biodiversa+
European Biodiversity Partnership

EUROPEAN PARTNERSHIP

# Report on the harmonisation and interoperability of datasets across regions and countries

Co-funded by the European Union

# Document Information

| | |
|---|---|
| **Grant Agreement number:** | 101052342 |
| **Project acronym** | Biodiversa+ |
| **Project full name** | The European Biodiversity Partnership |
| **Biodiversa+ duration:** | 7 years |
| **Biodiversa+ start date** | Start date: 1st October 2021 |
| **More information about Biodiversa+** | Website: www.biodiversa.org<br>Email: contact@biodiversa.org<br>@BiodiversaPlus<br>Biodiversa+ |

| | |
|---|---|
| **Deliverable title:** | D2.2 Report on the harmonisation and interoperability of datasets across regions and countries |
| **Authors:** | Alberto Basset (MUR), Senem Onen Tarantini (MUR), Hilde Eggermont (BelSPO), Cécile Mandon (FRB), Petteri Vihervaara (MoE_FI). |
| **Contributors** | Guillaume Body (OFB), Karel Chobot (NCA CZ), Andreas Hilpold (PROV_BZ), Triin Sellis (MoE_EE), Radoslav Stanchev (ExEA) |
| **Work package:** | WP2 Promote and support transnational biodiversity monitoring |
| **Task or sub-task title:** | Subtask 2.2.2: Harmonise databases and data interoperability |
| **Picture credits:** | © Pixabay (cover picture) |
| **Release date:** | April 2023 |

## What is Biodiversa+

Biodiversa+ is the new European co-funded biodiversity partnership supporting excellent research on biodiversity with an impact for policy and society. It was jointly developed by BiodivERsA and the European Commission (DG Research & Innovation and DG Environment) and was officially launched on 1 October 2021.

Biodiversa+ is part of the European Biodiversity Strategy for 2030 that aims to put Europe's biodiversity on a path to recovery by 2030.

The Partnership aims to connect science, policy and practise for transformative change. It currently gathers 80 research programmers and funders and environmental policy actors from 40 European and associated countries to work on 5 main objectives:

1. Plan and support research and innovation on biodiversity through a shared strategy, annual joint calls for research projects and capacity building activities
2. Set up a network of harmonised schemes to improve monitoring of biodiversity and ecosystem services across Europe
3. Contribute to high-end knowledge for deploying Nature-based Solutions and valuation of biodiversity in the private sector
4. Ensure efficient science-based support for policy-making and implementation in Europe
5. Strengthen the relevance and impact of pan-European research on biodiversity in a global context

More information at: https://www.biodiversa.eu/

# Table of contents

# Table of acronyms

| | |
|---|---|
| ABCD | Access to Biological Collections Data Task Group |
| CODATA | Committee on Data of the International Science Council |
| DwC | Darwin Core |
| EBVs | Essential Biodiversity Variables |
| EESVs | Essential Ecosystem Services Variables |
| eLTER | Europe Long-Term Ecosystem Research Network |
| EML | Ecological Metadata Language |
| EOSC | The European Open Science Cloud |
| FAIR | Findability, Accessibility, Interoperability, and Reuse of digital assets |
| INBO | Research Institute for Nature and Forest |
| KCBD | Knowledge Centre for Biodiversity |

# Introduction

The story of gathering data reaches out to Palaeolithic tribes' people (Kapcar, 2011) and that of developing computational and data analysis tools to approx. 3000 years BC, with the realisation of the abacus. In the last fifty years, data acquisition, computational storage and power capacity, analytical and modelling tools for data analysis and visualisation, eScience tools and services have increased impressively and at a continuously growing rate. Different types of sensors are producing data on both abiotic and biotic components of Earth's ecosystems at an unprecedented rate; at the same time, the research e-infrastructures recently developed at the national, European and international level are offering both storage and computational capacity and e-services and e-tools to run modelling on big data as never before. Nevertheless, we still have problems in converting the collected data into knowledge on mechanisms underlying biodiversity status, trends, dynamics and ecosystem functioning.

A critical point is that understanding the organisation of life in the Biosphere is a complex process, not easily achievable simply describing in detail ecosystem structures being based on the interlacing networks of interactions that every single individual establishes with the abiotic environment and with the individuals of same or different species sharing the same spatial unit at the same time. Interactions between conspecific individuals and co-occurring species are (i) not easily estimated, while species occurrence and population density could be; (ii) have a nested organisation in complex interaction networks with direct and indirect effects on individual energetics, behaviour and performances, which affect individual and population fitness; and (iii) require time-consuming experimental approaches for their description and their strength assessment, which limits data production. Moreover, the biological data and at a lower extent the abiotic data required to decode the biodiversity patterns of variation, suffer from dis-homogeneity in the sampling and quantification methodologies, from limited use of standardisation procedures and semantic description.

Finally, the principle of *open data* and *open science* are still not fully embraced by the scientific community and often data are still not described and documented following the FAIRness principles and, therefore, are not findable (F), available (A), interoperable (I) and reusable (R) (Wilkinson et al. 2016).

There is a strong effort to minimise these limitations through development of procedures of metadata standardisation, data and metadata harmonisation and interoperability. On these issues there is a growing attention of national, European and international Initiatives, research infrastructures and large projects to improve the standardisation and FAIRness of data and datasets towards a wider harmonisation and interoperability. This Biodiversa+ report is focused on data and datasets harmonisation and interoperability.

Data harmonisation is a process that aims to transform data from different sources in such a way that they fit together and provide users with a comparable view of data from different studies. This is a process modification of semantics and data structure and adjustment of differences among different labels, measurements, methods, and procedures in a way that is of no apprehension to the end user.

The capacity of using wider data on biodiversity has great importance for better modelling that provides guidelines for decisions on how to manage biological diversity in terms of production and conservation. At that point, we need to standardise data sets that are available in machine-to-machine processes.

Interoperability is the ability of applications to operate across otherwise incompatible systems; it gives the possibility for datasets/ metadata to be merged, and for services to interact, without repetitive manual operations, in the manner that the result is harmonious. It has a direct role in increasing the impact of research, encouraging innovations, and taking the lead on new collaborations between data users and creators. Analysis of big, harmonised datasets on biodiversity can provide better sight into the future of biodiversity, and trend identification. It can help acting and helps act, creating policies on a local, national, and global scales.

This report is a building block, based on the Biodiversa+ work, to contribute to the overall aim of supporting the harmonisation of biodiversity monitoring databases and the data interoperability across Europe.

This report:

1. analyses the state of the art on data and datasets harmonisation and interoperability at the European and global scale and advertises the EuropaBON biodiversity monitoring database[1], which is providing a mapping of how biodiversity data collected in monitoring schemes across Europe flows through different institutions and programs (section 1);
2. synthetically presents the outcomes of two workshops held on data architecture and data workflows harmonisation and interoperability with the contribution of both European/global projects, Initiatives and Research Infrastructures, and national Institutions (section 2);
3. addresses the potential role of Biodiversa+ in supporting data interoperability and harmonisation for better biodiversity monitoring in Europe, with a concrete portfolio of activities (section 3);
4. considers some suggestions on how to ensure deeper and more effective harmonisation and interoperability of biodiversity monitoring data and databases in Europe (section ‘Conclusions’).

[1] EuropaBON: https://europabon.org/?page_id=2513

# 1. State of the art on data/ dataset harmonisation and interoperability

The need to combine and analyse biodiversity data coming from different sources and with different standards represents one of the most challenging aspects. Indeed, biodiversity data are highly complex due to the number of involved data types, the heterogeneity of data collection procedures, the different formats used for their description and the level of accessibility. These might limit harmonisation and interoperability.

- Types of the data show variety as species distribution map, species occurrence records, species abundance, species threat status, species abundance trend, species behaviour / trait, habitat and vegetation maps, species genetic composition, community composition, ecosystem functions, intraspecific and intrapopulation trait data and their variability. (Underwood, Taylor, & Tucker, 2018)
- Source of data: *Primary data* are those data directly collected in the field or in the lab, through sampling or using sensors by researchers; data collections refer to sampling techniques and methodological procedures of sensor characteristics, which have an influence on the data. Data might be then checklists, occurrence data that give information about the location of an individual organism at a particular time, sampling event data that are collected via specific protocols for measuring (monitoring programs, research projects, etc.). *Secondary data* are those data that are produced by researchers or any type of data users through a process of data analysis and modelling, representing different types of synthetic descriptions of biodiversity data characteristics and which clearly refer to the procedure for primary biodiversity data analysis. They include also those data collected through data mining procedures from published, peer-reviewed and/or grey literature.
- Data format: Lately the scientific community tends to organise data in different digital formats such as binary file formats, comma separated values or raster data. However, many datasets still exist in PDF and even JPEG format (especially historic data) so that they are not available for machine-to-machine processes (Hebert et al., 2013).
- Data accessibility: The possibility to access biodiversity data can remain limited by licensing agreement, data security & policy and the cost of non-open access data.

Therefore, biodiversity data need proper description and harmonisation, to ensure full comparability of the primary data which can be openly shared. Multilateral or fully open systems that allow to share biodiversity data represent hence a necessary step forward to deepen knowledge and understanding on biodiversity organisation, status and conservation at any level of spatial scale, from the ecosystem to the global ones, providing enormous societal benefits (Aubry et al., 2022; Scholz et al., 2022; von Wettberg & Khoury, 2022).

Transforming datasets and metadata in a consensual format needs a general acceptance process such as using the same vocabularies, glossaries of terms, sampling and measurement techniques, sampling tools and probes and unit. Vocabularies and glossaries of terms are used to transform datasets and metadata in a consistent format. Vocabularies provide a common language used to refer to and describe the data and all steps and procedures involved in the

primary data collection, while glossaries of terms provide definitions and explanations for the terms used in the data. Vocabularies and glossaries ensure that all users understand the data in the same way, making it easier to share, compare, and analyse the data. They also help to ensure both accuracy (as incorrect terms or definitions can lead to misinterpretation of the data) and actual harmonisation of the data, as all steps in primary data collection and measurement are clearly reported by guaranteeing full data comparability.

Harmonised data are then potentially available for the process of full interoperability, boosting the process of converting collected data and information into new and/or deeper knowledge development. Data interoperability enables data to be used across multiple systems and applications, and for different stakeholders to access and use the same data, regardless of the source.

By following the EOSC Interoperability Framework (European Commission, 2021), interoperability can be analysed by four main points of view (as depicted in Fig 1):

- Technical: the ability of different information technology systems and software applications to communicate and exchange data by using the same data formats and communication protocol(s);
- Semantic: the ability of computer systems to transmit data with unambiguous, shared meaning. Data can be hence transferred meaningfully in a way that allows the receiving system to correctly understand and use the data exchanged (Heiler S., 1995). The usage of controlled vocabularies and relevant ontologies is a key to achieve semantic interoperability (Stocker et al., 2018).
- Organisational: the way in which organisations align their business processes, responsibilities and expectations to achieve commonly agreed and mutually beneficial goals.
- Legal: it deals, in particular, with how data should be re-used. Legal interoperability can be achieved when the conditions of use for each dataset are met, and when users can legally access and use each dataset without seeking authorization from data rights holders on a case-by-case basis. The ideal goal for legal interoperability is when datasets are positively identified as having no legal restrictions (RDA-CODATA Legal Interoperability Interest Group, 2016). However, a CC0 licence does not attribute the adequate acknowledgement that is very relevant in the scientific context. The CC-BY licence instead explicitly requires acknowledgement and attribution.

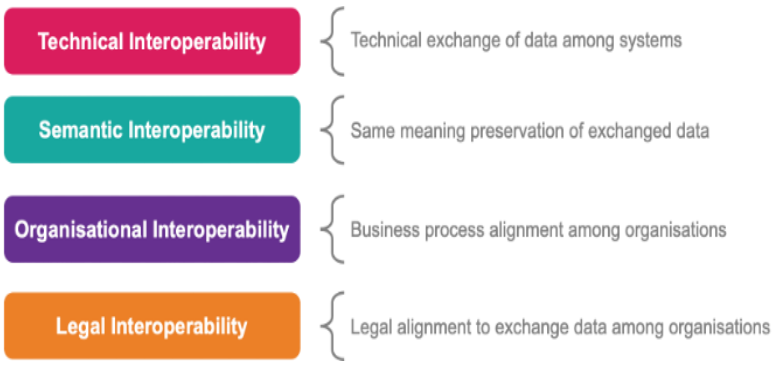



Fig.1 The interoperability levels according to the EOSC Interoperability Framework

As such, many factors influence data interoperability and interoperability across infrastructures. Moreover, the variety in data management practices, the complexity and diversity of the available tools and terminology resources led to an increase of heterogeneity at multiple levels severely hampering the data integration and interoperability.

