

An interoperability framework for biodiversity monitoring in Europe



Document Information

Grant Agreement number	101052342
Project acronym	Biodiversa+
Project full name	The European Biodiversity Partnership
Biodiversa+ duration	7 years
Biodiversa+ start date:	1 st October 2021
For more information about Biodiversa+	Website: https://www.biodiversa.eu/ Email: contact@biodiversa.eu LinkedIn: Biodiversa+

Deliverable title	D2.18 Report on best solutions for data interoperability and data storage in Europe
Dissemination level	Public
Authors	Yann Le Franc (e-SDF), Benjamin Cogrel (Ontopic)
Contributors	Reviewers: Senem Onen Tarantini (MUR), Martina Pulieri (MUR), Maxime Coupremagne (BBPf-BELSPO), Michele Bresadola (Eurac Research, BOZEN), Petteri Vihervaara (SYKE)
Work package title	WP2 Promote and support transnational biodiversity monitoring
Task or sub-task title	Sub-task 2.2.2 Harmonise databases and data interoperability
Lead partner	MUR
Date of publication	October 2026
Disclaimer	Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.

Cover page illustration: © [Kristaps Ungurs \(Unsplash\)](#)

What is Biodiversa+

The European Biodiversity Partnership, Biodiversa+, supports excellent biodiversity research with impact for policy and society. Connecting science, policy and practice for transformative change, Biodiversa+ is part of the European Biodiversity Strategy for 2030, which aims to put Europe's biodiversity on a path to recovery by 2030. Co-funded by the European Commission, Biodiversa+ brings together partners from research funding, research programming and environmental policy across European and associated countries to work on five 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 high-end knowledge for deploying Nature-based Solutions and valuing biodiversity in the private sector.
4. Ensure efficient science-based support for policymaking and implementation in Europe.
5. Strengthen the relevance and impact of pan-European research on biodiversity in a global context.

More information: <https://www.biodiversa.eu/>

Table of contents

Executive Summary	5
1. Introduction	6
1.1. Problem statement	6
1.2. What is an Interoperability Framework?	7
1.3. Objective of the Framework	8
1.4. What are the key design principles?	8
1.5. How did we design this Interoperability Framework?	9
1.5.1. Understanding the current biodiversity landscape	9
1.5.2. Investigating the existing Interoperability Frameworks	11
2. The biodiversity monitoring Interoperability Framework	15
2.1. Models and Formats	15
2.1.1. Data Models	16
2.1.2. Semantic artefacts: knowledge models	19
2.1.3. Formats	22
2.2. Metadata schemas	23
2.3. Persistent Identifiers	24
2.4. API and Access	26
3. Putting the Framework in Practice	29
3.1. Findability/discoverability	29
3.2. Data integration	35
4. Conclusions	37
5. References	40
6. Appendix	43

Executive Summary

Biodiversity monitoring generates a vast amount of data every year. This data is essential for evidence-based policymaking, for guiding research and for meeting society's needs. Yet much of it remains underused. The organisations that produce it rely on different standards, formats and access methods, which makes it hard to bring data together, compare it and combine it at European scale.

To remove this barrier, the Biodiversa+ partnership proposes an **Interoperability Framework for Biodiversity Monitoring**. The goal is simple: to allow data from many different sources to flow, connect and be easily reused, for the benefit of policymakers and researchers alike.

The framework does not start from scratch. It builds on existing European and international interoperability frameworks, notably the European Open Science Cloud (EOSC) and the Cross-Disciplinary Interoperability Framework (CDIF), and focuses on the implementation of the FAIR Principles, which aim to make data Findable, Accessible, Interoperable and Reusable.

The report sets out 24 recommendations: 16 generic and 8 operational. They follow a **stepwise approach**. The first steps are simple actions that can be taken quickly, such as adopting common ways of describing and sharing data. Later steps introduce more advanced technical solutions that let computer systems automatically exchange and link information.

However, the report stresses that technology alone is not enough. Real, lasting interoperability also requires **shared governance**, **investment in training and skills**, and **long-term funding** for the infrastructures involved. These are areas where institutions and policymakers have direct responsibility.

Implementing this framework would strengthen Europe's ability to draw on all its biodiversity monitoring data to inform policy and track progress towards its environmental targets. It would also help build a global Biodiversity Monitoring Commons as part of the open science ecosystem.

1. Introduction

1.1. Problem statement

Biodiversity and biodiversity monitoring domains are complex fields with many initiatives at different scales, geographical coverages, constraints and more importantly information systems. One of the main aims of Biodiversa+, the European Biodiversity Partnership, is to enable the interoperability (see definition below in section “What is an Interoperability Framework?”) between the different information systems to support accelerated reporting at different scales to provide core information for the establishment of policies at different levels (regional, national and European).

In a previous report from Biodiversa+¹ (Basset et al., 2023), a first analysis of the practices related to interoperability in biodiversity, including analysis of the landscape of biodiversity initiatives and their approach to publish and share their data either in Open Access or under strict control of access for sensitive information, has been conducted.

This report highlighted the heterogeneity of initiatives (national, subnational, global networks, European, international, citizen science, etc.) as well as the heterogeneity of technical approaches to publishing biodiversity monitoring data. It shows the diversity of the technical capabilities by mapping each of the organisations identified in Hardisty et al. (2019) with their functional capabilities i.e. data collection and organisation, data discovery and access, data analysis and/or visualisation and data preservation. They emphasised the lack of satisfactory standards (due to the diversity of existing standards and non-standards approaches) for structuring metadata and sharing data. Although standards such as Ecological Meta Data Language (EML, Jones et al., 2019) and Darwin Core (DwC, Wiczorek et al., 2012) are required for publishing data to the Global Biodiversity Information Facility, (GBIF), these standards are not necessarily used for the publication on their platform and formats vary often which hampers the aggregation of different data sources. The heterogeneity of technical approaches and maturity of the current biodiversity monitoring ecosystem hampers the integration of diverse datasets from different biodiversity monitoring data sources.

This is a key challenge as biodiversity monitoring data must be aggregated and integrated from multiple and heterogeneous sources of data at various scales. Indeed, multiple stakeholders involved in biodiversity monitoring require aggregation and integration of data to support the establishment of biodiversity preservation policies at different scales as well as impact studies for infrastructure and many other societal usages. In addition, this data is valuable to understand the evolution of biodiversity, increase our knowledge of ecosystems and therefore they should be strongly linked with scientific research. Achieving interoperability between different biodiversity monitoring infrastructures is not only a technical necessity but also a crucial element for effective policy implementation. European policies such as EU Biodiversity Strategy for 2030 and the Nature Restoration Law require well structured, harmonised, and spatially explicit biodiversity data to support evidence-based decision-making, target setting, and impact assessment. By facilitating data integration across institutions and countries, the interoperability framework can directly enhance reporting capacities, reduce duplication of efforts, and ensure coherence with legal obligations. Addressing this challenge requires a lot of resources both human and technical.

