complete workflow as well as model parametrisation will be documented in a FAIR way, following the protocol suggested by Zurell et al. (2020). While this is specifically for species distribution models, we will adapt the respective parts also for temporal analyses.

4.2. BATs list

The BATs are structured around the key use cases for the HD and BD, organized across the terrestrial, freshwater, and marine realms. The following tables summarize the relevant tasks, input data, analytical methods, outputs, and potential challenges for each realm-specific BATs.

BAT for assessing distribution and status of habitat types

Task	Assess distribution and status of habitat types monitored under Annex I of the HD or based on the EUNIS habitat type classification

Data	● Existing map products on the distribution of habitat types ● Satellite remote sensing time series for inferring status and changes of habitat types
Methods	Use existing map products to provide visualizations and summary statistics on the distribution of EUNIS habitat types. Use satellite time series information to infer, where possible, information on the status and changes in habitat types.
Outputs	● Maps of Annex I / EUNIS habitat type distributions ● Summary statistics on habitat type distributions (e.g., per Member state or Natura2000 site) ● Where possible, information on status and changes in habitat types derived from satellite time series
Challenges	● Lack of available map products which follow the Annex I habitat type classification ● Potential complexities or data loss when crosswalking EUNIS habitat classifications to Annex I types. ● Potential lack of extensive ground-truthing data which may be necessary for inferring status and changes in habitat types via satellite time series ● High computational costs associated with processing fine-resolution satellite time series data

BAT for assessing spatial distributions of terrestrial species (HD)

Task	Assess the distribution of terrestrial species listed in the annexes of the HD and BD: current and future potential distributions
Data	Occurrence data: ● GBIF records for terrestrial species listed in the HD Annexes (II, IV) ● Global occurrence data is important to characterize the full environmental niche of the species ● Regional occurrence data (e.g., within Europe) can then be used to build regional SDMs that use outputs of global models as an input (nested distribution modeling approach) Predictor data: ● CHELSA Bioclim: 19 bioclimatic variables as input to global SDMs ● Chelsa Bioclim + other environmental predictors as inputs to regional SDMs. Potentially useful variables include datasets on human landscape modification, transportation infrastructure (roads & railways), soil conditions, land cover, topographic variables ○ The selection of specific datasets depends on their integration in the BMD data catalogue
Methods	1. Current potential distribution: Apply nested or hierarchical SDMs. Contrast potential habitat suitability predictions from SDMs with

	observed (realized) distributions from occurrence data to assess range filling and identify suitable but unoccupied habitat. 2. Future potential distributions: Project SDMs using future climate scenarios (e.g., SSP pathways) to estimate potential range shifts. For bird species, where applicable, create separate predictions for breeding and wintering ranges.
Outputs	Maps of current species occurrences; maps of current and future potential distributions based on SDMs; analysis of projected species range shifts under future climate scenarios.
Challenges	● Data gaps and bias: Occurrence data from GBIF often have spatial, temporal, and taxonomic biases that should be accounted for in models and trend analyses. ● Model uncertainty: Projections of future distributions carry uncertainty related to climate model selection, species' dispersal assumptions, and model parameterization.

BAT for assessing temporal trends for terrestrial species (HD)

Task	Assess temporal trends in the occurrence of terrestrial species listed in the annexes of the HD and BD.
Data	Occurrence data: ● GBIF records for terrestrial species listed in the HD Annexes (II, IV) Predictor data: ● Data on spatiotemporal trends in sampling effort, for example number of occurrences for target-group taxa ● For occupancy models: environmental predictor data similar to SDMs (climate, soil, topography, etc.)
Methods	Trends could be calculated using: 1. Occupancy models, which separately model species occurrence and the observation process. 2. The Frescalo method, which relies on local frequency scaling to correct for spatiotemporal sampling biases 3. Simple summaries of species occurrence number trends relative to spatiotemporal indicators of sampling effort. For bird species, where applicable, create separate predictions for breeding and wintering ranges.
Outputs	Graphs and statistics summarizing trends in species occurrences/ richness per spatial unit (Natura2000 site, Member state) over time
Challenges	● Thorough calculation of population trends usually requires systematically collected data (i.e., survey-type data), including information on absences

	and sampling effort. The lack of this information will make inferences of trends less robust.