As data volumes grow, interoperable methods of accessing and working with biodiversity data will continue to be essential to allow the many technical, scientific, and also administrative fields to work together. Standardised methods for direct access to on-line biodiversity data will also continue to rapidly mature. In order to ensure that the full potential of biodiversity data is realised, it is essential to support common, interoperable data formats, tools and delivery standards. Implementing EBVs, for example, requires the global cooperation of biodiversity researchers and research infrastructures to serve comparable data sets and analytical capabilities, implying their interoperability.

To foster data interoperability and to adhere to FAIR principles, it is important that data providers adopt common standards to be able to publish data and metadata in standard forms by allowing research infrastructures to combine and federate content across providers. Several technical barriers exist but in general, sharing data depends on the consistent use of agreed standards. Like in many other research domains, “one standard does not fit all” for all the different research infrastructures, which generally collect and store data in the forms and combinations that best meet their needs, that’s why many community-driven standards for biodiversity data exist for different purposes. With the final aim to make biodiversity data more widely findable, accessible and to ensure they can be integrated and reused for different purposes, data providers should consider the most suitable way to expose their data. This means that they should decide which standards would be best recognised by others so that they can ensure adequate data transformation techniques, able to obtain data represented by such standards.

Several initiatives with the corresponding supporting standards have been proposed in literature, but this represents a challenging topic that is in constant evolution and that typically follows the main recommendations for best practices in data sharing among biodiversity data providers (e.g., the annual reports of the TDWG).

The main adopted data standards in the biodiversity domain are:

- Darwin Core, DwC (Darwin Core Task Group, 2009; Wieczorek et al., 2012);
- Access to Biological Collections Data, ABCD (Access to Biological Collections Data Task Group, 2005; Holetschek et al., 2012);
- Network Common Data Form, NetCDF(https://docs.unidata.ucar.edu/netcdf-c/current/index.html).

**Darwin Core (DwC)** is one of the most commonly used data standards in the biodiversity community and it is maintained by the DwC maintenance group. It includes a glossary of terms intended to facilitate the sharing of information about biological diversity by providing identifiers, labels, and definitions. DwC is primarily based on taxa, their occurrence in nature as documented by observations, specimens, samples, and related information. But the simplicity of this standard, established and maintained by Biodiversity Information Standards

(TDWG), has significant limitations when it comes to shaping data from different sources. Indeed, within the TDWG community, DwC extensions are continuously developed to allow data providers to express data elements that go beyond the current version of the DwC and that allow it to support richer and more complex types of biodiversity data.

**The Access to Biological Collections Data (ABCD)** schema is an evolving comprehensive standard for the access to and exchange of data about specimens and observations (primary biodiversity data). The ABCD schema attempts to be comprehensive and highly structured, supporting data from a wide variety of databases and it is compatible with several existing data standards. Moreover, since it defines relations between terms, ABCD is a step towards an ontology for biological collections.

**The Network Common Data Format (NetCDF)** is a self-describing, machine-independent data format that is meant to represent and store array-oriented data. It supports the creation, access, and sharing of array-oriented scientific data and is used by a wide range of biodiversity domains (e.g., marine, atmospheric, soil, etc.) that want a standard way to store data so that they can be efficiently shared and reused.

Biodiversity data have to be also suitably described, including appropriate information like those associated to data source and ownership for example. This implies that also a recognised metadata standard has to be adopted to guarantee the interoperability.

The main adopted metadata standards in the biodiversity domain are:

- Ecological Metadata Language, EML (Jones et al., 2019; Fegraus et al., 2005);
- ISO 19115 (ISO, 2019).
- Data Catalog Vocabulary (DCAT)
- JavaScript Object Notation for Linked Data (JSON-LD)
- PPSR Core

The choice on which one to use depends on specific research needs.

**The Ecological Metadata Language (EML)** is a metadata standard that records information about ecological datasets in a series of modular and extensible XML document types (https://eml.ecoinformatics.org). It is in widespread use in the earth and environmental sciences, and increasingly in other research disciplines as well. Ecological Metadata Language (EML) provides high quality metadata specification to dataset derived from ecological, environmental science. The Ecological Metadata Language (EML) is a standard for describing ecological data. It provides a way to document the context, content, and structure of data sets so that they can be understood by humans and machines alike. EML includes conventions for documenting information about the origin of data, experimental design, sampling methods used in collecting it, as well as how it was processed and analysed. This helps ensure that the data remains understandable over time even if standards or technologies change. Additionally, EML also makes it easier to share metadata across different systems and platforms. EML used by the National Ecological Observatory Network, the Long-Term Ecological Research Network, and the NOAA Office of Ocean Exploration and Research.

**The ISO 19115** is an internationally-adopted schema for describing geographic information and services. It provides information about the identification, the extent, the quality, the spatial and temporal schema, spatial reference, and distribution of digital geographic data.

**Data Catalog Vocabulary (DCAT):** DCAT is a metadata standard for describing data sets stored in a data catalogue. It is used to provide information about the structure and content of datasets, as well as the sources they come from. It is an RDF vocabulary designed to facilitate interoperability between data catalogues published on the Web. DCAT includes conventions for documenting information about the origin of data, experimental design, sampling methods used in collecting it, as well as how it was processed and analysed. This helps ensure that the data remains understandable over time even if standards or technologies change. Examples of initiatives that use DCAT include the European Data Portal and the UK Open Data Initiative.

**JavaScript Object Notation for Linked Data (JSON-LD):** JSON-LD is a metadata standard for describing data sets stored in a linked data format. It is used to provide information about the structure and content of datasets, as well as the relationships between them. JSON-LD is based on the JSON data format and can be used to describe datasets in a more human-readable format. It provides an easy way to represent and exchange structured data on the web, such as biodiversity information. JSON-LD includes conventions for describing relationships between different types of data using terms from existing vocabularies, such as those used by Darwin Core or EML.

**PPSR Core** is a set of global, transdisciplinary data and metadata standards for use in Public Participation in Scientific Research (Citizen Science) projects. These standards are united, supported, and underlined by a common framework illustrating how information is structured within the citizen science domain. This allows data to be used across platforms and projects in a consistent manner, furthering the research goals of the scientific community. At the global level, the PPSR core vocabularies have been adapted by the World Wide Web Consortium (W3C) in its Linked Data Platform. This platform is used to facilitate the development of applications that use linked data and to enable the creation of a global data space. At the European level, the PPSR core vocabularies have been included in the European Commission's European Data Portal, which provides access to open data from European public sector organisations. The portal enables users to discover, access, and reuse datasets from across the European Union. Additionally, the vocabularies have been integrated into the European Commission's Open Data Portal, which provides access to open data from European public sector organisations.

Starting from the work of Hardisty et al., 2019, Table 1 includes the main international Research Infrastructures and initiatives that offer multiple services related to the whole data lifecycle. In particular, the following service categories will be analysed:

- C: data collection and organisation;
- D: data discovery and access;
- A: data analysis and/or visualisation;
- P: data preservation.

Table 1: International Research Infrastructures and initiatives that provide multiple services related to the dataset's lifecycle (*of major relevance for European biodiversity)

| Infrastructure | Services | Endpoint |
|---|---|---|
| ACTRIS Data Centre | D, P | https://actris.nilu.no/ |
| AnaEE Data Portal | D, P | https://data.anaee.eu/ |
| *AQUACOSM | D, P | https://d.aquacosm.eu/ |
| Atlas of Living Australia (ALA) | C, D, A, P | https://www.ala.org.au |
| Arctic Data Center | D, P | https://arcticdata.io |
| Biodiversity Heritage Library (BHL) | C, D, A, P | https://www.biodiversitylibrary.org/ |
| Cary Institute of Ecosystem Studies | D, P | https://www.caryinstitute.org |
| *Catalogue of Life (CoL) | C, D | http://www.catalogueoflife.org/ |
| *Copernicus Services | D, A, P | https://www.copernicus.eu/en/copernicus-services/ |
| Cornell Lab of Ornithology | C, D, P | https://ebird.org/ |
| *DANUBIUS Data Portal | D, A, P | https://gis.geoecomar.ro/danubius/dataportal/ |
| Data Observation Network for Earth (DataONE) | D, P | https://www.dataone.org/ |
| Dryad Digital Repository | D, P | https://datadryad.org/ |
| Earth Data Analysis Center (EDAC) | C, D, A, P | https://edac.unm.edu/ |
| EMSO ERIC | D, P | https://data.emso.eu/ |
| Encyclopedia of Life (EoL) | D | http://eol.org/ |
| Environmental Data Initiative | C, D, A, P | https://environmentaldatainitiative.org/ |
| Environmental System Science Data Infrastructure for a Virtual Ecosystem (ESS-DIVE) | C, D, P | https://ess-dive.lbl.gov |
| ESA Data Registry | C, D, P | https://www.esa.org |

| Infrastructure | Services | Endpoint |
|---|---|---|
| *Europe Long-Term Ecosystem Research Network (eLTER Europe) | C, D, P | https://elter-ri.eu |
| *European Environment Agency (EEA) Data and Maps | D, P | https://www.eea.europa.eu/data-and-maps/ |
| *European Marine Observation and Data Network (EMODnet) | C, D, A, P | https://emodnet.ec.europa.eu/en/ |
| Forest Ecosystem Monitoring Cooperative (FEMC) | D, P | https://www.uvm.edu/femc |
| *Global Biodiversity Information Facility (GBIF) | D, A, P | https://www.gbif.org/ |
| Global Biotic Interactions (GloBI) | D, A | https://www.globalbioticinteractions.org/ |
| Gulf of Mexico Research Initiative | D, P | https://data.gulfresearchinitiative.org |
| IAGOS Data Portal | D, P | https://iagos.aeris-data.fr/ |
| *ICOS Data Portal | D, P | https://www.icos-cp.eu/data-services/ |
| IEDA: EarthChem Library | C, D, P | https://www.earthchem.org/ecl/ |
| IEDA: Marine-Geo Digital Library | C, D, P | https://www.marine-geo.org/ |
| IEDA: US Antarctic Program Data Center | C, D, P | https://www.usap-dc.org/ |
| Integrated Digitized Biocollections (iDigBio) | D, P | https://www.idigbio.org/ |
| JERICO-RI Catalogue | D, P | https://www.jerico-ri.eu/jerico-ri-catalogue/ |
| Knowledge Network for Biocomplexity | C, D, P | https://knb.ecoinformatics.org/ |
| *LifeWatch ERIC | D, P | https://metadatacatalogue.lifewatch.eu |
| Map of Life (MoL) | D, A | http://mol.org/ |
| National Ecological Observatory Network (NEON) | C, D, P | https://data.neonscience.org/ |

| Infrastructure | Services | Endpoint |
|---|---|---|
| National Specimen Information Infrastructure | C, D | http://nsii.org.cn/ |
| NOAA NCEI Environmental Data Archive | C, D, A, P | https://www.ncei.noaa.gov/ |
| PANGAEA Data Publisher for Earth and Environmental Science | C, D, P | https://www.pangaea.de/ |
| Partnership for Interdisciplinary Studies of Coastal Oceans (PISCO) | C, D, A, P | https://www.piscoweb.org/ |
| Programa de Pesquisa em Biodiversidade (PPBio) | C, D, P | https://ppbio.inpa.gov.br/en |
| Rolling Deck to Repository (R2R) | C, D, P | https://www.rvdata.us/ |
| SDI - geo-spatial data catalogue | D, P | https://sdi.eea.europa.eu |
| SeaDataNet | C, D, P | https://www.seadatanet.org/ |
| SIOS Data Access Portal | D, P | https://sios-svalbard.org/ |
| Sistema de Informação sobre a Biodiversidade Brasileira (SiBBr) | C, D, A, P | http://www.sibbr.gov.br/ |
| South African National Biodiversity Institute (SANBI) | C, D, A, P | http://www.sanbi.org/ |
| speciesLink | D | http://www.splink.org.br/ |
| TERN Australia | D, P | https://www.tern.org.au/ |
| TRY Plant Database | C, D, P | https://www.try-db.org/ |
| USGS Science Data Catalog | D, P | https://data.usgs.gov/datacatalog/ |
| VertNET | D | http://vertnet.org/ |

# 2. Biodiversa+ approach to data harmonisation and interoperability

To contribute to the overall aim to achieve a deeper harmonisation and interoperability of biodiversity monitoring data and databases across Europe, Biodiversa+ organised two workshops on biodiversity monitoring data interoperability and harmonisation aimed at reinforcing the connection of the Institutions deputed on monitoring biodiversity at the National level with a few key International Initiatives acting on biodiversity data harmonisation and interoperability.

To connect with national and sub-national biodiversity monitoring databases and initiatives, Biodiversa+:

- Launched a survey and bilateral follow-up interviews in April – June 2022 with all the Biodiversa+ partners involved in biodiversity monitoring to understand how the biodiversity monitoring national landscape is structured.
- Organised a workshop on the 1st of September 2022 on biodiversity monitoring data harmonisation and interoperability with a focus at European and global levels
- Launched a survey focussing on data interoperability and harmonisation for biodiversity monitoring in October / November 2022. The Biodiversa+ partners involved in biodiversity monitoring and other initiatives were invited to complete this survey (see the Annexe).
- Organised a workshop on the 4th of November 2022[2] on data interoperability and harmonisation with a focus on the (sub-)national scale.