¹ <https://www.biodiversa.eu/wp-content/uploads/2023/05/D2.2-Report-data-interoperability.pdf>

To reduce this burden, a solution is to adopt a common set of practices and technical approaches that will allow the interconnection of the resources and the creation of services or tools allowing a smooth and reliable exchange of information. Therefore, it is essential to define an interoperability framework that will ease the process of integrating biodiversity monitoring and surveillance data at the European level. To provide an initial framework, the report also investigated interoperability approaches, and particularly the European Open Science Cloud (EOSC) Interoperability Framework (Corcho et al., 2021). This report extends this analysis by integrating other existing interoperability frameworks with the aim to establish a framework by reusing the recommendations that are deemed relevant from the different interoperability frameworks (see section “Investigating the existing Interoperability Frameworks”). First, this report provides a definition for the concepts of interoperability and interoperability framework. It then provides the objectives of the interoperability framework. The following sections provide an overview of our approach to design this interoperability framework, i.e. what are the key design principles and how we designed this interoperability framework. Then, generic recommendations of the Biodiversa+ interoperability framework and then extends this description with more practical recommendations are presented.

1.2. What is an Interoperability Framework?

The aim of this Biodiversa+ report is to provide an interoperability framework for biodiversity monitoring organisations across Europe and the rest of the world to better understand the status and trends of biodiversity. Before defining the concept of interoperability framework, we should first consider **what is interoperability**. The European Commission provides a generic definition of interoperability²:

Interoperability is defined (...) as the ability of disparate and diverse organisations to interact towards mutually beneficial and agreed common goals, involving the sharing of information and knowledge between the organisations, through the business processes they support, by means of the exchange of data between their respective ICT systems.

This definition provided by the European Commission is quite broad as it considers the global level of organisations which includes different aspects such as legal, IPR, ...

A more operational definition, focused on information systems, has been proposed in the context of the Biodiversa+ project (Basset et al., 2023):

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.

An **interoperability framework** is a structured approach that defines the rules, standards, and protocols necessary to ensure that different systems, organisations, or technologies can be interoperable. It **provides a common set of guidelines** that enable interoperability across diverse platforms, ensuring that they can interact without compatibility issues³.

² <https://interoperable-europe.ec.europa.eu/collection/joinup/glossary/term/interoperability>

³ <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:52017DC0134>

1.3. Objective of the Framework

The objectives of this framework in the context of biodiversity are as follows:

- Ease the findability and accessibility of a large diversity of biodiversity monitoring datasets at the European level,
- Ease data integration between biodiversity monitoring datasets from multiple sources to facilitate biodiversity reporting at any scale in Europe.
- Enable the connection of biodiversity monitoring resources with the European biodiversity research community, the European Open Science Cloud and global international communities such as GBIF.

These objectives are summarised in **figure 1** below.

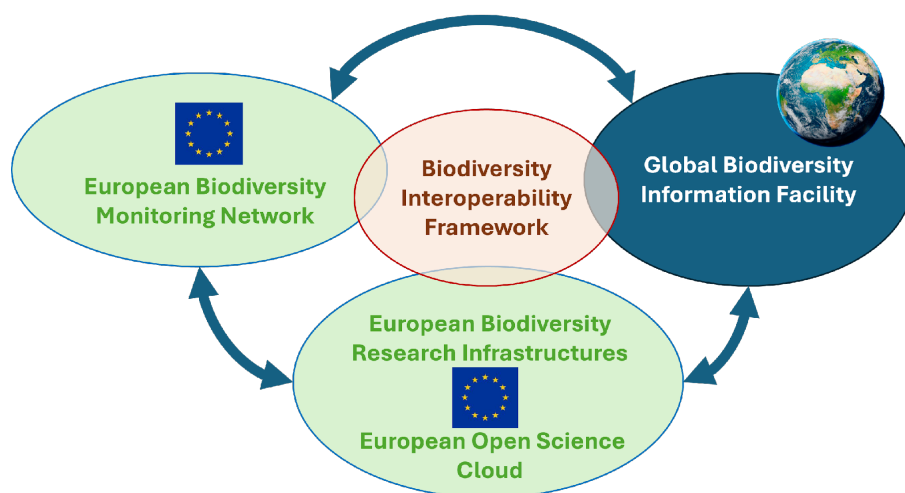


Figure 1. The Interoperability Framework for Biodiversity Monitoring as the glue to connect the European Biodiversity Monitoring Network with European Research Infrastructures and EOSC and global organisations as GBIF.

1.4. What are the key design principles?

To design this interoperability framework, we chose to focus on the following specific design principles or requirements:

- Reuse existing interoperability frameworks.
- Alignment with FAIR Principles.
- Practical Framework to ease the implementation.
- Alignment with the current interoperability approaches used by Research Infrastructures.
- Follow the evolution of standards in biodiversity.

The FAIR principles (Wilkinson et al., 2016) are a set of guidelines to enable data to be Findable, Accessible, Interoperable and Reusable organised in 15 principles. These principles have been designed for both humans and machines. They aim to address the problem of the automatization of data management operations in face of the ongoing “data deluge” (machine actionability).

To summarise, the principles can be grouped around 4 main technical considerations:

1. Models and formats i.e. using standards and semantic artefacts (ontologies, controlled vocabularies, thesauri, etc.) and common formats for (meta)data;
2. Metadata covering different aspects (generic, domain specific, provenance, licence and access rights...);
3. Globally Unique Persistent and Resolvable Identifiers;
4. Open and standardised APIs with access controls.

1.5. How did we design this Interoperability Framework?

In this section, we are describing our approach to develop this framework of interoperability.

Our first step was to investigate the existing interoperability frameworks to identify relevant recommendations. In the meantime, we also investigated the technical landscape both at the level of the European Research Infrastructures and at the level of the biodiversity monitoring and surveillance initiatives listed in D2.2. (Basset et al., 2023).

1.5.1. Understanding the current biodiversity landscape

We organised two expert workshops to collect detailed information about a sample of the existing organisations both biodiversity monitoring and European research infrastructures in biodiversity. These two workshops enabled us to identify gaps and current trends in these communities.

The first workshop aimed at getting an overview of the existing technical systems of European and International Research Infrastructures and their interoperability approaches and to have a better understanding of the existing interoperability frameworks developed in the context of research. The following organisations were presented: SeaDataNet, GBIF, eLTER, Pangaea, GFBio/NFDI4Biodiversity, LifeWatch.

To harmonise the presentation of each research infrastructure, we asked the organisations to answer specific questions on the following topics: Persistent Identifier (PIDs), metadata, semantic artefacts and mappings, data formats, access and API, services supporting interoperability.

The presentations revealed a diversity of approaches to achieve interoperability and the implementation of the FAIR principles. Despite this diversity of approaches, a general trend among the organisations was the move toward adopting formats such as JSON-LD (JSON for Linked Data) or RDF (Resource Description Framework), and the use of semantic artefacts. All the research infrastructures offer semantic artefacts catalogues, use EML metadata schema in most cases and DCAT (Data Catalog Vocabulary, a W3C standard designed to facilitate interoperability between data catalogues published on the web). Many of the organisations provide Persistent Identifiers such as PID Handle or DOIs. Their APIs are documented using the OpenAPI standard and access control is provided to restrict access to sensitive data.

During the discussion, several points have been raised such as the need for converters and brokering services, the reduced use of semantic artefacts as a key to harmonise metadata descriptions, federated queries, geospatial standards, PIDs policies and more.