BAT for assessing spatial distributions of freshwater species (HD)

Task	Assess the distribution of freshwater species listed in the annexes of the HD and BD: current and future potential distributions

Data	Occurrence data:
	● GBIF records for freshwater species listed in the HD Annexes (II, IV)
	Predictor data:
	● CHELSA Bioclim
	● Land cover/use
	● Elevation
	● Hydrographic data (https://hydrography.org/hydrography90m/hydrography90m_layers)
	    ● River basin and subcatchment
	    ● Flow accumulation
	    ● Stream slope
	    ● Stream order
	    ● Stream power and compound topographic index
	● Environment90m dataset - contains harmonized data of the above predictors by subcatchment in table format (txt) https://hydrography.org/environment90m/environment90m_layers)
	● Other factors: Human footprint index or settlement index
	
	The selection of specific datasets depends on their integration in the BMD data catalogue.

Methods	1. Current potential distribution: Apply nested or hierarchical SDMs. Contrast potential habitat suitability predictions from SDMs with observed (realized) distributions from occurrence data to assess range filling and identify suitable but unoccupied habitat. 2. Future potential distributions: Project SDMs using future climate scenarios (e.g., SSP pathways) to estimate potential range shifts.
Outputs	Maps of current species occurrences; maps of current and future potential distributions based on SDMs; analysis of projected species range shifts under future climate scenarios.
Challenges	● Data gaps and bias: Occurrence data from GBIF often have spatial, temporal, and taxonomic biases that should be accounted for in models and trend analyses. ● Model uncertainty: Projections of future distributions carry uncertainty related to climate model selection, species' dispersal assumptions, and model parameterization.

BAT for assessing temporal trends for freshwater species (HD)

Task	Assess temporal trends in the occurrence of freshwater species listed in the annexes of the HD and BD.
Data	Occurrence data: ● GBIF records for freshwater species listed in the HD Annexes (II, IV) Predictor data: ● Data on spatiotemporal trends in sampling effort, for example number of occurrences for target-group taxa ● For occupancy models: environmental predictor data similar to SDMs (climate, soil, topography, etc.)
Methods	Trends could be calculated using: 1. Occupancy models, which separately model species occurrence and the observation process. 2. The Frescalo method, which relies on local frequency scaling to correct for spatiotemporal sampling biases 3. Simple summaries of species occurrence number trends relative to spatiotemporal indicators of sampling effort.
Outputs	Graphs and statistics summarizing trends in species occurrences/ richness per spatial unit (Natura2000 site, Member state) over time
Challenges	Thorough calculation of population trends usually requires systematically collected data (i.e., survey-type data), including information on absences and sampling effort. The lack of this information will make inferences of trends less robust.

BAT for assessing spatial distributions of marine species (HD)

Task	Assess the distribution of marine species listed in the annexes of the HD and BD: current and future potential distributions
Data	Occurrence data: ● GBIF records for marine species listed in the HD Annexes (II, IV) ● OBIS (Ocean Biodiversity Information System; https://obis.org) for global marine species observations. ● EMODnet Biology (https://emodnet.ec.europa.eu/en/biology) for harmonized biological datasets across European marine regions. Predictor data: ● Bio-ORACLE v2.2 (https://www.bio-oracle.org/downloads-to-email-v2.php?version=2_2) and MARSPEC (https://marspec.weebly.com/) datasets for marine environmental variables (e.g., sea surface temperature, salinity, chlorophyll-a, current velocity, dissolved oxygen, pH); ● EMODnet Seabed Habitats (EMODnet Seabed Habitats) for benthic substrate type, depth zones, and habitat classification;