The first workshop was dealing with key European/Global Initiatives and organised with a participative approach producing feedback on the key challenges in biodiversity data harmonisation and interoperability and the expected/wished contribution of Biodiversa+. The second workshop was focused on the National experiences on monitoring biodiversity, taking advantage from the feedback of the first one.

## 2.1. For global and European biodiversity monitoring databases

The first workshop was held on the 1st of September 2022, as an online workshop[3] gathering 37 participants representing European/ global initiatives/ databases, (sub-)national initiatives/databases, Biodiversa+ Partners, representatives of European institutions and researchers. The workshop had four main purposes:

- get a better understanding of the data architectures and data workflows of four main European / global initiatives,
- understand main gaps about data harmonisation and interoperability,
- start reflecting on how Biodiversa+ could help improve interoperability

[2] Biodiversa+ (2022). Workshop on data interoperability and harmonisation. 4th of November 2022. https://www.biodiversa.eu/2022/11/05/workshop-on-data-interoperability-and-harmonisation/

[3] Biodiversa+ (2022). Biodiversity monitoring data interoperability and harmonisation. 1st of September 2022 workshop: https://www.biodiversa.eu/2022/09/06/biodiversity-monitoring-data-interoperability-and-harmonisation/

- allow national initiatives developing or having national biodiversity monitoring databases to get more familiar with these global and European initiatives.

During this workshop, the EuropaBON biodiversity monitoring database, GBIF, LifeWatch ERIC and BIOSCAN/iBol presented their data architecture and data flows. Essential Biodiversity Variables were also introduced. During a second stage of the workshop issues preventing biodiversity monitoring data interoperability and possible support from Biodiversa+ were identified (see section 3 of the report).

## 2.2. For national and sub-national biodiversity monitoring databases

### 2.2.1 Survey on biodiversity monitoring governance and interviews with the Biodiversa+ partners

In April – May 2022, Biodiversa+ conducted a biodiversity monitoring survey followed by bilateral interviews with all the Biodiversa+ Ministries of Environment, Environmental Protection Agencies and other interested Biodiversa+ partners. Through this exercise 23 countries were covered. The survey showed that there was an important need from the Biodiversa+ partners to get more collaboration/ solutions/ support for data management and interoperability (see Fig. 2). To help address this need, through this report, Biodiversa+ aims to share national and sub-national examples of good practices and challenges faced at national level for biodiversity monitoring data interoperability and harmonisation (see the Annexe). These showcases can then help other countries build on this experience and further enhance or develop national biodiversity monitoring databases that are more interoperable with other national/sub-national/ European and global databases.

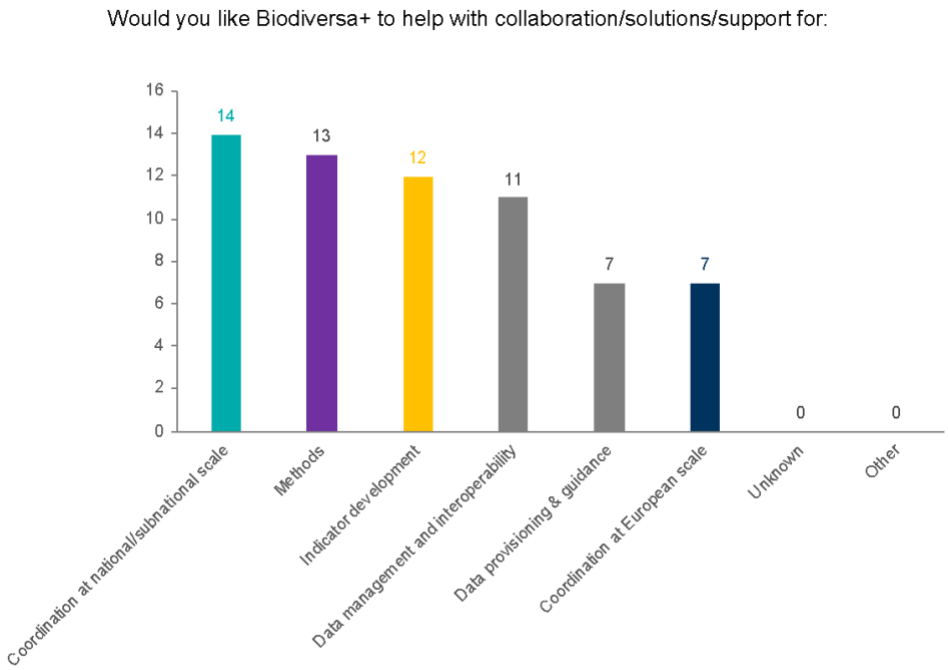



Fig. 2. Results from the Biodiversa+ biodiversity monitoring survey, survey respondents could tick a maximum of 3 options.

### 2.2.2 Biodiversa+ workshop and survey with third organisations

The second workshop was held on the 4th of November 2022, being built on the outcomes of the first. This workshop was an opportunity for (sub-)national initiatives/databases to introduce their data architecture and data workflows and to discuss on how Biodiversa+ could provide support to reinforce data interoperability and harmonisation. This workshop was split into 3 sections. During the first section a keynote speech on data interoperability was given by Hanna Koivula from CSC – IT Centre for Science, the "Meetnetten" Flemish webtool, and the German NFDI4 Biodiversity data architecture and workflows were presented. Next, the outcomes of a Biodiversa+ survey shared before the workshop were introduced. This survey allowed to collect 31 answers, 48% from Biodiversa+ Partners and 52% from (sub-)national and European databases or initiatives[4]. The survey was divided into two sections; the first part allowed the respondents to assess the outcomes of the first workshop and the second part aimed to collect descriptions of national biodiversity monitoring databases (see the Annexe of this report for more information on the survey).

[4] Organisations which answered the survey: Swedish Biodiversity Data Infrastructure (SBDI), Swedish University of Agricultural Sciences (SLU), Bavarian Natural History Collections (SNSB), Belgium Biodiversity Platform, Centre for Ecological Research and Forestry Applications (CREAF), European Environment Agency (EEA), Bulgarian Environment Executive Agency (ExEA), Finnish Environment Institute (SYKE), French museum of natural history (MNHN), German Center for Integrative Biodiversity Research (iDiv), German Federation for Biological Data / NFDI4Biodiversity project, German National Monitoring Center for Biodiversity, German Research Foundation (DFG), Italian National Institute for Environmental Protection and Research (ISPRA), Italian National Research Council, LifeWatch ERIC, Lund University, Croatian Ministry of Economy and Sustainable Development, Institute for Environment and Nature (MESD), Estonian Ministry of Environment, National Biodiversity Data Centre of Ireland, Naturalis Biodiversity Center from the Netherlands, Nature Conservation Agency CZ (NCA CZ), Norwegian Environment Agency, Norwegian Institute for Nature Research, Plant Science and Biodiversity Centre, Slovak Academy of Sciences, Autonomous Province of Bolzano (PROV_BZ), Public Service of Wallonia, Research Institute for Nature and Forest (INBO), University Kassel, Executive Agency for Higher Education, Research, Development and Innovation Funding of Romania.

# 3. Data interoperability and harmonisation challenges and possible support from Biodiversa+

## 3.1. Issues preventing deeper data harmonisation and interoperability

For data interoperability and harmonisation at global level, based on the outcomes of the first Biodiversa+ workshop, three main challenges were identified:

- The lack of satisfactory (meta)data standards
- The lack of strong enough capacity building and knowledge sharing opportunities
- The governance for biodiversity monitoring data interoperability

New technology needs were considered during the workshop as a cause and effect of standard needs and governance issues.

Building on the Biodiversa+ follow-up survey targeting Biodiversa+ partners, (sub-)national and global/European databases, these first three issues received good support: 58% of the respondents considered that these three main challenges were relevant. In the meantime, technology needs become prominent as much as standard needs (see Fig 3).

More details on these challenges are available below.

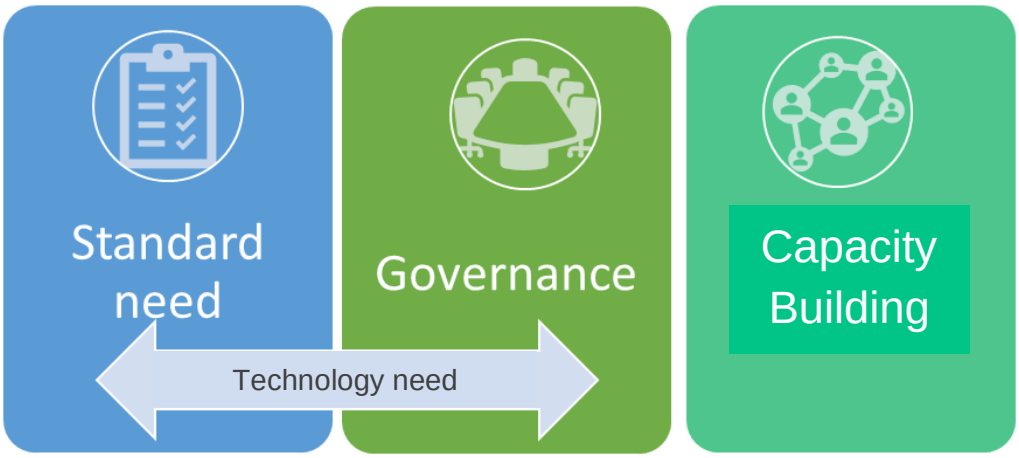



Fig 3: Main data interoperability challenges identified

### 3.1.1 The lack of satisfactory (meta)data standards

The recent proliferation of data standards has been a major boon to the advancement of digital technology, allowing for increased interoperability between systems and enabling data to be more effectively shared and used. However, this proliferation has also highlighted gaps in existing data standards, which are becoming increasingly apparent as new technologies emerge. In particular, the emergence of machine learning, big data analytics and other advanced technologies has highlighted several challenges in terms of data harmonisation and interoperability and metadata management.

These include domain-specific standards that are applied across different sectors; data collection by different agencies with different standards, preventing direct biodiversity data harmonisation and interoperability and the development of an increasing numbers of metadata standards without an effort, or with a limited effort, on their harmonisation with those already existing and widely adopted. Machine learning processes often lack transparency in terms of how they were created or what training sets were used; this impairs confidence on the produced results without human interpretation or validation. Additionally, there is a growing number of ontologies being developed which cannot easily be cross-walked or integrated with each other; this requires initial human investigation and sense-making before any kind of machine-readable repository can be created.

The sheer volume of big data generated by machine observations means that IT infrastructure needs to keep up with storage needs of such observations. Otherwise, large datasets will become unmanageable or unusable due to their size or complexity.

### 3.1.2 The lack of strong enough capacity building and knowledge sharing opportunities

The lack of clear guidance or mandate from funding agencies to adhere to specific standards was stressed as a main challenge for biodiversity monitoring data interoperability. Without such guidance, data sharing initiatives will often become mired in conflicting standards, leading to a lack of consistency and potential incompatibilities between different systems. Furthermore, the absence of a consistent standard leaves researchers struggling to keep up with the ever-changing landscape of technology and data formats. As such, it is vital that funding agencies provide clear guidance on what standards should be followed when collecting or sharing data so as to ensure greater consistency across different research projects. More generally, it was also raised that many data collectors are still unfamiliar with some concepts such as interoperability ontologies and semantics.

The lack of consistent use of provenance tools such as ORCID, ROR or DOIs was identified as another key obstacle for data interoperability. Provenance tools are essential for ensuring that data can be properly attributed back to its original source and for managing the lifecycle of datasets over time. Despite their importance, researchers often fail to use these tools consistently due to a lack of understanding around their purpose or even awareness about their existence in some cases. This means that it can be difficult for organisations or individuals trying to access datasets from multiple sources as they may not have access to all necessary provenance information needed for proper attribution. Capacity building efforts should focus on increasing awareness about these types of tools among researchers so they can begin using them more consistently when collecting and sharing datasets.

Another issue pointed out is the little attention that has been paid by capacity building initiatives towards poorer areas of Europe where resources are sparsely allocated compared with regions such as Western European or Northern American ones, where more resources exist per capita than elsewhere in Europe. This imbalance makes it difficult for people living in many European regions both to access high-quality training opportunities through formal channels

and to demonstrate the expertise gained through informal learning pathways, such as coding boot camps or online courses, which might not be recognised by employers as trustable sources of qualification and experience. These issues need urgent attention to prevent any kind of actual discrimination across Europe in the access to training opportunities related to data harmonisation and interoperability topics.

A last challenge which was identified is related to the semantic artefact development. Such development requires expert knowledge not only of one institution's own internal data architectures but of multiple different ones in order to achieve a full and seamless exchange between different systems. This means that the personnel of different organisations need to understand how these organisations store process metadata and order craft solutions will work universally rather than just within one organisation's own ecosystem making sure all stakeholders can benefit from this type of collaboration. To achieve this goal, capacity building efforts on providing proper training on the architectures used by various organisations in order to equip the personnel with skills needed bridge gaps between disparate systems, thereby enabling better collaboration across sectors.