The second part of the first workshop was dedicated to presentation of different existing interoperability frameworks: the EOSC Interoperability Framework, CDIF, the FAIR Digital Object “framework” (FDO), the FAIR IMPACT project. This session enabled us to better understand the interoperability framework landscape, described in the next section. The EOSC Interoperability Framework is a high-level framework while CDIF is more practical as it emerges from 11 use-cases from 13 scientific communities, including Biodiversity. The FAIR Digital Object is promising as a model to exchange data between stakeholders. However, there are no consensual specifications to implement Fair Digital Objects (FDOs). Finally, the FAIR IMPACT project presented their ongoing work to map the different interoperability frameworks.

The second workshop was focused on biodiversity monitoring organisations at both subnational and national levels and on the FAIR principles.

In the first part of the workshop, we invited 5 organisations: the Biodiversity Monitoring South Tyrol (Autonomous Province of Bolzano, Italy), the National Biodiversity Portal of Czech Republic (Czech Republic), Patrinat and the Information System on Natural Heritage (SINP, France), KESE (Estonia) and Syke (Finland). We used the same topics as the first expert workshop to harmonise the presentations and collect information. The presentations revealed the heterogeneity of the organisations in terms of size, maturity levels and technical implementation. In many cases, national organisations, which provide richly described datasets, use the DCAT model for structuring their collection metadata and EML for the domain specific metadata. However, we cannot generalise the use of EML and DCAT to the whole biodiversity monitoring landscape. They publish their data through their own platforms which do not necessarily provide programmatic access through an API. In many cases, the focus of the organisations is dual: 1- aggregate data from a wide variety of data providers and 2- provide complex visualisations for the public, policy makers and researchers. These national organisations also publish their data into the GBIF platform, therefore using DwC-A as standard for data exchange. Some biodiversity monitoring organisations are in line with most of the FAIR principles except for the use of Globally Unique Persistent and Resolvable Identifiers. Only a few organisations provide DOIs for their datasets. These organisations spend a lot of effort in the aggregation of the different data sources either using a well-defined process and semi-automated information exchange based on existing standards.

All the organisations develop their infrastructure and services in silo, without real collaboration with other national initiatives. Furthermore, one of the main issues for the harmonisation is the lack of resources and access to the appropriate skills. Becoming interoperable with the other national biodiversity monitoring organisations would require a substantial investment for many organisations. Therefore, an ideal interoperability framework should have a low barrier of entry. Although the technical work is of importance, one of the main challenges will be to establish interoperability at the organisation level. A first step in this direction could be to organise regular meetings to discuss interoperability and exchange on the technical solutions developed in the context of each organisation.

During the second half of the workshop, we discussed interoperability and FAIR principles. Although the organisations present had some knowledge of the FAIR principles, it appeared that they did not have a clear understanding of the implementation requirements. For instance, the use of semantic artefacts as an enabler of interoperability.

1.5.2. Investigating the existing Interoperability Frameworks

This definition of interoperability, proposed in the context of the European Interoperability Framework (EIF, European Commission, 2017)⁴, spans both the organisational and the technical aspects. The EIF is a set of guidelines and recommendations developed by the European Commission for public administrations in Europe on how to set cross-border and cross-sector interoperable digital solutions based on open standards and specifications.

Several Interoperability Frameworks have been proposed to align information systems in different contexts. The EIF has been designed to allow exchange of information between e-government information systems at European level to support the Open Data movement. Other Interoperability Framework have been developed in the context of the Open Science movement and in particular the development of the European Open Science Cloud (EOSC). The aim of EOSC is to provide a web of FAIR data and Services⁵.

EOSC Interoperability Framework and FAIR Digital Object

From this context, the first Interoperability Framework for EOSC has been proposed by the EOSC Executive Board Working Group (WG) FAIR and Architecture. This framework derives from the European Interoperability Framework (EIF) and provides a set of high-level recommendations to address specific problems hampering interoperability derived from interviews with different scientific communities. The recommendations are covering the four levels of interoperability, defined in the EIF:

- Organisational
- Legal
- Technical
- Semantic

To support the implementation of the recommendations, the framework proposes an architecture diagram with considerations for services that are necessary for interoperability. The core technical concept used to define this architecture is the FAIR Digital Object (FDO). The concept of the FAIR Digital Object has been proposed and defined in the “Turning FAIR into reality”⁶ report (European Commission, 2018) and is considered as “the atomic entity for the FAIR ecosystem” (Corcho et al., 2021). **Figure 2** presents the proposed “layered” structure for FDOs. At the centre of the model lies the Digital Objects which should be any research products (data, code, ...). To be FAIR, these Digital Objects must be represented with standard formats, persistent identifiers, metadata and documentation. This model has been further developed in the context of expert workshops (Wittenburg et al., 2019; De Smedt, 2020) and is now driven by a community effort in the context of the FAIR Digital Object Forum⁷

⁴ <https://data.europa.eu/doi/10.2799/78681>

⁵ <https://digital-strategy.ec.europa.eu/en/policies/open-science-cloud>

⁶ <https://data.europa.eu/doi/10.2777/1524>

⁷ <https://fairdo.org/>

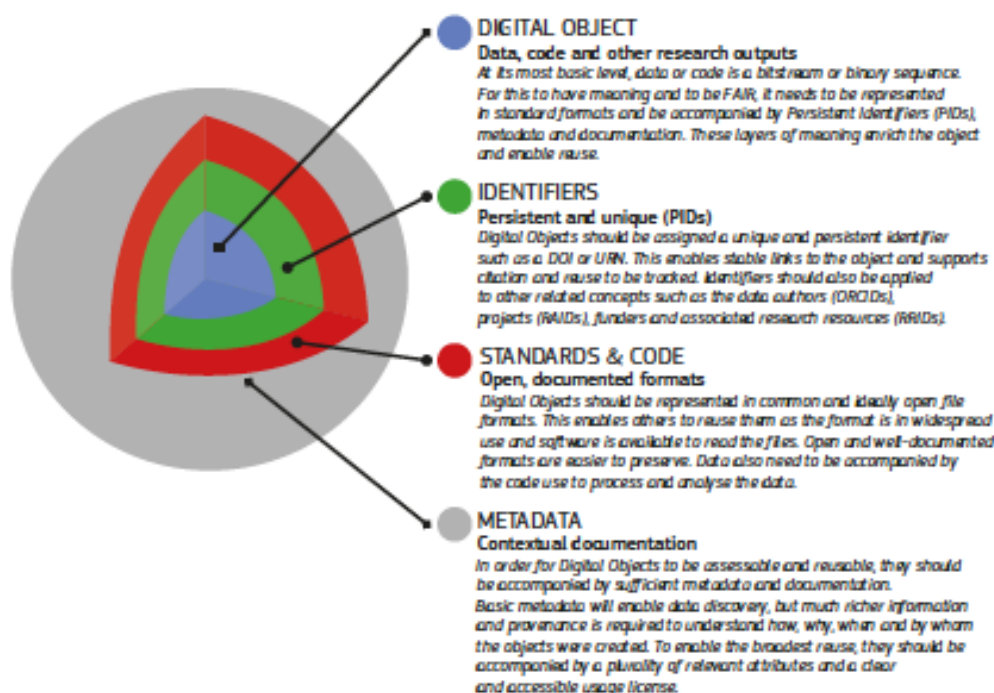


Figure 2. Layered representation of a FAIR Digital Object, extracted from the “Turning FAIR into reality”⁸ report (European Commission, 2018).