	● EMODnet Chemistry marine chemical data (https://emodnet.ec.europa.eu/en/chemistry); ● EMODnet Physics in situ ocean physics time-series data and vertical profiles (https://emodnet.ec.europa.eu/en/physics); ● Copernicus Marine Service (CMEMS - https://marine.copernicus.eu/) for physical and biogeochemical parameters at high spatial and temporal resolution.
Methods	1. Current potential distribution: Apply nested or hierarchical SDMs. Contrast potential habitat suitability predictions from SDMs with observed (realized) distributions from occurrence data to assess range filling and identify suitable but unoccupied habitat. 2. Future potential distributions: Project SDMs using future climate scenarios (e.g., SSP pathways) to estimate potential range shifts.
Outputs	Maps of current species occurrences; maps of current and future potential distributions based on SDMs; analysis of projected species range shifts under future climate scenarios.
Challenges	● Data gaps and bias: Occurrence data from GBIF often have spatial, temporal, and taxonomic biases that should be accounted for in models and trend analyses. ● Model uncertainty: Projections of future distributions carry uncertainty related to climate model selection, species' dispersal assumptions, and model parameterization.

BAT for assessing temporal trends for marine species (HD)

Task	Assess temporal trends in the occurrence of marine species listed in the annexes of the HD and BD.
Data	Occurrence data: ● GBIF records for marine species listed in the HD Annexes (II, IV) ● OBIS (Ocean Biodiversity Information System; https://obis.org) for global marine species observations. ● EMODnet Biology (https://emodnet.ec.europa.eu/en/biology) for harmonized biological datasets across European marine regions; Predictor data: ● Bio-ORACLE v2.2 (https://www.bio-oracle.org/downloads-to-email-v2.php?version=2_2) and MARSPEC (https://marspec.weebly.com/) datasets for marine environmental variables (e.g. sea surface temperature, salinity, chlorophyll-a, current velocity, dissolved oxygen, pH); ● EMODnet Seabed Habitats (EMODnet Seabed Habitats) for benthic substrate type, depth zones, and habitat classification; ● EMODnet marine chemical data (https://emodnet.ec.europa.eu/en/chemistry);

	• EMODnet in situ ocean physics time-series data and vertical profiles (https://emodnet.ec.europa.eu/en/physics); • Copernicus Marine Service (CMEMS - https://marine.copernicus.eu/) for physical and biogeochemical parameters at high spatial and temporal resolution.
Methods	Trends could be calculated using: 1. Occupancy models, which separately model species occurrence and the observation process. 2. The Frescalo method, which relies on local frequency scaling to correct for spatiotemporal sampling biases 3. Simple summaries of species occurrence number trends relative to spatiotemporal indicators of sampling effort.
Outputs	Graphs and statistics summarizing trends in species occurrences/ richness per spatial unit (per Member State, and Natura2000 sites) over time
Challenges	Thorough calculation of population trends usually requires systematically collected data (i.e., survey-type data), including information on absences and sampling effort. The lack of this information will make inferences of trends less robust.

BAT for assessing spatial distributions of birds (BD)

Task	Assess the distribution of bird species listed in Annex I of the BD: current and future potential distributions

Data	Occurrence data:
	• GBIF records for bird species listed in the BD Annex I
	• Global occurrence data is important to characterize the full environmental niche of the species
	• Regional occurrence data (e.g., within Europe) can then be used to build regional SDMs that use outputs of global models as an input (nested distribution modeling approach)
	Predictor data:
	• CHELSA Bioclim: 19 bioclimatic variables as input to global SDMs
	• Chelsa Bioclim + other environmental predictors as inputs to regional SDMs. Potentially useful variables include datasets on human landscape modification, transportation infrastructure (roads & railways), soil conditions, land cover, topographic variables
	    ○ The selection of specific datasets depends on their integration in the BMD data catalogue
	    ○ For seabirds, additional marine environmental variables could be incorporated

Methods	1. Current potential distribution: Apply nested or hierarchical SDMs. Contrast potential habitat suitability predictions from SDMs with observed (realized) distributions from occurrence data to assess range filling and identify suitable but unoccupied habitat.