### 3.1.3 Governance related challenges

Lots of existing data is collected in templates which lose out on raw information. For example, mandated reporting requirements from EU institutions often require data collection into predefined templates which break down complex information into simpler forms but lose out on granular detail, meaning that much important contextual information gets lost along the way. Such practices make it difficult for researchers trying to carry out detailed analyses since they do not have access to all necessary raw information needed to draw meaningful conclusions. Capacity building efforts should focus on providing training around best practices with regards both collection storage as well developing methods analysing pre-existing templates to gain back some contextual insight into historical datasets.

Another challenge identified is that there is a lack of satisfactory harmonisation on what to monitor at European level and that the roles and missions of the different initiatives involved in biodiversity monitoring and data interoperability and harmonisation should be clearer.

Finally, it was raised that the fragmentation of the biodiversity monitoring landscape was difficult to handle. Having a one stop shop provided by IT infrastructures, connected with national nodes, creating / asking to apply standards, providing a central register for those standards could help.

Building on the Biodiversa+ survey on data harmonisation and interoperability, the above-listed issues were ranked by priority level (Fig 4). The above issues that are considered as the most important are first that there is not enough capacity building and knowledge sharing and then the governance issues.

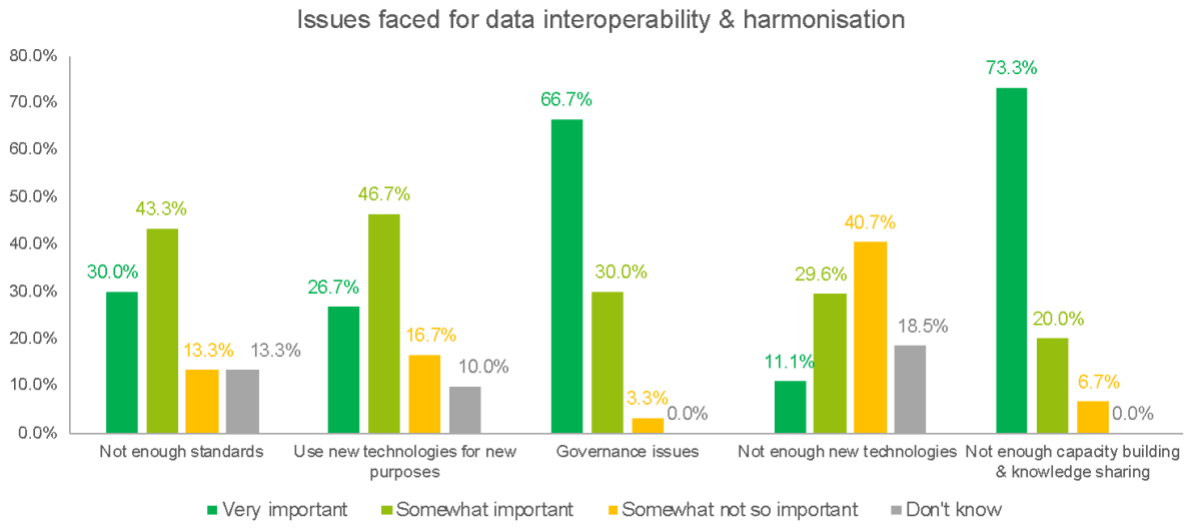



Fig 4: How important are the above data interoperability and harmonisation issues according to the Biodiversa+ survey respondents.

## 3.2 Possible support by Biodiversa+ to tackle data interoperability issues

### 3.2.1 Support to encourage best practices, knowledge sharing and capacity building

A key activity to develop capacities and best practices on data interoperability and harmonisation would be to develop a guide. This guide could:

- Explain how to establish data management plans for biodiversity monitoring initiatives
- Detail what it means to use controlled vocabularies from different sub-disciplines in different technical metadata standards.
- Document the key existing standards and data types and promote data interoperability
- Include a map to visualise how different tools for data interoperability work together

Such a guide would have to be updated over time to document the evolving landscape of biodiversity monitoring data interoperability and should.

Handbooks could be another concrete way to guide data management in a consistent manner at different points in the cycle, such as collecting data, analysing data, and publishing results. In addition, it would be beneficial to develop a standards case to help communicate which standards and vocabularies could be used for data management.

Developing or mapping existence data management tools that have the potential to help standardise data management practices throughout the scientific community. By providing a toolkit that includes data management tools, semantic data, and data validation tools,

Biodiversa+ can help researchers manage their data, as well as share and validate their research findings more easily.

In order to contribute to the harmonisation and interoperability of data across disciplines, Biodiversa+ could work to build capacity within different research communities to manage data effectively and efficiently. This can be done through a number of strategies, including:

1.Including the necessity of calculating a budget for data management into proposals for research calls as well as R&I Projects from the side of the funder. Otherwise, such efforts are often left out of budgets and plans. In the Biodiversa+ research calls, the funders could explore including a dedicated budget for data management into the research proposals[5].

2.Creating a document of strategies of different funders in Biodiversa+ how they deal with costs for data management.[6]

3.Providing training and educational opportunities on data management practices to researchers. Create online courses/instructions on standards and methodologies for different stakeholders and also courses for lecturers to enable them to teach standards to their students.

4.Working with other data-sharing organisations to create a comprehensive data sharing platform.

5.Developing software tools to make data management easier.

6.Working with data Stewardship initiatives to improve data management practices.

7. Explore set up of expert teams to provide hands-on support to data collectors, or facilitate interaction with such teams

8. Raise awareness and understanding of existing tools and standards. Several options to raise awareness could be explored: eg. Biodiversa+ could communicate on GBIF to encourage data sharing, and communicate on the guides that may be created.

### 3.2.2 Governance

A key role for Biodiversa+ in the European landscape, in collaboration with other key initiatives, would be to facilitate the integration of biodiversity monitoring data across sectors (e.g., agriculture, forest, nature, water) to create co-benefits and improve cost effectiveness, with specific focus on key priority needs. Additionally, as raw data are important for reporting towards the EU directives but are often not shared (except for species occurrence data), Biodiversa+ partners could share their raw data. Yet, it would help to first have common

[5] It was raised by the German funder DFG during the 2nd Biodiversa+ workshop that DFG is aware of data management costs:
https://www.dfg.de/en/research_funding/principles_dfg_funding/research_data/index.html
[6] Such activities are of relevance for the Biodiversa+ WP1.

standards and have a common platform to gather the data. In addition, a FAQ on questions and needs stated in EU global environmental policies could be developed.

Biodiversa+ could also help to identify the inter-linkages between the different actors to help move towards EU/global biodiversity monitoring networks. Biodiversa+ could play a role in coordinating the interactions between existing national initiatives, and in helping to navigate the scientific and biodiversity monitoring communities in this complex landscape. For example, Biodiversa+ could play a role in encouraging cross-working at national level among nodes from different multi-country networks and could facilitate coordination between stakeholders to agree on specific standards or workflows. Lastly, with the support from Biodiversa+, it was encouraged to have a European one-stop shop portal for biodiversity data which would be "close to home" through national nodes.

### 3.2.3 (Meta)data standards

To overcome the lack of satisfactory (meta)data standards and the proliferation of standards, several ways forward were identified.

Biodiversa+ can contribute to standard needs by providing guidance on how to harmonise data across different research landscapes and can also help develop standards for "results" of monitoring schemes and establish cooperation with local and European initiatives that offer services aimed to reach FAIR data. Furthermore, developing guides for data harmonisation of different data types and metadata standards can help standardise data across different research initiatives. Promoting standardisation "programme" metadata (i.e. data / dataset metadata / programme metadata) can help to improve the usability of data across different research projects.

Biodiversa+ could also contribute: i.) to approach a consensus on strength and weak points of technical standard fitting for purpose in biodiversity data and what extensions to these standards mean; ii.) to establish common best practices and guidelines (or building blocks) for data management plans of upcoming monitoring initiatives. Additionally, specific funding and support for the development of data standards (TDWG, RDA...) could be facilitated. The Biodiversa+ Call for proposals could also aim to capitalise on increasing the profile of new technologies (e.g. through Artificial Intelligence). Fostering data interoperability and semantic meaning to allow interdisciplinary use of data was encouraged. For this, in a training course, support to incorporate expertise on semantics and ontologies was identified as a possible activity.

Finally, to overcome the lack of satisfactory (meta)data standards, identifying minimum metadata sets of information that can be translated in all the major existing metadata schemes would be of high added-value. It would also be vital to store metadata evidences for eg. store records of field observations and use citizen science to document knowledge of the species and its distribution. This information can be used to inform conservation and management decisions and to help understand the species ecology.

**Knowledge sharing and capacity building**

- Sharing guidelines and best practices to tackle data interoperability and tool for calculating budget for data management
- Working across disciplines to build capacity within different research communities
- Helping create and mobilising an expert team to data harmonisation
- Mapping existing tools

**Governance**

- Easing the coordinating interactions between national level nodes and multi-country networks
- Collaboration between key European and national initiatives

**(Meta)Data Standards**

- Providing guidance on how to harmonise data for: data; dataset metadata; programme metadata
- Supporting the use of standards for monitoring protocols, vocabularies and results
- Funding programmes for data management plans
- Incorporate expertise and trainings

Fig 5: Summary of the possible support from Biodiversa+ to tackle data interoperability challenges

# Conclusions

Through the networking and workshop activities developed as part of its activities on biodiversity monitoring dataset harmonisation and data interoperability, Biodiversa+ has achieved the objectives below towards a reinforcement of biodiversity monitoring data and datasets harmonisation and interoperability in Europe:

- Analyse the state of the art of current development on biodiversity data harmonisation and interoperability and global, European, national and sub-national level and contribute to mapping the main actors involved and their complementarities;
- Contribute to support the collaboration of Biodiversa+ with strategic project initiatives, as EuropaBON, European Research Infrastructures, as eLTER-RI and LifeWatch ERIC, and global International Initiatives, as GBIF and Catalogue of Life on the theme dealing with biodiversity monitoring data and datasets harmonisation and interoperability in Europe;
- Contribute on this theme to support the collaboration between these European and Global actors with the national and sub-national Biodiversa+ partners directly involved in running the biodiversity monitoring programmes;
- Highlight shared needs of the different organisations which contributed to the Biodiversa+ work on data interoperability and harmonisation. The input provided represents valuable recommendations for the future development of the Biodiversa+ activities on biodiversity monitoring data and datasets harmonisation and interoperability in Europe. Such needs can be summarised as follow
    - Supporting European scale coordination mechanisms (e.g., projects, Institutions, Initiatives, ERIs) to boost the adoption of common standards and vocabularies ensuring the FAIRification of biodiversity data for common understanding, sharing and re-use;
    - Extending use and the alignment of semantic approaches and tools developing shared roadmaps for the harmonisation and integration of biodiversity data
    - Assessing biodiversity data workflows, from raw observation to indicators and knowledge: completing the mapping of National and EU databases and defining data architecture plans optimal solutions.
    - Developing capacity building actions supporting key stakeholders with best practices and training to optimise the use of the semantic interoperability (FAIR) for biodiversity data

# References

Aubry, S., Frison, C., Medaglia, J. C., Frison, E., Jaspars, M., Rabone, M., van Zimmeren, E. (2022). Bringing access and benefit sharing into the digital age. *PLANTS, PEOPLE, PLANET, 4*(1), 5-12. doi: https://doi.org/10.1002/ppp3.10186

Access to Biological Collections Data Task Group (2005) Access to Biological Collection Data (ABCD). Biodiversity Information Standards (TDWG), http://www.tdwg.org/standards/115

Darwin Core Task Group (2009) Darwin Core. Biodiversity Information Standards (TDWG) http://www.tdwg.org/standards/450

European Commission, Directorate-General for Research and Innovation, Corcho O., Eriksson M., Kurowski K. (2021) EOSC interoperability framework: report from the EOSC Executive Board Working Groups FAIR and Architecture. Publications Office. Available at: https://data.europa.eu/doi/10.2777/620649

Fegraus E.H., Andelman S., Jones M.B., Schildhauer M. (2005) Maximising the Value of Ecological Data with Structured Metadata: An Introduction to Ecological Metadata Language (EML) and Principles for Metadata Creation. The Bulletin of the Ecological Society of America, 86: 158-168. https://doi.org/10.1890/0012-9623(2005)86[158:MTVOED]2.0.CO;2

Hebert, P. D., DeWaard, J. R., Zakharov, E. V., Prosser, S. W., Sones, J. E., McKeown, J. T., La Salle, J. J. P. o. (2013). A DNA 'Barcode Blitz': Rapid digitization and sequencing of a natural history collection. *8*(7), e68535. https://doi.org/10.1371/journal.pone.0068535