Global Open Research Commons (GORC)

Another Interoperability Framework has emerged from the work of Global Open Research Commons Interest Group⁹ from the Research Data Alliance¹⁰. The GORC group considers interoperability in the broader context of “Open Science Commons” or “Data Commons”, which provide a “shared virtual space or platform that provides a marketplace for data and services”. Such “Commons” already exist such as the European Open Science Cloud, the Australian Research Data Commons, ... The development of these international platforms must be coordinated to enable a network of interoperable platforms. A Working Group was spun out of the Interest Group to establish the first International Common Model (GORC-IM, Woodford et al., 2024). This working group also provided an initial definition of a “Commons”, as follows (Jones et al., 2023): “a global trusted ecosystem that provides seamless access to high-quality interoperable research outputs and services.”

In this context, the Interoperability Framework is defined as a global common set of practices to support national and cross-national data exchange and integration. The GORC-IM model is based on the typology defined by the community and outlining the essential elements of a Global Open Research Commons. This typology is summarised in the figure 3 below and covers the technical aspects/elements in blue and dark blue, the human and organisational elements in white, connected through the central element of Interoperability and Standards which is the interface of both the human and technical elements. Each of these elements is subdivided into categories (e.g. Interoperability is split into Technical Interoperability,

⁸ <https://data.europa.eu/doi/10.2777/1524>

⁹ <https://www.rd-alliance.org/groups/global-open-research-commons-ig/activity/>

¹⁰ <https://www.rd-alliance.org/>

Organisational Interoperability and Legal Interoperability) and subcategories (e.g. technical interoperability is subdivided into syntactic interoperability, semantic interoperability and other interoperability). The categories and subcategories have attributes which can be subdivided into features. The model is shared as an excel spreadsheet which contains guidelines associated with the different elements, categories and subcategories¹¹. The GORC model provides a comprehensive set of attributes to be considered when developing a Commons. It provides a relevant framework for establishing a Commons for the biodiversity monitoring community, that would cover both organisational interoperability and technical interoperability. Two elements cover the topics of “interoperability” and “standards and conventions”. The attributes for these two elements overlap with the other investigated interoperability frameworks. It is important to note that this model is not prescriptive and is based on existing good practices.



Figure 3. Essential elements of a Research Commons, extracted from Woodford et al., 2024.

Cross Disciplinary Interoperability Framework

A third Interoperability framework has been proposed in the context of the WorldFAIR project¹², the Cross Disciplinary Interoperability Framework (CDIF)¹³. This framework has been developed using a bottom-up approach based on 11 case studies from 13 different domains, including biodiversity. Based on a common methodology, the 13 different domains have made propositions to support interoperability which have been aggregated and further refined, with the support of a dedicated group of external experts. The

¹¹ <https://doi.org/10.15497/RDA00119>

¹² <https://worldfair-project.eu/>

¹³ <https://zenodo.org/records/11236871>

aim of this framework is to provide practically grounded technical recommendations to support cross-disciplinary interoperability for data aggregation across domains. To support data integration, CDIF leverages DDI-CDI¹⁴, the new standard of the DDI Alliance¹⁵. The purpose of this standard is to describe data and processes with a domain-neutral specification. It covers most data structures commonly used i.e. wide, long, multi-dimensional and key-value pairs and provides a mechanism to interoperate heterogeneous and distributed datasets from multiple domains at the level of the datum itself. CDIF also integrates together existing work to avoid duplication of efforts and supports the adoption of existing recommendations and approaches. This interoperability framework is designed to support the implementation of the FAIR Principles and proposes specific profiles and technical recommendations for different functional aspects: discoverability, access, controlled vocabularies, core data description.

As the main objectives of the Interoperability Framework for Biodiversity Monitoring are to (1) ease the findability and accessibility of a large diversity of biodiversity monitoring datasets at the European level, and (2) ease data integration between datasets from multiple sources to facilitate biodiversity reporting at any scale in Europe, we chose to consider the CDIF practical approach which is more in line with the aim of data integration. To support the implementation of CDIF, we consider the connection between necessary services defined in the EOSC-IF. In this document, we do not consider specifically the GORC Model as the interoperability and the standards and conventions provide high level consideration while the CDIF provides more prescriptive recommendations for interoperability. However, we recommend considering this model to build the interoperability framework at organisational level.

¹⁴ <https://ddialliance.org/ddi-cdi>

¹⁵ <https://ddialliance.org/>

2. The biodiversity monitoring Interoperability Framework

Based on our analysis, we describe here a proposal for an interoperability framework for the biodiversity monitoring and surveillance community. This framework proposes recommendations based on recommendations from existing interoperability framework, adapted and contextualised for the biodiversity community. These recommendations heavily rely on existing standards and propose some customisation that could prove useful. This interoperability framework is not a requirement document, but it rather aims at providing a list of things to consider for becoming interoperable. None of these recommendations are mandatory. They do not have to be implemented all at once; technical choices can be different but must be able to work with the solutions proposed here. Interoperability is a challenging and elaborated journey; organisations should consider applying these recommendations based on the resources available and therefore can start with a small number of recommendations which will bring them closer to full interoperability. Furthermore, this interoperability framework is not set in stone. It should be adapted regularly to integrate new standards and technologies fostering interoperability.

To present the recommendations, we chose to use the following categorisation: models and formats, metadata, persistent identifiers and access and API. These categories can be easily aligned with individual FAIR principles and have been used in the context of the FAIR-IMPACT project to group recommendations with FAIR principles. This alignment is presented in the table below (Table 1).

Table 1. Mapping between the technical categories and the FAIR principles

Models and formats	Interoperability principles
Metadata	Findability and Reusability principles
Persistent IDentifiers (PIDs)	Findability
Access and API	Accessibility

This interoperability Framework has been designed upon the CDIF Interoperability Framework and can be considered as a proposition for a specialised CDIF profile for biodiversity monitoring. Although most of the recommendations are in line with CDIF, we also include additional practical recommendations which highlight the necessary adaptations for the implementation of the framework.

2.1. Models and Formats

In this section, we address the representation of information i.e. both data models and semantic artefacts that can structure both the data and the metadata. Data models describe how the data is organised in the dataset while semantic artefacts provide the meaning of the content of the dataset and help harmonise the description of the concepts measured by the data. Although presented separately, these two models are tightly connected with each other.

2.1.1. Data Models

Biodiversity generates highly diverse datasets. In this report, we focus on two distinct types: **occurrence data**, which are primary observations of species, and **Essential Biodiversity Variables (EBVs)**, which are a first level of synthesized indicators made of the aggregation of occurrence records.

For occurrence data (raw data)

The harmonisation of biodiversity data standards, particularly for species occurrence data, is driven by Biodiversity Information Standards (TDWG, formerly the Taxonomic Databases Working Group), with the strong contribution of the international organisation GBIF. TDWG defined a standard to exchange data across systems, the Darwin Core Archive (DwC - A). This standard proposes to structure the data around the star model (see [Figure 4](#)) interlinking well-defined tables and providing an XML representation of the data structure and associated metadata based on the Ecological Metadata Language (EML). While effective for common use cases, this structure has imposed significant limitations when it comes to modelling more complex relationships, such as species interactions and environmental DNA (eDNA) data. To provide more flexibility, TDWG is working on a new Darwin Core standard called the Darwin Core Data Package (DwC-DP, see [Figure 5](#)). This new data model leverages the Data Package standard¹⁶ (previously called Frictionless Data Package in its first version¹⁷) developed by the Open Knowledge Foundation¹⁸ which provides a container for datasets and a JSON description of the data package content i.e. data source, metadata and data source structure.