	2. Future potential distributions: Project SDMs using future climate scenarios (e.g., SSP pathways) to estimate potential range shifts. 3. Where applicable, create separate predictions for breeding and wintering ranges.
Outputs	Maps of current species occurrences; maps of current and future potential distributions based on SDMs; analysis of projected species range shifts under future climate scenarios.
Challenges	● Data gaps and bias: Occurrence data from GBIF often have spatial, temporal, and taxonomic biases that should be accounted for in models and trend analyses. ● Model uncertainty: Projections of future distributions carry uncertainty related to climate model selection, species' dispersal assumptions, and model parameterization.

BAT for assessing temporal trends for birds (BD)

Task	Assess temporal trends in the occurrence of terrestrial species listed in the Annex I of the BD.
Data	Occurrence data: ● GBIF records for terrestrial species listed in the BD Annex I Predictor data: ● Data on spatiotemporal trends in sampling effort, for example number of occurrences for target-group taxa ● For occupancy models: environmental predictor data similar to SDMs (climate, soil, topography, etc.) ● For seabirds, additional marine environmental variables could be incorporated
Methods	Trends could be calculated using: 1. Occupancy models, which separately model species occurrence and the observation process. 2. The Frescalo method, which relies on local frequency scaling to correct for spatiotemporal sampling biases 3. Simple summaries of species occurrence number trends relative to spatiotemporal indicators of sampling effort. For bird species, where applicable, create separate predictions for breeding and wintering ranges.
Outputs	Graphs and statistics summarizing trends in species occurrences/ richness per spatial unit (Natura2000 site, Member state) over time
Challenges	● Thorough calculation of population trends usually requires systematically collected data (i.e., survey-type data), including information on absences and sampling effort. The lack of this information will make inferences of trends less robust.

eDNA-based BAT for marine species

Task	Environmental DNA based monitoring of hard-bottom marine biodiversity and an autonomous reef monitoring structure.
Data	● Marker genes (16S/18S rRNA, ITS and COI) ● High throughput sequence (PEMA Sequence Retriever) ● WoRMS: An authoritative classification and catalogue of marine names (https://www.marinespecies.org/)
Methods	● Three existing eDNA based workflows for identification of species in marine environments can be improved. All three workflows share a common core analysis structure: The process begins by employing the PEMA pipeline to analyze the raw eDNA sequencing data. The resulting Operational Taxonomic Units are then linked to species records. This step determines the status of each identified species at the sample site. The features of the workflow: ○ Workflow 1 - DNA-based trend for the taxa: does Nonmetric Multidimensional Scaling of the PEMA output data. ○ Workflow 2 - DNA-based trend in the distributional range of the taxa: analyses the geographical expansion of the identified species. ○ Workflow 3 - DNA-based taxonomic relatedness: measures the phylogenetic and taxonomic relatedness. ● The three workflows use the following algorithms / pipelines: ○ Pipeline for Environmental DNA Metabarcoding Analysis (PEMA, Zafeiropoulos et al., 2020) is used to process raw sequencing data from environmental samples and provide information on the community composition of organisms present. It supports the downstream analysis of four marker genes (16S/18S rRNA, ITS and COI), by allowing the user to train the classifiers with custom reference databases (https://github.com/hariszaf/pema). ○ Data analysis by calling the taxondive (Clarke and Warwick, 2001) (solution provided by vegan), which finds indices of taxonomic diversity and distinctness, which are averaged taxonomic distances among species or individuals in the community (https://github.com/vegandevs/vegan). ○ metaMDS (Kruskal, 1964): Nonmetric Multidimensional Scaling with stable solution from random starts, axis scaling and species scores (https://github.com/vegandevs/vegan).
Outputs	Clustering into Operational Taxonomic Units (OTUs) or inferring Amplicon Sequence Variants (ASVs). Figures of indices of taxonomic diversity and distinctness, which are averaged taxonomic distances among species or individuals in the community.

	By integrating with the Biodiversity meets Cubes framework, the species occurrence data from our workflows will be transformed into standardized data cubes, making them readily usable across other virtual research environments.
Challenges	Reference database incompleteness: The single biggest hurdle. The taxonomic assignment step is only as good as the reference databases (e.g., SILVA, UNITE). PEMA uses tools like Swarm and VSEARCH for clustering and error mitigation, but these errors can never be fully eliminated.