Hardisty A.R., Michener W.K., Agosti D., Alonso García E., Bastin L., Belbin L., Bowser A., Buttigieg P.L., Canhos D.A.L., Egloff W., et al. (2019) The Bari Manifesto: An interoperability framework for essential biodiversity variables. Ecological Informatics. Vol. 49. pp 22-31. https://doi.org/10.1016/j.ecoinf.2018.11.003

Heiler S. (1995) Semantic interoperability. ACM Computing Survey, Vol. 27, No 2, pp. 271-273. https://doi.org/10.1145/210376.210392

Holetschek J., Dröge G., Güntsch A., Berendsohn W.G. (2012) The ABCD of primary biodiversity data access. Plant Biosyst. An Int. J. Deal. with all Asp. Plant Biol., 146, pp. 771-779, https://doi.org/10.1080/11263504.2012.740085

ISO (2019) International Organisation for Standardisation, https://www.iso.org/standard/53798.html

Jones M.B., O'Brien M., Mecum B., Boettiger C., Schildhauer M., Maier M., Whiteaker T., Earl S., Chong S. (2019) Ecological Metadata Language version 2.2.0. KNB Data Repository. https://doi.org/10.5063/F11834T2

Kapcar, A. (2011). *Tally sticks of the Stone Age*. https://doi.org/10.1111/j.1949-8594.1909.tb01371.x

RDA-CODATA Legal Interoperability Interest Group (2016) Legal Interoperability of Research Data: Principles and Implementation guidelines, https://doi.org/10.5281/ZENODO.162241

Stocker M., Paasonen P., Fiebig M., Zaidan M.A., Hardisty A. (2018). Curating Scientific Information in Knowledge Infrastructures. Data Science Journal, 17, p. 21, https://doi.org/10.5334/dsj-2018-021

Scholz, A. H., Freitag, J., Lyal, C. H. C., Sara, R., Cepeda, M. L., Cancio, I.,. Overmann, J. (2022). Multilateral benefit-sharing from digital sequence information will support both science and biodiversity conservation. *Nature Communications, 13*(1), 1086. https://doi:10.1038/s41467-022-28594-0

Underwood, E., Taylor, K., & Tucker, G. (2018). The use of biodiversity data in spatial planning and impact assessment in Europe. Research Ideas and Outcomes, 4, http://doi:10.3897/rio.4.e28045

von Wettberg, E., & Khoury, C. K. (2022). Biodiversity data: The importance of access and the challenges regarding benefit sharing. PLANTS, PEOPLE, PLANET, 4(1), 2-4. doi: https://doi.org/10.1002/ppp3.10241

Wieczorek J., Bloom D., Guralnick R., Blum S., Döring M., Giovanni R., Robertson T., Vieglais D. (2012) Darwin Core: an Evolving Community-developed Biodiversity Data Standard. PLoS One, 7, p. e29715, https://doi.org/10.1371/journal.pone.0029715

Wilkinson, M., Dumontier, M., Aalbersberg, I. *et al.* (2016).The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* 3, 160018 https://doi.org/10.1038/sdata.2016.18

# Annex: Mapping of biodiversity monitoring databases & initiatives

The descriptions of the initiatives below are extracted from the Biodiversa+ survey on data interoperability and harmonisation launched in October – November 2022. The respondents had an opportunity to describe the main objectives of their databases, its data architecture and workflows as well as what they identified as best practices and challenges that they faced.
Based on these answers, it is interesting to highlight that 7 of the databases identified are then uploaded their data on GBIF, yet several initiatives also acknowledged that their data or access to their data is not fully open.
For the national and sub-national databases, it can be flagged that several of them mentioned that their data are used:

- For national, European or other international requirements: 8
- For research purposes: 6
- To advice decision making: 5
- To conserve or protect biodiversity: 4
- To produce indicators: 4

Other uses of biodiversity data were also mentioned, including for example that data are used for ecosystem accounting and to create methods to map ecosystems.
When it comes to the recurring best practices, the respondents mentioned that it was very positive to have clear metadata descriptions, for all datasets and that it was positive to publish data on GBIF through the Darwin Core.
Recurring challenges faced by the databases are the lack of interoperability or compatibility between their databases and other databases, the difficulty to clean data which is time and human resources consuming as well as that having different monitoring methods or practices on the ground makes it hard to compare data. Changes need to occur in the databases, this is also a challenge that was identified.

## Global and European databases

### EuropaBON biodiversity monitoring database

| | |
|---|---|
| Name of initiative | EuropaBON monitoring database |
| Brief presentation of the initiative and its main goals | The EuropaBON monitoring database is a web-based platform to collect/record the current biodiversity data workflows across Europe. This database is aimed to understand how biodiversity data collected in monitoring schemes across Europe flows through different institutions and programs and gets processed to produce Essential Biodiversity Variables (EBVs) and Ecosystem Services Indicators (ESS) other EU policy-relevant indicators. From here on, we use the term 'integration initiatives to refer to each one of these full biodiversity data workflows. The main goal of the database is to report integration initiatives at the European level, but also at other scales (e.g. National), particularly if there is a potential to produce a policy-relevant EBV or indicator, or at least the potential to be integrated into other European-level integration initiatives. |
| Scale of the initiative and of their dataset | European |

| Name of initiative | EuropaBON monitoring database |
|---|---|
| Overview of data architecture and workflows of the database | The main elements of the database are the nodes of the monitoring networks, named "'integration initiatives": these are initiatives/ projects/ programmes that integrate biodiversity data from monitoring schemes and process them to produce aggregated biodiversity data or Essential Biodiversity Variables (EBVs), Essential Ecosystem Services Variables (EESVs) and/or other European policy-relevant indicators. Each integration initiative is composed of three distinctive elements:<br>1. Integration Nodes: institutions/projects/platforms integrating/processing biodiversity data to generate EBVs, EESVs or any other indicators with potential relevance to environmental policy, particularly at the European level. One single integration initiative can be composed of many nodes operating at different scales (e.g., subnational, national, European-level) in a coordinated manner. One single institution can act as an integration node of different integration initiatives.<br>2. Biodiversity Data: biodiversity monitoring initiatives/schemes responsible for the collection of biodiversity-related information, including data collected from the field as well as remote sensing data.<br>3. Datastreams: these represent data flows and connect IntegrationNodes across different scales. These contain two types of information: information related to the data flows between nodes (dataset) and whether this has been integrated/processed ("products": EBVs, EESVs, indicators) or not (raw and aggregated data) (data process). If the data has been integrated/processed, it also contains information about the integration method (e.g., statistical modelling, expert opinion). A data stream that does not connect with another integration node represents a final Product (EBVs, EESVs, Indicator) that can eventually be up taken by policy or not.<br>The database has been developed using technologies licensed as free software. A PostgreSQL database has been chosen to store all data. The website has been developed using technologies licensed as free software. The Backend has been programmed using the Python language, taking advantage of the powerful Django web framework, and the Frontend has been developed in HTML, JS, jQuery, and some libraries as Leaflet and d3.js for mapping and data visualisations. The whole we-platform including the database is stored in a VPS located in Frankfurt. The deployed website uses Nginx web server technology. |

| | |
|---|---|
| Name of initiative | EuropaBON monitoring database |
| What is data used for? | The main objective of the EuropaBON monitoring database is to be able to explicitly describe current workflows of monitoring efforts delivering biodiversity information in Europe. Monitoring efforts are understood in this deliverable as projects designed to track biodiversity change in time in a spatially explicit context. To do this, and in contrast to previous efforts, the present database has developed a novel network framework allowing to map the flow that data experiences since it is collected on the field until it is used in different aggregation steps across different spatial and temporal scales.<br><br>To describe policy-relevant monitoring initiatives and data workflows, the database is not aimed to include all existing national and subnational monitoring projects, but to be comprehensive in including all representative networks that, at European scale, coordinate and integrate biodiversity data and allow obtaining representative biodiversity estimates of change. By, building on the current approach, the database proposed by the EuropaBON project would be able to be expand in the future and gradually grow by including biodiversity networks at progressively lower spatial scales (i.e., national and subnational) and eventually contribute to a more detailed assessment of the monitoring panorama in Europe. |
| Data access and compatibility with the FAIR principles | Some changes have to be made to follow all the FAIR principles |
| Best practices of the initiative to ensure interoperability of their datasets | This database is publicly available to everybody through a website. Everybody can have access and register new information.<br><br>Visualisation, mapping and query modules are useful to navigate through the database. |
| What are the main challenges faced related to interoperability of datasets? | It has to implement an API to allow external automated access to the database. The database will be published in a structured JSON format. There's already a similar output existing in the website database for coordinators using CSV files. It will be important to publish last updated timestamps in every data structure to help external integrations updates.<br><br>It has to be discussed which fields can be publicly accessible. To improve interoperability, it will be also important to better study available data standards, such as PPSR Core. Most of the EuropaBON database fields should be mapped to the closest (and simplest if possible) standard to our database. |
| The website of the initiative, contact details | EuropaBON website: https://europabon.org<br><br>EuropaBON monitoring database: https://monitoring.europabon.org/monitoring<br><br>Contact person: David Martí Pino<br><br>Email: d.marti@creaf.uab.cat |

## Essential Biodiversity Variables Data Portal

| | |
|---|---|
| Name of initiative | Essential Biodiversity Variables Data Portal |
| Brief presentation of the initiative and its main goals | A data and metadata standard and an infrastructure for the mobilisation and discovery of EBV cubes. These are spatiotemporal biodiversity monitoring information products |
| Scale of the initiative and of their dataset | International |
| Overview of the data architecture and workflows of their database(s) | EBV data cubes |
| What are data used for? | Data are used for spatio-temporal assessment of biodiversity change. |
| Data access and compatibility with the FAIR principles | The platform is in open access and follows FAIR principles. It also follows GEO Data Management Principles |
| Best practices of the initiative to ensure interoperability of their datasets | The standard defines a common format to organise, document and seamless analysing EBV data cubes |
| Main challenges faced related to interoperability of datasets | - Documentation of data provenance.<br>- Metadata completeness |
| Website | Website: https://portal.geobon.org/home |

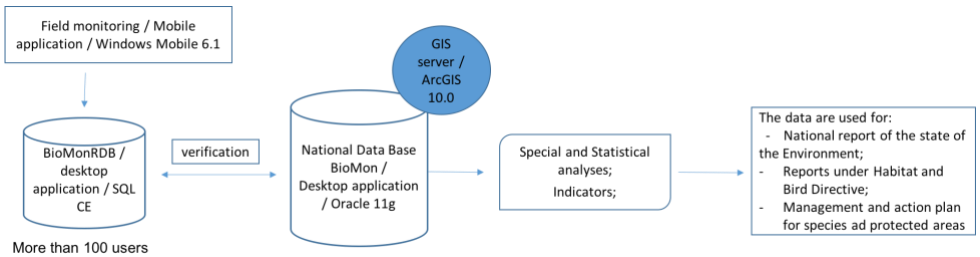
# (Sub-)National databases and initiatives

## Belgium - Flanders

| | |
|---|---|
| Name of initiative | Meetnetten.be |
| Presentation of the initiative and its main goals | Meetnetten.be is a suite, operationalised since 2018, of monitoring networks through which the Flemish Government is collecting high-quality information on 78 priority plant and animal species. These are species on which Flanders has to report to Europe in the framework of the Habitats and Birds Directives, but also other species that are important for the Flemish nature policy. |
| Scale of the initiative and of the dataset | Sub-national |
| Overview of the data architecture and workflows of their database(s) | Meetnetten.be collects data from Citizen scientists and from biodiversity monitoring network coordinators who define which species should be monitored and with which protocols and identify volunteers to perform monitoring activities.<br>All the collected data then go to a large database and one of the key aims of Meetnetten is to then share these data with the researchers and that data get published. To help the scientists find the data that they need, a start scheme was created allowing to connect all the institutes and data collected by volunteers through Meetnetten.be via an ODBC access. The results are also disseminated to the Agency for Nature and Forests by a map service in high resolution. The data are also all published through GBIF in the Darwin core standard but with a lower resolution. Meetnetten.be is indeed not able to publish all the data in high resolution to GBIF. |
| What are data used for? | • To report and to protect nature.<br>• For research |
| Data access and compatibility with the FAIR principles | Yes |
| Best practices to ensure interoperability of their datasets | • To publish data to GBIF in DwC |
| Main challenges faced related to interoperability of datasets | • Different methods of monitoring, making the data not always 100% comparable |

| | |
|---|---|
| Name of initiative | Meetnetten.be |
| Website of the initiative, contact details | Websites:<br>• www.meetnetten.be (in Dutch)<br>• https://www.gbif.org/dataset/search?q=meetnetten and https://www.gbif.org/occurrence/search?collection_code=meetnetten (on GBIF) |