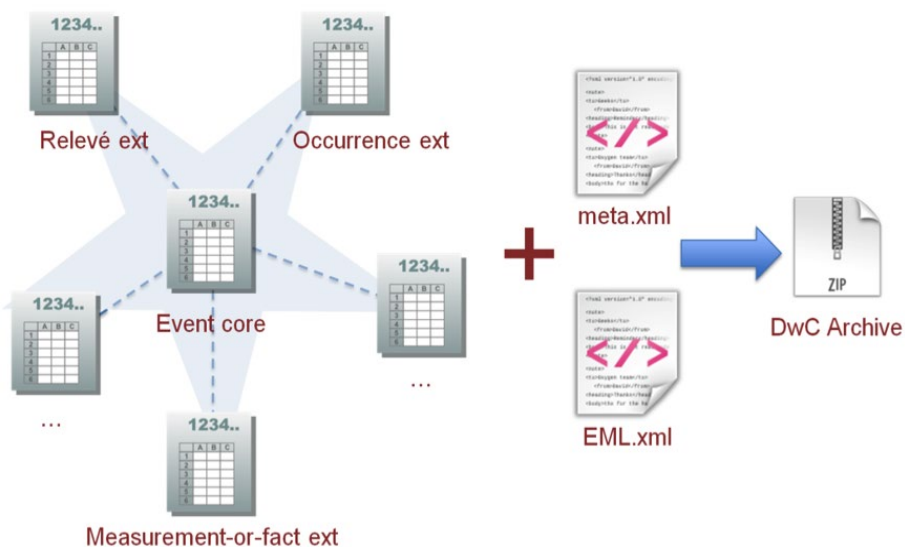


Figure 4. Darwin Core Archive (DwC-A) model (from <https://www.gbif.org/sampling-event-data>)

This new framework relaxes the previous constraints and supports a much wider and richer representation of data such as organism interactions, molecular data, species traits, camera trap

¹⁶ <https://datapackage.org/>

¹⁷ <https://specs.frictionlessdata.io/data-package/>

¹⁸ <https://okfn.org/>

observations, and structured monitoring. In particular, the Humboldt Extension¹⁹, ratified by TDWG in 2024, introduces standardized terms to describe inventory scope, sampling effort, and monitoring protocols, making it especially valuable for integrating long-term and spatially explicit biodiversity monitoring datasets. The DwC-DP accommodates increased complexity by providing 71 different tables²⁰, enhancing its flexibility and applicability to modern biodiversity data needs. Additionally, DwC-DP embeds data validation mechanisms, including requirements for mandatory fields, controlled vocabularies, and expected data types, ensuring improved data quality and consistency across datasets. This new model provides more flexibility to define more complex data structures. The description of the data structure is serialised in a JSON file, and the data model provides metadata using the EML standard. The Data Package (DP) standard allows for the inclusion of extra tables to enable greater extensibility. However, it is important to note that repositories such as GBIF are likely to not adopt these additional custom tables if they fall outside the recognized schema

Generic Recommendation 1

Adopt the upcoming data model to be proposed by TDWG: DwC-DP. We recommend prioritising the use of newly available DwC-DP tables over creating custom extensions whenever possible.

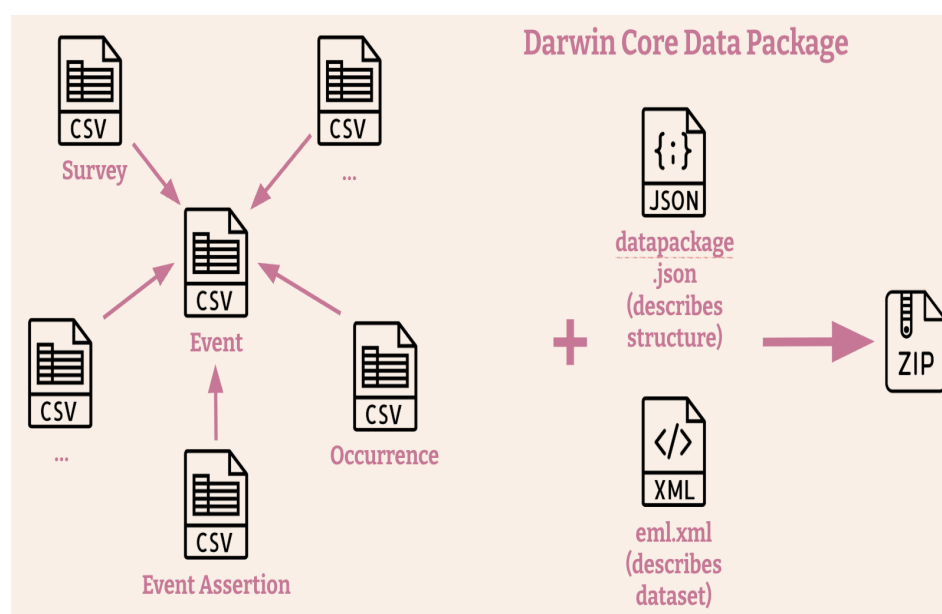


Figure 5. Darwin Core Data Package (DwC-DP) (from <https://gbif.github.io/dwc-dp/> - CC0)

¹⁹ <https://www.tdwg.org/community/osr/humboldt-extension>

²⁰ As of June 2nd 2025

For Essential Biodiversity Variables

Regarding Essential Biodiversity Variables (EBVs), the first proposed standard is EBVCube format²¹, promoted by the GEO BON²² community and leveraging the Climate and Forecast Conventions²³ and the Attribute Convention for Data Discovery²⁴. It contains geospatial raster data stored in NetCDF files and is compatible with EML for the metadata description.

NetCDF (Network Common Data Form) is a set of software libraries and machine-independent data formats that support the creation, access, and sharing of array-oriented scientific data²⁵. The EBV data cube has four dimensions: latitude, longitude, time and the measured entity (e.g. species occurrences) as shown in figure 6.

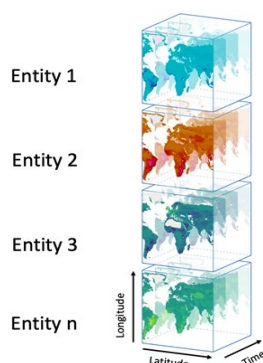


Figure 6. EBV data cube and its four dimensions

The EBVCube enables the possibility to organise the raster data into a hierarchy to group several datasets together. A schematic representation of the EBVCube hierarchical model is shown in figure 7 below.