5. Technical implementation

5.1. Spatial and temporal analyses (habitat types)

For assessing the distribution and status of habitat types, recently created data products based on EUNIS classification can be used (Si-Moussi et al., 2025; Picek et al., 2025). Based on these map products, visualizations and summary statistics for user-defined areas of interest can be provided. To make habitat type information more relevant to monitoring efforts under the HD, a translation of EUNIS-based data to the Annex I habitat type classification would be desirable. For this purpose, available crosswalks should be explored, although direct 1-to-1 translations will likely only be possible in certain cases. In addition, these habitat type maps can be linked to species occurrence data from GBIF to infer the presence of species per habitat type, including species targeted by the relevant EU regulations (i.e., HD, BD). Linking map products to satellite time series data (e.g., Sentinel-2) may provide additional opportunities to assess the status and changes in habitat types, for example based on the calculation and trend classification of vegetation indices (e.g., NDVI) based on satellite time series algorithms. Key challenges for this type of analysis include a potential lack of suitable ground-truth reference data, which may be necessary to properly characterize the status of habitat types from satellite images, as well as the high computational demand associated with processing large volumes of satellite time series data. Thus, the implementation of these BAT features will depend on the design and configuration of the BMD data space and computational infrastructure.

5.2. Spatial analyses (species-level)

Species distribution models will be built using state-of-the-art approaches for modeling invasive species. Specifically, considering the following aspects will be important:

1. Accounting for spatial sampling bias: Given the use of GBIF data mainly containing opportunistically sampled presence-only data, the models should account for spatial sampling bias. This can be achieved, for example, via specific background-sampling approaches (e.g., target-group background sampling, Barber et al. 2023).
2. Minimizing impacts of niche truncation: When SDMs are fitted within spatial extents that do not encompass the whole species range, the estimated environmental niche can be truncated, making the model less reliable when projecting across space and time. To avoid this issue, global occurrence data (encompassing the entire range) should be used. Spatially nested SDMs offer an effective approach to address this issue (Guisan et al. 2025). The combination of global and

regional models has also been shown to be effective for improving SDMs of IAS and understanding invasion dynamics (Gallien et al., 2012).

3. Use of well-performing SDM algorithms: Flexible machine learning algorithms such as Maxent, random forest, and boosted regression trees yield the highest predictive performance for presence-only data in most cases (Valavi et al. 2022). However, hyperparameter optimization is important (see below).
4. Optimization (tuning) of algorithm hyperparameters: Optimizing algorithm hyperparameters is important to help regularize models and avoid overfitting. This can be achieved based on cross-validations of the model. To simulate model transfer, spatial/environmental block cross-validation can be used (Roberts et al. 2017; Valavi et al. 2018)
5. Identification of model extrapolation: The training data available for building SDMs often do not capture the full domain of environmental conditions for which predictions should be made. This causes model extrapolation. Regions where extrapolation occurs should be flagged as having higher uncertainty. Extrapolation can be detected, for example, using a so-called MESS analysis (univariate extrapolation (Elith et al. 2010)), or based on the Mahalanobis distance in order to also capture multivariate extrapolation (new combinations of variable values (Mesgaran et al. 2014)).
6. Considering multiple climate scenarios and models: For SDM projections to future climate conditions, options should be provided to compare different SSP (emission) scenarios. Bioclimatic variables based on different climate models should either be combined into ensemble predictions (i.e., averaging across all climate models), or the range/uncertainty across predictions based on individual climate models should be provided.

5.3. Temporal analyses (species-level)

Without systematically collected data on species abundance, which most likely will not be available for this use case, approaches that robustly estimate population trends are limited. A robust approach for opportunistically sampled presence-only data (i.e., corresponding to most data available via GBIF) are occupancy models. While occupancy models require so-called detection histories containing presence-absence data collected through repeated visits to a set of study sites (e.g., collected through surveys), opportunistic occurrence data can be aggregated into detection histories for their use in occupancy models (Isaac et al. 2014). Occupancy models employ a hierarchical modeling approach to separately model species occupancy (presence/absence) and the observation process (detection/non-detection). In the context of opportunistic data, non-detections can be inferred from visits to a site (e.g., pixel in a raster grid) without the target species being detected, while the length of species lists compiled by observers ("list length") can be used as a proxy for sampling effort. For details on the implementation of occupancy models for opportunistic data, see, for example: Kéry et al (2010), van Strien et al. (2013), Dennis et al. (2017). In R, occupancy models can be fit, for example, using the unmarked package (https://cran.r-project.org/web/packages/unmarked/index.html).