## Bulgaria

| | |
|---|---|
| Name of initiative | BioMon – Information system at the National Biodiversity Monitoring System |
| Brief presentation of the initiative and its main goals | The Biodiversity Monitoring Information System is a distributed system composed of a Central National Database (BioMon), which is administered by ExEA and regional databases (BioMonRDB) – installed in different users responsible for monitoring biodiversity in Bulgaria. Primary data can be collected by mobile application – BioMon Mobile working on Windows mobile 6.1 operating system.<br>The information system has the following main functionalities: creating electronic forms, creating users, creating monitoring sites, which, depending on monitoring methodologies are - points, routes or sampling sites (plots); data validation process; exports in different formats – .shp, excel tables, printing and doc.<br>Users of the Information System are experts from regional environmental inspections, experts from national and natural parks, scientists and volunteers.<br>This system is being updated. |
| Scale of the initiative and of the dataset | National |
| Overview of the data architecture and workflows of their database(s) | The main datasets in the Information system are species and habitats from National and European interests from almost all main biological groups (fungi, mosses, vascular plants, invertebrates, fish, amphibians and reptiles, birds, mammals, invasive alien species):<br>Architecture of the current Information system – three components<br>1. Central database (BioMon) – desktop application located in ExEA server. This is a database based on Oracle 11g.<br>2. Regional databases (BiomonRDB) – open-source database - SQL CE database<br>3. Mobile application (BioMon mobile) for collecting primary data from field monitoring observations. It works on Windows Mobile 6.1.<br>  <br><br>Fig 6: Data architecture and workflow in the current BioMon database: |

| | |
|---|---|
| Name of initiative | BioMon – Information system at the National Biodiversity Monitoring System |
| What are data used for? | - Primary data from field monitoring are used for analyses for the state of species and habitats on local (mostly geographical area – mountain or river for example) and national level.<br>- Production of different indicators and reports which are part of National Report of the state of the Environment in Bulgaria.<br>- Prepare national reports under Habitat and Birds directive.<br>- Research: students and scientists used the datasets for publications and theses.<br>More information on:<br>● Brown bears (Ursus arctos)<br>● Wintering birds<br>● Chamois (Rupicapra rupicapra balcanica) |
| Data access and compatibility with the FAIR principles | Primary data are not open but, as a governmental institution, BioMon will provide them when they are asked. |
| Best practices of the initiative to ensure interoperability of their datasets | ● Primary data can be exported in different data formats for any kind of analyses - .shp file for special analyses and .xml.excel file for statistical analyses. |
| Main challenges related to interoperability of datasets | ● The current system is not interoperable<br>● There is ongoing work on future data information systems to have tools and data formats as export in csv file format and we want to integrate Darwin core as a standard to store and share our data with external databases or Information systems. |
| The website of the initiative, contact details | 1. The main concept of Bulgarian Biodiversity Monitoring system, species and monitoring sites - https://eea.government.bg/bg/bio/nsmbr/osnoven-dokument-na-nsmbr<br><br>2. Field guide and field protocols - https://eea.government.bg/bg/bio/nsmbr/prakticheskо-rakovodstvo-metodiki-za-monitoring-i-otsenka<br><br>3. Information system for monitoring of Biodiversity - https://eea.government.bg/bg/bio/nsmbr/inf-system |

## Czech Republic

| | |
|---|---|
| Name of initiative | National Biodiversity Portal of Czech Republic |
| Brief presentation of the initiative and its main goals | The system of Nature Directives Monitoring in the Czech Republic is developed to fulfill the obligation of Nature Directives related to surveillance and reporting, and therefore encompasses a wide spectrum of activities, namely: habitat mapping (field survey entire state territory) and species monitoring programmes. |
| Scale of initiative and dataset | European |
| Overview of the data architecture and workflows of their database(s) | The National Biodiversity Portal, encompasses Species Occurrence Database (universal species biodiversity system) and Habitat Mapping Database (results of continuous habitat mapping). The Species Occurrence Database centralises available data from species monitoring and other surveys, including citizen science projects or independent research activities. Habitat Mapping Database is solely built upon habitat mapping. |
| What are data used for? | • EU obligations, national indicators<br>• Red Lists<br>• Regional and local nature conservation needs (incl conservation areas management plans and praxis or decisions of local authorities. |
| Data access and compatibility with the FAIR principles | Yes. Open data in case of habitat database, registration needed for detailed data access, interpreted data (10km grid) free. |
| Best practices to ensure interoperability of their datasets | • Open data character,<br>• A clear metadata description,<br>• An easy access<br>• Validation procedures. |
| Main challenges faced related to interoperability of datasets | • Human resources<br>• The data cleaning of different imports is a huge burden - but in the process of centralisation, which should be understood as a key part of interoperability, being the main factor of database quality - therefore the key element of database attractivity and importance - it is necessary. |

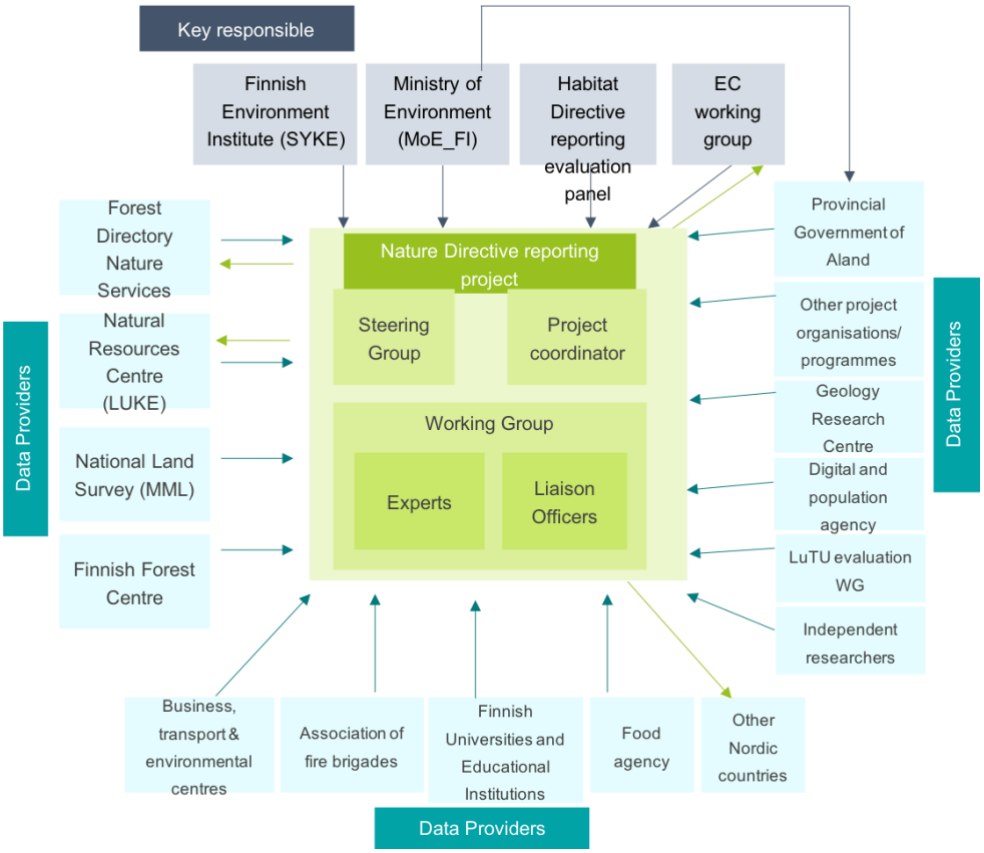
| | |
|---|---|
| Name of initiative | National Biodiversity Portal of Czech Republic |
| The website of the initiative, contact details | Website: https://www.nature.cz/web/en<br>Database access: portal.nature.cz/nd (in Czech only)<br>Contact: karel.chobot(at)nature.cz |

## Estonia

| | |
|---|---|
| Name of initiatives | Environmental Monitoring System (KESE)<br>Estonian Nature Information System (EELIS)<br>PlutoF, Nature Observations Database (LVA) |
| Brief presentation of the initiatives and their main goals | In Estonia environmental data are collected, including wildlife data in KESE. EELIS has been created (upgraded at the moment) and is widely used by the decision makers. PlutoF supports non-governmental databases. |
| Scale of initiatives and datasets | National |
| Overview of the data architecture and workflows of their database(s) | For KESE, environmental data (water, habitat and species) is inserted and held in the system.<br>EELIS is used by the decision makers and planners. Information about protected species, habitats, protected areas, waterbodies, salmonid rivers, nature observations, dams, international areas are inserted, updated and held there by the Environmental Agency which is governmental organisation.<br>The environmental Portal is a website, where users can find various environmental information and access to Estonian Nature Information System (EELIS). Different datasets using analytical tools can be found there – for example hunting data, bird ringing data and soon data of protected areas. This system is still being developed and new data will be available for the public on an ongoing basis.<br>In PlutoF the users can manage their full data lifecycle: from data management plan to the publishing and archiving their datasets in machine readable format.<br>Finally, the LVA non-governmental supporting database provides the possibility to forward information about protected species to the EELIS, after the Environmental Board specialist has checked the information. |
| What are data used for? | • Data are used for decision making at local and national level, for example giving permits (developing, mining, etc), planning processes, monitoring, analyses, preparation of the rules of the protected areas, use of water, assessing the impact on natural values, international reporting, etc.<br>• Data are also used to create the methods to map and assess ecosystem services and carry out analytical studies. |
| Data access and compatibility with the FAIR principles | Most data are compatible with the FAIR principles, yet this is still work in progress. |
| Best practices to ensure interoperability of datasets | Ringing birds: if a ringed bird is seen in Estonia, it will be registered in the database (KESE) and the information will also go to the person who saw the bird and as well to the person who ringed the bird. But methods for wildlife monitoring are mostly created taking into account the local conditions, available experts and budget. |

| | |
|---|---|
| Name of initiatives | Environmental Monitoring System (KESE)<br>Estonian Nature Information System (EELIS)<br>PlutoF, Nature Observations Database (LVA) |
| Main challenges faced related to interoperability of datasets | • Changing existing methods and systems is complex. Estonia has just finished the development of KESE and EELIS is being upgraded at the moment. Such changes, in addition to being expensive, are time and energy consuming work.<br>• Compatibility between the different databases should and could be improved. |
| The website of the initiative, contact details | Websites:<br>• KESE portal: https://keskkonnaportaal.ee/ (in Estonian)<br>• KESE registration: https://register.keskkonnaportaal.ee/register (in Estonian)<br>• KESE: https://kese.envir.ee/kese/welcome.action (in Estonian)<br>• PlutoF: https://lva.keskkonnainfo.ee/default.aspx?state=1;877954539;est;lvadb;;&lang=eng<br>• PlutoF: https://plutof.ut.ee/ |

## Finland

| | |
|---|---|
| Name of initiative | Finnish Ecosystem Observatory (FEO) |
| Brief presentation of the initiative and its main goals | FEO provides a national cooperation model for biodiversity and ecosystem monitoring, and to improve interoperability of the ecosystem data. |
| Scale of initiative and dataset | National |
| Overview of the data architecture and workflows of their database(s) | Illustrative material is not yet available for the entire FEO. Yet below is available a graph illustrating how the Habitat Directive reporting is working in Finland.   Fig 7: Workflows for habitat reporting in Finland |
| What are data used for? | So far, FEO is in a building phase. FEO uses cases for ecosystem data cover:<br>• assessment of threatened habitats and directive reporting<br>• carbon-neutral land use<br>• biodiversity & ecosystem data for municipalities,<br>• indicator development,<br>• ecosystem accounting. |

| | |
|---|---|
| Name of initiative | Finnish Ecosystem Observatory (FEO) |
| Data access and compatibility with the FAIR principles | It depends on the data owner; most of the data collected by SYKE are FAIR. |
| Best practices of the initiative to ensure interoperability of their datasets | • Metadata descriptions were developed and improved. |
| Main challenges faced related to interoperability of datasets | • Poor metadata descriptions of some data providers remains still a challenge<br>• un-FAIR policies of data owners |
| Website | Website: www.feofinland.fi |