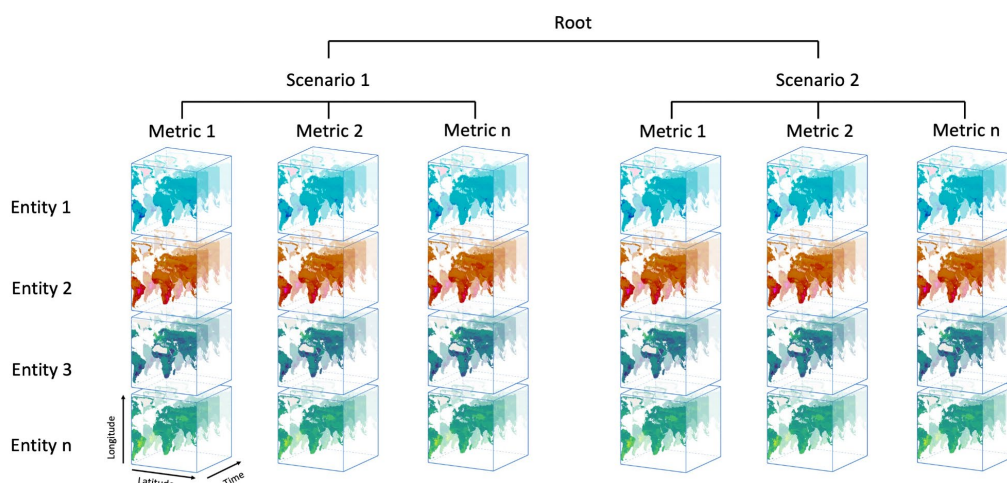


Figure 7. EBV Cube hierarchical structure (from <https://github.com/EBVcube/EBVCube-format>)

²¹ <https://github.com/EBVcube/EBVCube-format>

²² <https://portal.geobon.org/>

²³ <https://cfconventions.org/Data/cf-conventions/cf-conventions-1.8/cf-conventions.html>

²⁴ https://wiki.esipfed.org/Attribute_Convention_for_Data_Discovery_1-3

²⁵ <https://en.wikipedia.org/wiki/NetCDF>

However, not all the EBVs are well-suited to be represented as rasters with a fixed geospatial grid. For instance, the geospatial dimensions may be administrative areas instead of geospatial grids or the EBVs are aggregated for a single location (i.e. geolocalisation variable is a constant) and only vary according to time (i.e. time series). To the best of our knowledge, no alternative standard currently exists for these cases. If data is expected to be of small or moderate size, we suggest considering the RDF Data Cube²⁶ ontology as a possible basis for this new standard. This W3C recommendation has been developed to exchange multi-dimensional statistical dataset in web friendly format and to publish on the web by leveraging the linked data framework based on the W3C RDF standard²⁷. Linked data²⁸ is an approach to publishing data on the web, enabling datasets to be linked together through references to common concepts and relying on HTTP, URIs and RDF. This model is very generic, and it can be used for other data sets such as survey data, spreadsheets and OLAP data cubes.

Generic Recommendation 2

Adopt EBVCube for raster data with spatial coverage. For non-spatial datasets investigate other models such as RDF Data Cube as a possible complementary data model.

2.1.2. Semantic artefacts: knowledge models

Semantic artefacts have been defined in the context of the FAIRsFAIR project²⁹ as follows (Le Franc et al., 2020):

“Semantic artefact is defined in this work as a machine-actionable and -readable formalisation of a conceptualisation, enabling sharing and reuse by humans and machines. These artefacts may have a broad range of formalisation, from loose set of terms, taxonomies, thesauri to higher-order logics. Moreover, semantic artefacts are serialised using a variety of digital representation formats, e.g., RDF Turtle, OWL-RDF, XML, JSON-LD.”

These artefacts play a major role in harmonising the concepts described by or describing the data, therefore, having a major impact on findability, interoperability and reuse of the datasets. In the FAIR principles, the use of semantic artefacts improves interoperability, but they must be themselves FAIR (principle “I2 - (Meta)data use vocabularies that follow the FAIR principles”). The FAIR Semantics recommendations, a first set of recommendations on how to make semantic artefacts FAIR has been proposed in the context of the FAIRsFAIR project (Le Franc et al., 2020; Hugo et al., 2021 and Le Franc et al., 2022). Other recommendations have been proposed by Garijo et al., 2020, Cox et al., 2022 and Xu et al., 2024.

Among these recommendations, semantic artefacts must be documented with proper metadata and should be published in dedicated Semantic Artefact Catalogs (SAC, Corcho et al., 2024). Two main tools

²⁶ <https://www.w3.org/TR/vocab-data-cube/>

²⁷ <https://www.w3.org/TR/2004/REC-rdf-primer-20040210/>

²⁸ https://en.wikipedia.org/wiki/Linked_data

²⁹ <https://www.fairsfair.eu/>

have been developed to evaluate the FAIRness of Semantic artefact: O'FAIRe (Amdouni et al. 2022) and FOOPS! (Garijo et al., 2021).

Taxonomies, controlled vocabularies, and thesauri are commonly used in the field of biodiversity to ensure consistent terminology across datasets. These are typically modelled using the Simple Knowledge Organization System standard (SKOS³⁰), a standard developed by the W3C. SKOS provides a simple, structured way to represent concepts and their relationships (such as broader, narrower, or related terms) in a format that computers can understand and process. Notable examples include the Environmental Thesaurus (EnvThes³¹) for harmonising data resulting from observations and measurements of ecosystem processes, and the Traits Thesaurus³², aiming at standardising traits data and metadata for aquatic organisms. Several formal ontologies, based on the Web Ontology Language (OWL³³) standard, are also in use in biodiversity, such as the Extensible Observation Ontology (OBOE³⁴, Schildhauer et al., 2016), Biological Collections (BCO³⁵) and the Environment Ontology (ENV-O³⁶). These two last ontologies take part in the Open Biological and Biomedical Ontology (OBO) foundry³⁷, which provides a solid methodology for building ontologies in biological sciences and is very well established in the biomedical domain. A general study of semantic artefacts in environmental sciences can be found in Di Muri et al., 2024.

Darwin Core terms³⁸ are also semantic artefacts as they have URLs and are backed by machine-readable and -actionable metadata serialised in RDF. If the terms are individually defined, they are still missing an ontology to organise them. However, an important step in that direction has been achieved with the creation of the DwC Conceptual Model³⁹ (see figure 8) during the design of the upcoming DwC-DP publishing data model. This conceptual model could serve as the foundation of an official formal ontology for Darwin Core.