A second robust approach for analyzing temporal trends based on opportunistic occurrence data is the Frescalo method (Goury et al. 2025). Instead of the site-level, Frescalo operates at a broader, neighborhood scale (typically around 10-100 km size) to correct for uneven recording effort. Its core principle is local frequency scaling, which adjusts the observed frequency of a species based on how well-recorded a given area and time period were, using other species as a reference. For details on the implementation of the Frescalo method for unstructured data (Goury et al. 2025).In R, the Frescalo method can be implemented using the sparta package (https://github.com/BiologicalRecordsCentre/sparta).

Finally, trend indices could be generated by setting occurrence record numbers in relation to indicators of sampling effort (for example, number of records for target-group taxon, or number of observers), could be calculated (Knape et al. 2022). In addition, a resampling-based abundance index could be calculated (Zbinden et al., 2014).

The three described approaches differ in their levels of data availability requirements (occupancy models > Frescalo > trend indices) and which of them can be used in BATs thus depends on data availability for different regions/taxa. For comparability and consistency, using the same approaches across target IAS and sites (Natura2000 sites and/or Member states) would be desirable. Which of these approaches will be feasible will depend on the spatiotemporal availability of occurrence records for IAS, as well as for reference species used to infer sampling effort.

5.4. User interface and outputs

User interfaces for the BATs should be easily accessible to non-technical users. The selection of target species and/or target spatial units (e.g., Natura2000 sites, Member states) should thus be implemented via a user-friendly dashboard, for example using (searchable) dropdown menus, or similar. The selection of target species should be informed by the EU- and national-level species lists, depending on the area of interest specified by the user. Given the non-technical user audience, BATs should run largely automatized based on minimal user input (i.e., not requiring manual parametrization of models or analyses).

Outputs can be grouped into visual and non-visual outputs (e.g., graphs/maps vs. numbers), as well as displayed outputs and downloadable outputs. Displayed outputs could be shared with users via the visualisation engine developed in Work Package 4. Occurrence data points and SDM prediction maps can be displayed in interactive web maps. Similarly, graphs for temporal trends could be displayed in interactive plots.

Downloadable outputs could be provided to users via the Single Access Point (SAP) developed within BMD (Work Package 6). These may include:

- Continuous and binary (thresholded) distribution maps, for example in a GeoTIFF format
- Species occurrence points, for example in a CSV or Shapefile format
- Tables containing summaries of occurrence numbers and distribution area trends, for example in a CSV format
- Underlying R scripts that run the implemented analyses (only if this is requested by stakeholders/users)

BATs should be accompanied by non-technical explanations of the implemented analyses. Uncertainties associated with the outputs should be clearly stated, also regarding the general limitations of SDMs. For example, presence-only SDMs only predict the potential distribution of a species (environmentally suitable habitat), while the realized distribution of a species is furthermore influenced by other factors (

biotic interactions with other organisms and dispersal limitations). When calculating range sizes based on model predictions (i.e., binarizing continuous predictions of habitat suitability), the influence of threshold selection should be indicated through the comparison of multiple thresholds.

6. Conclusions and future steps

This MS26 milestone presents the conceptual framework for integrating key use cases based on the HD and BD into the BATs. By aligning regulatory requirements with standardized data, design, and outputs, the proposed BATs covers habitat information, current and future species distributions, and temporal species occurrence trends.

Stakeholder feedback will play a central role in refining and validating the BATs. Input from Natura 2000 site managers, national environmental agencies, and NGOs will inform the design of BATs, including the utility of results, workflow structure, and user interface. This iterative process will ensure the BATs are practical, user-friendly, and directly supportive of evidence-based decision-making.