## France

- EBVOSC

| | |
|---|---|
| Name of initiative | EBVOSC |
| Brief presentation of the initiative and its main goals | The aim of EBOVSC is to operationalise Essential Biodiversity Variables (EBV) and associated biodiversity indicators by targeting the highest levels of FAIRness (Findable, Accessible, Interoperability, Reusable) for both data and source code implementation, so that data and tools can be widely shared and reused. EBVOSC aims to demonstrate that a better mobilisation of such data can readily generate EBVs and associated biodiversity indicators through automated and regular updating.<br>EBOVSC has several objectives:<br>● Provide an open and transparent comprehensive EBV operationalisation pilot to tackle IT challenges such as data structuration and sharing, source code review standardisation and dissemination that delays the ability to respond quickly to face current biodiversity and climate emergencies. Main target: biodiversity scientific community.<br>● Build on existing international standards, approaches and initiatives regarding data and workflows, thus benefiting communities in life, climate, and earth sciences as well as the humanities community, by linking biodiversity indicators to socio-economic measurements. Main target: broad research community.<br>● Operationalise the EBV concept in a FAIR and transparent way to increase people's awareness on the biodiversity and climate crises through trusted indicators. Main target: society and stakeholders. |
| Scale of initiative and dataset | International |
| Overview of the data architecture and workflows of their database(s) | The French Biodiversity e-infrastructure "PNDB" is gathering all research data related to biodiversity<br>Dataset example: https://data.pndb.fr/view/doi%3A10.48502%2F8bb5-pk85 and see how EML metadata specifications work to add terminological resources (ontology terms, thesaurus terms, ...) inside the metadata to standardise notably attributes description. |
| What are data used for? | Data are mainly used for research purposes. |
| Data access and compatibility with the FAIR principles | In the EBVOSC data catalogue, they are giving access to datasets under the CC-BY 4.0 licence. This is the only licence accepted in France for research data as open data. |

| | |
|---|---|
| Name of initiative | EBVOSC |
| Best practices of the initiative to ensure interoperability of their datasets | The use of the EML metadata standard is considered as a best practice because it can be translated into other metadata standards such as: ISO19139, DCAT which are less detailed than EML. In addition, data standards such as DwC oblige to describe each data attribute so that it can be used notably thanks to the possibility to link each attribute description to a (or several) terminological resources allowing "machine actions" to convert it "easily"". EML metadata standard allows data to be accessible in the semantic web / linked open data through its expression as JSON-LD objects. |
| What are the main challenges that you face related to interoperability of datasets? | - At the European level EML metadata standards are mainly focussed on ISO19139 standards due to the INSPIRE directive.<br>- In the European Union, DCAT is being promoted yet it is less detailed than the ISO19139 standards and this triggers a loose quality "by default"<br>- When explaining and showing EML interest to organisations like TDWG or European infrastructure, it is really complicated to be heard and understood.<br>- We need to develop tools and services related to the use of existing standards, it is also needed to map these as there is a lack of available funding. |
| The website of the initiative, contact details | Websites:<br>● National biodiversity data centre – homepage: https://www.pndb.fr/ (in French only)<br>● National biodiversity data centre – browser: https://data.pndb.fr/<br>● EBVOSC project: https://www.pndb.fr/fr/activites/projets-techniques-et-scientifiques/ebvosc |

- **Information system on natural heritage (SINP)**

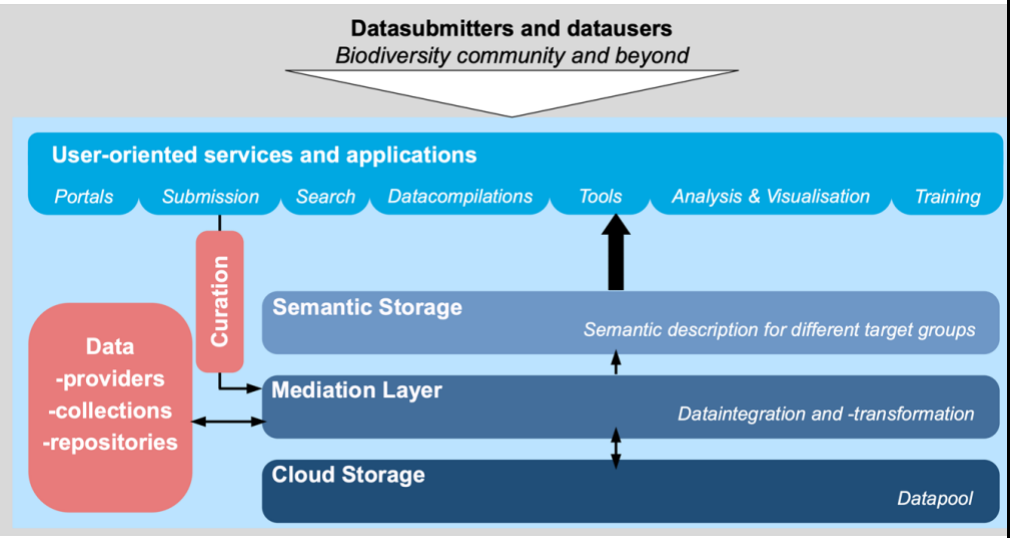
| | |
|---|---|
| Name of initiative | Information system on natural heritage (SINP) |
| Brief presentation of the initiative and its main goals | SINP is a network of actors who share naturalist information they own according to a set of common methods and rules.<br>SINP is a partnership between the French Ministry of Environment, the French Biodiversity Agency, the national natural history museum (MNHN), associations, territorial collectivities, public and private institutes, and decentralised public services. It aims at enhancing a synergy among actors working on the production, management, treatment, valorisation, and diffusion of geolocalised data corresponding to natural heritage (biodiversity and geodiversity).<br>SINP and international GBIF data are exchanged both ways through the French GBIF national node point. |

| | |
|---|---|
| Name of initiative | Information system on natural heritage (SINP) |
| Scale of initiative dataset | National<br>Sub-national |
| Overview of the data architecture and workflows of their database(s) | SINP data are stored in the databases according to the formats defined for the different programmes and themes covered. These data are transmitted by the different partners according to an architecture defined in a concerted way. A simplified overview of the data architecture and workflows is available here: https://inpn.mnhn.fr/programme/donnees-observations-especes/presentation by clicking on “Simplified SINP architecture as regards data exchange”. |
| What are data used for? | The data are used for various public policies: evaluations, monitoring, European reporting, etc. |
| Data access and compatibility with the FAIR principles | ● Data are released under an open license.<br>● Data are sent to GBIF, which complies with FAIR principles. |
| Best practices of the initiative to ensure interoperability of their datasets | ● SINP standard is compatible with Darwin Core (mapping documents and joint working groups between SINP and GBIF France to make both standards aligned).<br>● Data are published to GBIF in DwC (+ EML metadata) |
| Main challenges faced related to interoperability of datasets | ● Version change management<br>● Integration of standards in the various data entry/management tools<br>● Understanding of standards by data producers |
| Website and contact details | General presentation of SINP: https://inpn.mnhn.fr/informations/sinp/presentation (in French)<br>Data portal for species/ OpenObs: https://openobs.mnhn.fr<br>Patrinat GBIF data publisher page: https://www.gbif.org/publisher/1928bdf0-f5d2-11dc-8c12-b8a03c50a862 |

## Germany

- **NFDI4Biodiversity**

| | |
|---|---|
| Name of initiative | NFDI4Biodiversity - Consortium for Biodiversity, Ecology and Environmental Data within the German National Research Data Infrastructure |
| Presentation of the initiative and its main goals | NFDI4Biodiversity is a consortium funded by the German Science Minister Conference as part of the German Research Data Infrastructure NFDI. It is dedicated to mobilising biodiversity and environmental data for collective use.<br>The consortium consists of 50 scientific institutions, museums, natural history societies, state offices, and other institutes and expert groups. The partners pool their scientific and technical expertise to provide a broad service portfolio for handling biodiversity and environmental data. The cooperation is guided by the knowledge that stakeholders in science, politics, nature conservation and landscape management need reliable data to be able to develop better contributions to the conservation of biodiversity.<br>To this end, the consortium partners offer added value to the professional community, specifically:<br>- Access to modern technologies and a comprehensive stock of biodiversity and environmental data<br>- Methods and tools for archiving, publishing, searching and analysing data that are suitable for everyday use and have been tried and tested in practice<br>- An expert forum for the safe and competent handling of data for diverse and responsible use |
| Scale of initiative and dataset | National |
| Overview of the data architecture and workflows of their database(s) | The consortium partners of NFDI4Biodiversity hold individual databases and data-related services. From these, data is mobilised for a cloud-based Research Data Commons infrastructure with storage layer, semantic layer and an application layer with services for the end user wishing to access and analyse data. The architecture will be provided by e-mail.<br>The scope of the initiative is national, e.g. NFDI4Biodiversity is nationally funded and works with national researchers, institutions and initiatives to mobilise their data. The scale and scope of the datasets is NOT necessarily national, there are lots of data from international projects involved. Partners regularly provide data for international networks. All German GBIF nodes are part of the Consortium. |

| | |
|---|---|
| Name of initiative | NFDI4Biodiversity - Consortium for Biodiversity, Ecology and Environmental Data within the German National Research Data Infrastructure |
| |  <br>Fig 8: Data workflows in NFDI4Biodiversity. |
| What are data used for? | • Data are used for research purposes. |
| Data access and compatibility with the FAIR principles | Within the consortium, NFDI4Biodiversity provides several data products, including data portals. The mobilised data for the central Research Data Commons infrastructure are not in any case fully open, but made available according to the FAIR Principles - findable, accessible, interoperable, and reusable. |
| Best practices of the initiative to ensure interoperability of their datasets | • In the NFDI4 Biodiversity consortium, there are negotiations with groups of data providers, which datasets they can make available in a harmonised way. A best practice example is the delivery of data from seven German Natural History Collections to a common data portal (www.gfbio.org). The data pipelines have been published and can be implemented by other collections, who wish to plug into the portal (https://gfbio.biowikifarm.net/wiki/Category:Data_Publishing).<br>• Develop an existing tool to visualise, analyse and transform data, where users can combine data from the portal with their own data (see tutorial: https://www.youtube.com/watch?v=zhM717Mpw2c&t=90s, the tool is currently being reconstructed)<br>• Use of the GFBio terminology service for semantic enrichment of platforms (https://terminologies.gfbio.org/)<br>• Support of stakeholders with the implementation of local data management platforms, such as BEXIS and Diversity Workbench. |

| | |
|---|---|
| Name of initiative | NFDI4Biodiversity - Consortium for Biodiversity, Ecology and Environmental Data within the German National Research Data Infrastructure |
| Main challenges faced related to interoperability of datasets | - Divergent practices at the stage of data collection make it difficult to combine data afterwards. |
| Website and contact details | Website: www.nfdi4biodiversity.org<br>Email: contact@nfdi4biodiversity.org |

- **Flora of Bavaria**

| | |
|---|---|
| Name of initiative | Flora of Bavaria |
| Brief presentation of the initiative and its main goals | The Flora of Bavaria initiative aims to describe Bavaria's flowering plants and ferns, including naturally occurring, newly naturalised, invasive, and also extinct species. The project records and documents the flora of Bavaria over time and space with own taxonomic reference backbone list (14.000 names for c. 4.500 taxa) and currently 16 mio occurrence records<br>Data are published as grid-based and help to understand immigration or loss of plants and the basis of species protection.<br>In this project, volunteer monitoring experts, individual projects, societies and associations are involved. Regional species experts and citizen scientists are called upon to support the project with their own knowledge and observations.<br>Data workflows within the Flora of Bavaria initiative are visualised under<br>Data pipeline on observation data in the context of the Flora of Bavaria at https://wiki.bayernflora.de/web/Datenfluss Flora von Bayern (in German). |
| Scale of initiative and dataset | Sub-national |
| What are data used for? | Data on occurrence of plants in the Flora of Bavaria, broad user spectrum, partly for conservation issues. |
| Data access and compatibility with the FAIR principles | Data from the Flora of Bavaria are open, presented via GBIF, see https://www.gbif.org/dataset/64dabd3c-4f34-4520-b9dd-d227a0bf1582 and https://www.gbif.org/dataset/8ea4250e-0ff0-44f8-812e-bffc3b9ba2a4 |

| | |
|---|---|
| Name of initiative | Flora of Bavaria |
| Best practices of the initiative to ensure interoperability of their datasets | • Data are published as two ABCD structured data packages<br>• In parallel, data are published via own data portal under Botanischer Informationsknoten Bayern (https://daten.bayernflora.de/de/index.php) (with dynamic maps and download functions)<br>• Data pipelines are established for technical interoperability within a cross-border cooperation with the Czech Republic for the flora of the Bohemian Forest (see Novotný et al. 2022, https://doi.org/10.3897/BDJ.10.e87254). |
| Main challenges faced related to interoperability of datasets | It is a long-term project with the known challenges:<br>• Heterogeneous datasets involved<br>• Changing taxon concepts and nomenclature over the time<br>• Changing status concepts and conservation issues over the time (native, non-native)<br>• Changing and diverse geographic granularity over subprojects and time<br>The project design of one of the two published major data packages is at first that of traditional floristic grid-based monitoring projects on a voluntary basis, that means overrepresentation of certain taxa and geographical regions within Bavaria. The second package is the result of three guided monitoring initiatives led by the Bavarian environmental agency each with defined design and objectives. Description of the complex situation is available under Datenbereitstellung – Bayernflora (https://wiki.bayernflora.de/web/Datenbereitstellung) (in German). |
| The website of the initiative, contact details | Website: www.bayernflora.de<br>Email: bayernflora@snsb.de |