³⁰ <https://www.w3.org/TR/skos-primer/>

³¹ <http://vocabs.lter-europe.net/EnvThes>

³² https://ecoportal.lifewatch.eu/ontologies/TRAITS_THES

³³ <https://www.w3.org/TR/owl2-overview/>

³⁴ <https://bioportal.bioontology.org/ontologies/OBOE>

³⁵ <https://bioportal.bioontology.org/ontologies/BCO>

³⁶ <http://www.environmentontology.org/>

³⁷ <https://obofoundry.org/>

³⁸ <https://dwc.tdwg.org/terms/>

³⁹ <https://gbif.github.io/dwc-dp/#darwin-core-conceptual-model>

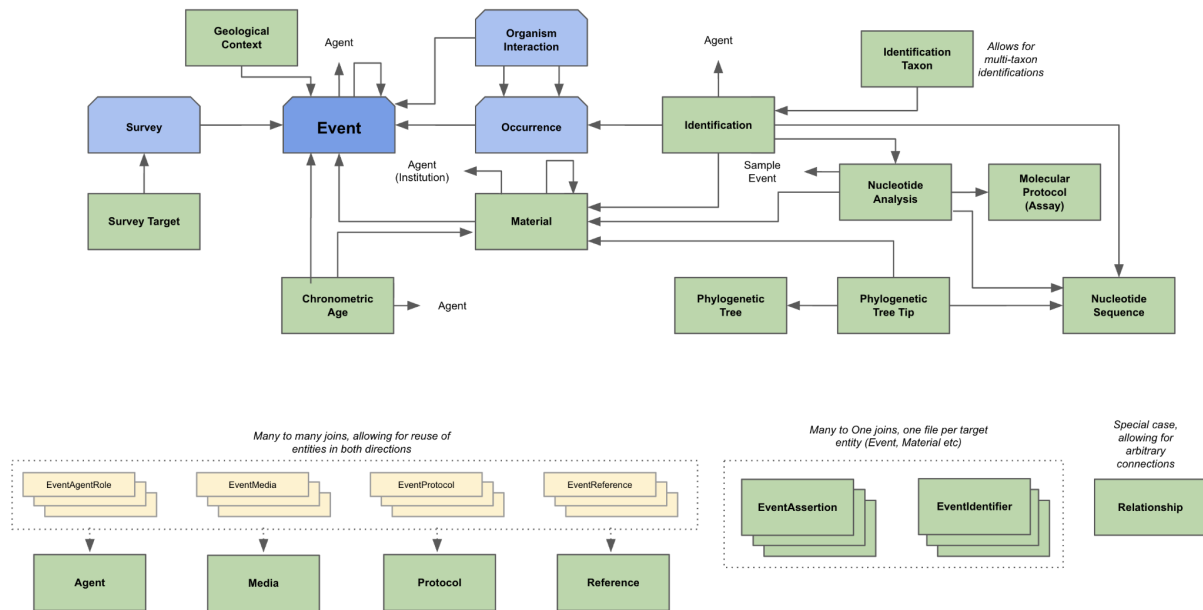


Figure 8. DwC Conceptual model from <https://gbif.github.io/dwc-dp/> (CC0)

Generic Recommendation 3

Use semantic artefacts to align metadata and data to common vocabularies and models.

Generic Recommendation 4

Semantic artefacts in biodiversity must be FAIR.

With the explosion of semantic artefacts emerges the need to build bridges between these models also called mappings. These mappings enable to define relations between concepts from different semantic artefacts. These relations can be relations of equivalence or more. As stated in the FAIR Semantics recommendations, these mappings must be themselves FAIR as they provide the glue to build a global semantic layer between information systems. The project FAIR-IMPACT⁴⁰ proposed a set of recommendations to make mappings FAIR^{41,42} (Le Franc et al., 2025) and provides a Practical Mapping Framework⁴³ (Juty et al., 2023) which describes a coherent process for handling mappings. The work of FAIR-IMPACT is being pursued in the context of the RDA FAIR Mapping WG⁴⁴. This work proposes to use a dedicated model to share and exchange mappings: Simple Standard for Sharing Ontological

⁴⁰ <https://fair-impact.eu/>

⁴¹ <https://doi.org/10.5281/zenodo.15111167>

⁴² <https://mapping-commons.github.io/rda-fair-mappings/recommendation/>

⁴³ <https://doi.org/10.5281/zenodo.15310718>

⁴⁴ <https://www.rd-alliance.org/groups/fair-mappings-wg/activity/>

Mappings⁴⁵ (SSSOM, Matentzoglou et al., 2022). This work has highlighted the diversity of mapping types which can link not only semantic artefacts together but also metadata schema, data models, APIs, and many various cases where it is necessary to build bridges between different information entities of heterogeneous information systems.

To support the FAIRification of mappings and collections of mappings, also called crosswalks, the FAIRCORE4EOSC project has developed the Metadata Schema and Crosswalk Registry (MSCR)⁴⁶ which enables the publication of metadata schema and semantic artefacts, the creation of mappings and their publication.

Generic Recommendation 5

Mappings should be shared and exchanged using SSSOM and any proposed extensions from the RDA FAIR Mapping WG.

Generic Recommendation 6

Mappings whatever their type must be FAIR.

2.1.3. Formats

The data and knowledge models mentioned before coming with a diversity of formats, including CSV, JSON, XML and RDF. In many situations one wants to combine these heterogeneous data sources together within a common format, especially when dealing with metadata. RDF, the standard to implement the concept of Linked Data⁴⁷, is considered the most suitable format for that. It has been designed to interconnect data sources on the web and is at the basis for the Linked Open Data Cloud⁴⁸. The practice of interconnecting distributed and heterogeneous data sources using RDF is promoted by CDIF and implemented in many research infrastructures. To reduce the burden of adopting RDF by existing systems, JSON-LD⁴⁹, a W3C Recommendation, has been proposed to extend systems leveraging JSON. It allows developers to embed semantic context directly within JSON documents by adding a context section, which links data fields to the associated ontologies using web identifiers (URIs). This enables semantic enrichment with minimal changes to the existing JSON structure.

Generic Recommendation 7

Use JSON-LD as a common format for metadata publication and data structure description.

⁴⁵ <https://mapping-commons.github.io/sssom/>

⁴⁶ <https://faircore4eosc.eu/eosc-core-components/metadata-schema-and-crosswalk-registry-mscr>

⁴⁷ <https://www.w3.org/DesignIssues/LinkedData>

⁴⁸ <https://lod-cloud.net/>

⁴⁹ <https://www.w3.org/TR/json-ld11/>

Semantic artefacts are often not provided into a machine-readable and machine-actionable format which dramatically hampers their use and the development of services to leverage them (Di Muri et al., 2024). These invaluable assets should be shared using the existing standard models and their serialisations.

Generic Recommendation 8

Provide conceptual models and classifications for biodiversity monitoring aligned with RDFS, SKOS or OWL models and serialised in one of the RDF or OWL-specific formats (e.g. Turtle, RDF/XML, OWL/XML, etc.).

2.2. Metadata schemas

The most widely used metadata schema in the biodiversity community is the Ecological Metadata Language (EML)⁵⁰. This standard is also used by repositories like GBIF as part of Darwin Core Archive and Darwin Core Data Package. EML is provided as a set of XML schemas tags providing administrative metadata and domain-specific metadata. EML has been designed in a modular way enabling its extension in breadth and depth. The core EML module or container is aligned with Dublin Core⁵¹, general-purpose metadata standard used across domains to describe basic properties of digital resources (e.g., title, creator, date) EML builds on this foundation with 4 additional modules: dataset, literature, software, protocol. These four modules share general information about the resources, including the possibility to add semantic annotations, and other specific information regarding the resource. The metadata schema can be extended with the inclusion of supplementary metadata information in the Additional Metadata module. The different resource modules import different submodules to provide a rich and in-depth metadata description. It covers different types of data e.g. data table, spatial raster, spatial vector and provides in-depth information about the data distribution with the eml-physical module.

It is important to note that in the context of Darwin Core Archive (DwC-A) and Darwin Core Data Package (DwC-DP), EML is not the main source for describing the structure of data tables (table names, column definitions, data types, relationships), the main source is a separate file known as the structural metadata file. EML is also suitable for describing Essential Biodiversity Variables (EBVs), as demonstrated by the GeoBON portal⁵². In practice, most of the EML metadata is automatically extracted from the metadata embedded within the EBVCube⁵³ packages.