The next steps include initial prototyping of BATs, stakeholder co-design workshops led by WP1, and technical integration, in line with the project roadmap. Continuous involvement of stakeholders throughout these phases will ensure the BATs not only meet reporting requirements under Article 17 (HD) and Article 12 (BD) but also effectively address the monitoring needs of end users.

7. Acknowledgements

The authors gratefully thank Wojciech Mróz for his helpful feedback and suggestions. We also acknowledge the European Commission for its financial support through the EU Horizon Europe Research and Innovation Action (Grant Agreement No. 101181294, BMD project).

8. References

Barber, R. A., Ball, S. G., Morris, R. K., and Gilbert, F., 2022. Target-group backgrounds prove effective at correcting sampling bias in Maxent models. *Diversity and Distributions*, *28*(1), 128–141. https://doi.org/10.1111/ddi.13442

Clarke, K. R., and Warwick, R. M., 2001. A further biodiversity index applicable to species lists: variation in taxonomic distinctness. *Marine Ecology Progress series*, *216*, 265-278. https://doi.org/10.3354/meps

Dennis, E. B., Morgan, B. J., Freeman, S. N., Ridout, M. S., Brereton, T. M., Fox, R., et al., 2017. Efficient occupancy model-fitting for extensive citizen-science data. *PloS one*, *12*(3), e0174433. https://doi.org/10.1371/journal.pone.0174433

EEA., 2020. State of nature in the EU. European Environment Agency.

Elith, J., Kearney, M., and Phillips, S., 2010. The art of modelling range-shifting species. Methods in *Ecology and Evolution*, *1*(4), 330–342. http://www.respond2articles.com/MEE/

European Commission., 2021. EU Biodiversity Strategy for 2030.

Gallien, L., Douzet, R., Pratte, S., Zimmermann, N. E., and Thuiller, W., 2012. Invasive species distribution models–how violating the equilibrium assumption can create new insights. *Global Ecology and Biogeography*, *21*(11), 1126–1136. https://doi.org/10.1111/j.1466-8238.2012.00768.x

Gantioler, S., Rayment, M., ten Brink, P., McConville, A., Kettunen, M., and Bassi, S., 2014. The costs and socio-economic benefits associated with the Natura 2000 network. *International Journal of Sustainable Society*, *6*(1-2), 135–157. https://doi.org/10.1504/IJSSOC.2014.057894

Goury, R., Bowler, D. E., Harrower, C., Münkemüller, T., Vallet, J., Yearsley, J., et al., 2025. A practical guide to species trend detection with unstructured data using local frequency scaling (Frescalo). Preprint. https://ecoevorxiv.org/repository/view/9467/

Guisan, A., Chevalier, M., Adde, A., Zarzo-Arias, A., Goicolea, T., Broennimann, O., et al., 2025. Spatially nested species distribution models (N-SDM): An effective tool to overcome niche truncation for more robust inference and projections. *Journal of Ecology*, *113*(7), 1588–1605. https://doi.org/10.1111/1365-2745.70063

IPBES., 2019. Global assessment report on biodiversity and ecosystem services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services. E. S. Brondizio, J. Settele, S. Díaz, and H. T. Ngo (editors). IPBES secretariat, Bonn, Germany. 1148 pages. https://doi.org/10.5281/zenodo.3831673

Isaac, N. J., van Strien, A. J., August, T. A., de Zeeuw, M. P., and Roy, D. B., 2014. Statistics for citizen science: extracting signals of change from noisy ecological data. *Methods in Ecology and Evolution*, *5*(10), 1052–1060. https://doi.org/10.1111/2041-210X.12254

Jaureguiberry, P., Titeux, N., Wiemers, M., Bowler, D. E., Coscieme, L., Golden, A. S., et al., 2022. The direct drivers of recent global anthropogenic biodiversity loss. *Science Advances*, *8*(45), eabm9982. https://doi.org/10.1126/sciadv.abm9982

Kery, M., Royle, J. A., Schmid, H., Schaub, M., Volet, B., Häfliger, G., and Zbinden, N., 2010. Site-occupancy distribution modeling to correct population-trend estimates derived from opportunistic observations. *Conservation Biology*, *24*(5), 1388–1397. https://doi.org/10.1111/j.1523-1739.2010.01479.x