## Ireland

| | |
|---|---|
| Name of initiative | National Biodiversity Data Centre |
| Brief presentation of the initiative and its main goals | The National Biodiversity Data Centre works to make biodiversity data and information more freely available in order to better understand and assist the protection of Ireland's biodiversity.<br>The National Biodiversity Data Centre applies state-of-the-art information technology to manage data on Ireland's biodiversity. It also:<br>● Provides expertise to increase our understanding of Ireland's biodiversity.<br>● Provides coordination to encourage greater collaboration between partners.<br>● Communicates the evidence-base to inform decision-making.<br>● Supports partner organisations by offering shared-services and other resources.<br>● Builds capacity by provision of biodiversity training and training resources.<br>● Provides leadership to promote the conservation of biological diversity. |
| Scale of initiative and dataset | National |
| Overview of the data architecture and workflows of their database(s) | The National Biodiversity Data Centre serves as the national hub for the collation, storage, and dissemination of biodiversity data. Biodiversity Maps is a national portal that compiles verified biodiversity data from multiple sources and makes it freely available online. Biodiversity Maps serves as a growing repository of >6 million species occurrence records of over 17,000 species. The data is collated in over 170 datasets, and data from all Open Access datasets are shared with GBIF.<br>  <br><br>Fig 9: Data flows in of the National Biodiversity Data Centre of Ireland |

| | |
|---|---|
| Name of initiative | National Biodiversity Data Centre |
| What are data used for? | The data on Biodiversity Maps is used to<br>1. Inform decision-making - The National Biodiversity Data Centre facilitates and promotes the use of biodiversity data to inform public policy and decision-making through analysis, interpretation and reporting. This includes supporting reporting responsibilities under international conventions and legislation, providing evidence to support conservation management at local and regional levels and continued development of Ireland's National Biodiversity Indicators.<br>2. International collaboration - The National Biodiversity Data Centre serves as Ireland's Node of the Global Biodiversity Information Facility (GBIF). Data are also provided to other international initiatives dealing with Invasive Alien Species, Crop Wild Relatives, European Vegetation Studies, pan-European butterfly research and European distribution atlases.<br>3. Track change - The butterfly and bumblebee monitoring schemes have helped shed light on population changes since 2008, open data contributes to Red List assessments of Ireland's wildlife and data on invasive species helps track and catalogue these species. |
| Data access and compatibility with the FAIR principles | There are a mixture of Open Access and Restricted licenses within the database. |
| Best practices to ensure interoperability of datasets | ● Metadata is provided for all datasets.<br>● All data with an Open Access license can be downloaded as a CSV file from Biodiversity Maps.<br>● Publishing data to GBIF in DwC |
| Main challenges faced related to the interoperability of datasets | ● Dataset licence restrictions – some data are published to Ireland's mapping Portal with Restricted Access. Hence, such data are not shared further<br>● Data types - the National Biodiversity Data Centre is currently only sharing occurrence Data |
| Website and contact details | Website of BiodiversityMaps: https://maps.biodiversityireland.ie/<br>Data Manager: Michelle Judge mjudge@biodiversityireland.ie |

## Italy and the Autonomous Province of Bolzano

- Italian National Biodiversity Network

| | |
|---|---|
| Name of initiative | Italian National Biodiversity Network |
| Presentation of the initiative and its main goals | The Italian National Biodiversity Network provides and manages information on biodiversity related to the national territory through a network system that provides for the continuous population of relevant data in possession of national and regional bodies, including research bodies. |
| Scale initiative and dataset | National |
| Overview of the data architecture and workflows of their database(s) | The Italian National Biodiversity Network is a shared data management system consisting of a central node, which allows you to perform search and management operations on the data, and peripheral nodes (databases that have primary biodiversity data) aimed at guaranteeing consultation and "efficient integration of information on biodiversity, all without the physical transfer of the data, which always reside with the cooperating entities that hold the legal rights.<br>The databases owned by the individual nodes differ in structure (different fields) and architecture (different DBs, type Access, Oracle, Mysql, etc.), but they are able to communicate through the BioCASe Protocol.<br>The Network is able to ensure interoperability with similar international infrastructures (LifeWatch, GBIF, etc.) and with the National GeoPortal, in accordance with the provisions of the INSPIRE Directive (Legislative Decree 32/2010). |
| What are data used for? | The Italian National Biodiversity Network, through the aggregation of the current state of knowledge on biodiversity in Italy aims to improve the dissemination and sharing of biodiversity data, making the data available for pure research, for applied research, for education and for training, and to represent a strategic national tool for informed political decisions, which guarantee a sustainable use of the natural resources of our country. |
| Data access and compatibility with the FAIR principles | Sll dataset with an open creative common licence CC-BY (https://creativecommons.org/licenses/by/4.0/deed.it). They ensure that data are findable, accessible, interoperable and reusable, but it still need to improve our metadata quality and standard to fully respond to FAIR principles |
| Best practices to ensure interoperability of datasets | • Publish dataset following INSPIRE directive standards (on data and metadata)<br>• Publish our data using the OpenAPI standard |

| | |
|---|---|
| Name of initiative | Italian National Biodiversity Network |
| What are the main challenges that you face related to interoperability of datasets | • Lack of standards related to ontologies<br>• Missing of a common framework for capacity building and best practice sharing |
| Website and contact details | Website: https://www.nnb.isprambiente.it/<br>Email contact: Cristian Di Stefano cristian.distefano@isprambiente.it |

- **Autonomous Province of Bolzano**

| | |
|---|---|
| Name of initiative | Biodiversity Monitoring South Tyrol |
| Presentation of the initiative and its main goals | This initiative consists of a systematic data collection database of biodiversity including plants, lichens, butterflies, grasshoppers, invertebrates, birds and bats and abiotic parameters as soil chemicals and land-use management data. A development of limnological data is being planned for future activities.<br>All data are collected according to the Darwin Core Standard and sent to the main offices and public entities involved in spatial planning and conservation at local level or involved with European-level reporting. |
| Scale of initiative and dataset | Sub-national |
| Overview of the data architecture and workflows of their database(s) | The relational database is characterised by abundance data of different terrestrial taxa. This allows us to easily store and retrieve information on monitoring sites and periodic surveys. The architecture is structured in a storage part (SQL), a backend part and a front-end part with the possibility of client or mobile client use.<br>It is developed in non-open source SQL language and is coordinated by computer scientists who manage the information system at provincial level. From a GUI, it is possible to import and export data and to conduct queries. Data is automatically provided on existing databases of public bodies by direct writing, reducing human effort. Some features are currently under development.<br>Data is uploaded to the Museum of Natural Sciences portal http://www.florafauna.it/ where, through a GIS application, the data can be retrieved by the stakeholders or also be viewed, with some limitations, to the public. In addition, the complete raw data is accessible to other entities via the geobrowser (https://geoportale.retecivica.bz.it/). |

| | |
|---|---|
| Name of initiative | Biodiversity Monitoring South Tyrol |
| What are data used for? | • Research<br>• Developing local conservation strategies and spatial planning<br>• Habitats Directives reporting |
| Data access and compatibility with the FAIR principles | Data follows the FAIR principles. Datasets are accessible to all public entities and with limitations to the public |
| Best practices to ensure interoperability of their datasets | • The use of primary tables, and unique relationships between tables was identified as a best practice. Biodiversity Monitoring South Tyrol activates error restrictions via a GUI. Tables relating to the scientific names are processed by the provincial Museum of Natural Sciences and sent continuously updated to our system using common unique identifiers.<br>• Careful preparation of metadata to standardise the vocabulary and make it comprehensible to all users. |
| Main challenges related to the interoperability of datasets | • Homogenise different types of biological data in one single infrastructure (internal problem).<br>• At the level of interoperability between various entities, the biggest problem is related to the different development languages (often not open source) used by the various entities and often complicated to interconnect |
| Website and contact details | An interactive map with some results of the Biodiversity Monitoring South Tyrol work is available here: https://biodiversity.eurac.edu/results/<br>Taxonomic data and all the complete soil parameters measured are available in the database.<br>Contact to request specific data: biodiversity@eurac.edu and chiara.paniccia@eurac.edu. |

## Slovakia

| | |
|---|---|
| Name of initiatives | DataFloS – database to the initiative/long-term project Flora of Slovakia<br>Slovak Vegetation Database (SVD)<br>Non-native plant database (DASS)<br>Database of DNA barcodes of water biota in Slovakia; DNA barcoding of aquatic fauna of Slovakia (SKBAF)<br>Database of DNA barcodes of Elmidae |
| Brief presentation of the initiative and its main goals | **DataFloS**: database of the distribution data on vascular plants in Slovakia, currently under transfer to the structure of the database Pladias-SK, a clone of the well-established Czech database Pladias (https://pladias.cz/). The server with numerous reference indices for Pladias-SK was already installed and currently Slovakia is working on the user interface software that will enable the use of the database by the scientific community and public. DataFloS is linked to GBIF, specifically AlyBase (Database of names, chromosome numbers and ploidy levels of Alysae); CardaBase (Database of names, chromosome numbers and ploidy levels of Cardaminae)<br>**Slovak Vegetation Database**: this is a tool for monitoring habitats for nature conservation purposes. Data (phytosociological relevés; stored in Turbowin v2). GIVD ID: EU-SK-001; connected to European Vegetation Archive and sPLOT. More than 60 000 phytosociological relevés (http://www.ibot.sav.sk/cdf/index.html).<br>**Non-native plant database** (DASS). This database was designed to report the current state of the alien vascular flora of Slovakia.<br>**Database of DNA barcodes of water biota** in Slovakia (www.AquaBOL.sk): Biodiversity studies of water beetles include building a reference database of DNA barcodes (application of NGS sequencing and DNA 11 metabarcoding) of aquatic biota in Slovakia (www.aquabol.sk). Joint to the European Reference Genome Atlas Initiative (ERGA).<br>**Database of DNA barcodes of Elmidae** (www.elmidae.myspecies.info). This portal is dedicated to aquatic riffle beetles of the family Elmidae. The aim of this portal is to build up a comprehensive source of information on this beetle family from various fields like taxonomy, phylogeny, geography etc. This page is very new and still under construction. |
| Scale of initiatives and datasets | International & National |

| | |
|---|---|
| Name of initiatives | DataFloS – database to the initiative/long-term project Flora of Slovakia<br>Slovak Vegetation Database (SVD)<br>Non-native plant database (DASS)<br>Database of DNA barcodes of water biota in Slovakia; DNA barcoding of aquatic fauna of Slovakia (SKBAF)<br>Database of DNA barcodes of Elmidae |
| What are data used for? | • Records of species occurrences<br>• Research in systematics, taxonomy, ecology, phylogeny, evolution<br>• National and international collaboration in synthetic studies covering larger territories<br>• Cooperation with State Nature Conservancy (e.g. monitoring schemes)<br>• Records of occurrences of vegetation units<br>• Vegetation survey<br>• Research in vegetation ecology, ecosystem synanthropisation, remote vegetation survey<br>• Biodiversity and ecology of aquatic ecosystems |
| Websites | DataFloS – database to the initiative/long-term project Flora of Slovakia; AlyBase https://www.alysseae.sav.sk/; CardaBAse: https://www.cardamine.sav.sk/<br>Slovak Vegetation Database (SVD): https://ibot.sav.sk/cdf/<br>Non-native plant database (DASS): http://dass.sav.sk/en/<br>Database of DNA barcodes of water biota in Slovakia: www.AquaBOL.sk; DNA barcoding of aquatic fauna of Slovakia (SKBAF)<br>Database of DNA barcodes of Elmidae: www.elmidae.myspecies.info |

## Sweden

| | |
|---|---|
| Name of initiative | National Inventories of Landscapes in Sweden (NILS) |
| Presentation of the initiative and its main goals | NILS is an environmental programme monitoring biophysical conditions and changes in deciduous forests, grasslands, alpine habitats and seashores for reporting within the EU's Species and Habitats Directive. |
| Scale of initiative and dataset | National |
| Overview of the data architecture and workflows of their database(s) | NILS has several specialised inventories including:<br>● **NILS Grassland inventory** – A National inventory of all types of grasslands, but with an extra emphasis in grasslands with high nature values (started in 2020)<br>● **NILS Deciduous forests inventory** – A National inventory of deciduous forests complementing the data from the Swedish National Forest Inventory (started in 2020)<br>● **NILS Alpine inventory** – A National inventory of habitats in the Swedish alpine area, with a focus on nature types with high nature values (using new design since 2021)<br>● **THUF Sea shore inventory**– A National inventory of habitats along the Swedish seashore, with a focus on habitats included in appendix 1 in the EU Species- and Habitat directive (using new design since 2021) |
| What are data used for? | ● Monitoring biophysical conditions and changes in deciduous forests, grasslands, alpine habitats and seashores<br>● Reporting within the EU's Species and Habitats Directive |
| Data access and compatibility with the FAIR principles | Data are available on an online portal: https://landskap.slu.se/dv/ (in Swedish) |
| Best practices to ensure interoperability of their datasets | ● Requested data that are quality assured and easy to access and understandable (by metadata and the ability to ask someone that knows the data). |
| Main challenges faced related to interoperability of datasets | ● Get users of data to understand the data and its limitations and possibilities. |

| | |
|---|---|
| Name of initiative | National Inventories of Landscapes in Sweden (NILS) |
| Website and contact details | Website: https://landskap.slu.se/dv/<br>Email: Pernilla Christensen, pernilla.christensen@slu.se |