Knape, J., Coulson, S. J., van der Wal, R., and Arlt, D., 2022. Temporal trends in opportunistic citizen science reports across multiple taxa. *Ambio*, *51*(1), 183–198. https://doi.org/10.1007/s13280-021-01550-w

Kruskal, J. B., 1964. Nonmetric multidimensional scaling: a numerical method. Psychometrika, *29*(2), 115-129. https://doi.org/10.1007/BF02289694

Mesgaran, M. B., Cousens, R. D., and Webber, B. L., 2014. Here be dragons: a tool for quantifying novelty due to covariate range and correlation change when projecting species distribution models. *Diversity and Distributions*, *20*(10), 1147–1159. https://doi.org/10.1111/ddi.12209

Oeser, J., 2025. BMD WP5 Handbook (2nd draft, 30 June 2025). Unpublished internal document.

Picek, L., Leblanc, C., Joly, A., Bonnet, P., Palard, R. and Servajean, M., 2025. GeoPl@ ntNet: A Platform for Exploring Essential Biodiversity Variables. *arXiv preprint,* arXiv:2511.13790. https://doi.org/10.48550/arXiv.2511.13790

Rayment, M., Pieterse, M., Daly, E., Gerdes, H., Lago, M., Lang, S., Markandya, A., Nunes, P., Ding, H., Tinch, R. and Dickie, I., 2011. Estimating the Overall Economic Value of the Benefits provided by the Natura 2000 Network.

Roberts, D. R., Bahn, V., Ciuti, S., Boyce, M. S., Elith, J., Guillera-Arroita, G., et al., 2017. Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography*, *40*(8), 913–929. https://doi.org/10.1111/ecog.02881

Si-Moussi, S., Hennekens, S., Mücher, S., De Keersmaecker, W., Chytrý, M., Agrillo, E., Attorre, F., Biurrun, I., Bonari, G., Čarni, A. and Ćušterevska, R., 2025. EUNIS habitat maps: enhancing thematic and spatial resolution for Europe through machine learning. *Scientific Data*, *12*(1), 1976. https://doi.org/10.1038/s41597-025-06235-7

Valavi, R., Elith, J., Lahoz-Monfort, J. J., and Guillera-Arroita, G., 2018. blockCV: An r package for generating spatially or environmentally separated folds for k-fold cross-validation of species distribution models. *Biorxiv*, 357798. https://doi.org/10.1111/2041-210X.13107

Valavi, R., Guillera-Arroita, G., Lahoz-Monfort, J. J., and Elith, J., 2022. Predictive performance of presence-only species distribution models: a benchmark study with reproducible code. *Ecological Monographs*, *92*(1), e01486. https://doi.org/10.1002/ecm.1486

Van Strien, A. J., Van Swaay, C. A., and Termaat, T., 2013. Opportunistic citizen science data of animal species produce reliable estimates of distribution trends if analysed with occupancy models. *Journal of Applied Ecology*, *50*(6), 1450–1458. https://doi.org/10.1111/1365-2664.12158

Zafeiropoulos, H., Viet, H. Q., Vasileiadou, K., Potirakis, A., Arvanitidis, C., Topalis, P., and et al., 2020. PEMA: a flexible pipeline for environmental DNA metabarcoding analysis of the 16S/18S ribosomal RNA, ITS, and COI marker genes. *GigaScience*, *9*(3), giaa022. https://doi.org/10.1093/gigascience/giaa022

Zbinden, N., Kéry, M., Häfliger, G., Schmid, H., and Keller, V., 2014. A resampling-based method for effort correction in abundance trend analyses from opportunistic biological records. *Bird Study*, *61*(4), 506–517. https://doi.org/10.1080/00063657.2014.969679

Zurell, D., Franklin, J., König, C., Bouchet, P.J., Dormann, C.F., Elith, J., Fandos, G., Feng, X., et al., 2020. A standard protocol for reporting species distribution models. *Ecography*, *43*, 1261–1277. https://doi.org/10.1111/ecog.04960