Metadata schemas are crucial, but they need to be complemented by semantic artefacts to enhance the effectiveness of dataset searches. This is what the embedding of semantic annotations within EML enables. These annotations allow metadata elements such as attributes or keywords to be linked to well-defined concepts from external ontologies using persistent URIs. These annotations enable the generation of RDF triples, making the metadata interoperable and machine-actionable within Linked Data environments. For example, in an occurrence dataset published to GBIF, a column named habitat might contain free-text values like “temperate forest” or “wetland.” Using EML’s annotation element, the

⁵⁰ <https://eml.ecoinformatics.org/>

⁵¹ <https://www.dublincore.org/specifications/dublin-core/dcmi-terms/>

⁵² <https://portal.geobon.org/datasets>

⁵³ <https://github.com/EBVcube/EBVCube-format>

publisher could link these values to standardized concepts from the ENV-O ontology, such as http://purl.obolibrary.org/obo/ENVO_01000228 for “temperate forest”. This would allow machines to interpret the habitat consistently across datasets and support semantic search.

However, it is important to note that important repositories like GBIF do not yet expose or process such semantic annotations. With the emergence of DwC-DP and the DwC Conceptual Model, GBIF is nevertheless beginning to explore bringing more semantic capabilities to its platform. This initiative may open the door for semantic annotations in the future.

In addition to using EML, it is also recommended to publish metadata using the W3C standard DCAT. DCAT is particularly important for public administrations in the European Union, as they are required to publish their open data using this standard.

Generic Recommendation 9

Use EML and DCAT metadata schema to document biodiversity monitoring datasets.

Generic Recommendation 10

Enrich metadata with semantic annotation leveraging domain specific semantic artefacts.

Generic Recommendation 11

Publish metadata schema and associated mappings into a repository providing Globally Unique Persistent and Resolvable Identifiers and access to the metadata schema and the mapping via an API.

2.3. Persistent Identifiers

One of the essential requirements of the FAIR Principles for Findability is to use Globally Unique Persistent and Resolvable Identifiers (GUPRI) to identify both metadata and data.

The resolution of the GUPRI should enable stakeholders to get automatic access to metadata and data and guarantee the uniqueness of the information element. In the context of the FAIR Digital Object model, this persistent identifier should be associated with information on the type of objects and possible operations on the object.

In this document, we consider two main types of identifiers can ensure global uniqueness and resolvability:

- Handle service-based identifier⁵⁴ (ePIC, DOIs): emulate web-based resolution:
 - DOIs⁵⁵: often points to landing page: need extra information for resolution e.g. signposting;

⁵⁴ <https://www.rfc-editor.org/rfc/rfc3650.html>

⁵⁵ <https://www.doi.org/doi-handbook/HTML/index.html>

- ePIC⁵⁶: configurable to provide access to metadata and data with PID Kernel;
- Persistent URLs: content negotiation. In line with RDF, requires maintenance and naming convention rules.

Persistence is not a technical problem but an organisational problem: who is guaranteeing the persistence of the identifier and for how long? For this reason, specific organisations are supporting these different identifiers such as the DOI Foundation⁵⁷ or the DONA Foundation⁵⁸ and other organisations such as ePIC which will guarantee the persistence of these identifiers by the members of these organisations e.g. DataCite, CrossRef, etc.

The Handle service has been developed by CNRI and is now maintained by the DONA Foundation. It proposes both a namespace (e.g. 21.012/xyz-123) and an infrastructure and services to resolve the PIDs. The Handle system has been designed to function with UDP and TCP protocol and to emulate HTTP/HTTPS through tunnelling. To ensure the persistence, there is a cost for minting PIDs. Different PID providers are using the handle system for the resolution such as the Digital Object Identifier (DOI) or ePIC.

The DOI system has been extensively used for identifying publications and various research objects. The metadata associated with the DOI is defined by the different providers such as DataCite⁵⁹ and is called PID Kernel. Usually, DOIs point to a landing page which proves to be problematic to access metadata and data as it requires the parsing of the page to extract the links to the metadata and the data. A solution to address this issue is to use the FAIR Signposting approach using Web based standards to enrich the header of the webpages with additional controlled information including a list of links to additional resources⁶⁰.

In contrast, PIDs (handles) provided by organisations such as ePIC can be configured to have a specific PID Kernel which can include pointers to domain specific metadata and different distributions of the datasets. For both systems, the use of DOIs or PIDs has a cost for registering a prefix for organisations and for the minting and maintenance of the PIDs.

A less constrained approach and free approach is to use Persistent URLs or PURL. Originally, this was proposed by OCLC in 1995⁶¹. The PURL is a URL that is redirected to the location of the requested resource. A service has been developed to support the creation, management and resolution of PURL with the common namespace <https://purl.org>.

In 2016, OCLC had to shut down the operations of the service and then established a partnership with Internet Archive to maintain the original PURL services. Another solution based on the same principles has been proposed w3id.org, using GitHub to create and manage the PURL⁶². PURLs are widely used in the context of Linked data and rely on the scalability of the web.

⁵⁶ <https://www.pidconsortium.net/>

⁵⁷ <https://www.doi.org/>

⁵⁸ <https://www.dona.net/index>

⁵⁹ <https://schema.datacite.org/meta/kernel-4.6/>

⁶⁰ <https://signposting.org/FAIR/>

⁶¹ https://en.wikipedia.org/wiki/Persistent_uniform_resource_locator

⁶² <https://w3id.org/>

It is difficult to recommend one unique set of Persistent Identifiers. For example, DOIs should be used to publish datasets for citation purposes. To compensate for the drawback of the landing page, the use of FAIR Signposting should be considered. For semantic artefacts, we would advise using the PURL approach. Due to the possibility to resolve handles with HTTP/HTTPS protocols, these different approaches can be used together in a coherent framework such as Linked Data.

Generic Recommendation 12

Datasets and metadata records must be identified by a Globally Unique, Persistent and Resolvable Identifier.

2.4. API and Access

One of the objectives of the FAIR principles is to enable machine-actionability to design information systems where machines can support us with data management and manipulations. For this, it is necessary to provide programmatic access to the metadata and data to automate metadata harvesting for enhanced search capabilities and/or automate data integration and data analysis workflows. This is emphasised by the principle “A1- (Meta)data are retrievable by their identifier using a standardised communication protocol”.

Generic recommendation 13

Provide an API for programmatic access to data and metadata.

The most common form of API proposed by data services on the web is the HTTP REST API which provides access to the data and the metadata using HTTP standard methods and URIs to identify the resources. REST APIs provide a unique and unified access point for clients to get information from the servers. However, connecting to any REST API would require finding the API, reading the documentation of the API and to develop a script or adapt a client to collect the appropriate information. To support the documentation of APIs, the OpenAPI specification⁶³ has been proposed to harmonise the format of API documentation for human readability. OpenAPI descriptions provided as JSON or YAML are thus used to automatically generate documentation webpages. As these descriptions are structured in a machine-actionable format, they should be published to enable smart clients to easily connect with different APIs without changing the code. In a sense, API should be also FAIR. To support this idea, an extension of the OpenAPI specification has been proposed, smartAPI⁶⁴, which enables semantic annotations of the API description and metadata to support easier discovery of APIs and their specific calls. Such extended description supports FAIR API descriptions.

⁶³ <https://swagger.io/specification/>

⁶⁴ <https://smart-api.info/>

Generic Recommendation 14

Provide a machine-readable description of the API, accessible with a specific API call.

Through the API it is necessary to provide access to the API description but also to the metadata schema provided by the service. Having access to the metadata schema from the API will enable us to easily create mappings between the schema provided by the API and the metadata schema expected by the client.

Generic Recommendation 15

Provide access to the metadata schema used by the service to structure the information collected from the API.

Although it is essential to provide an open access to the metadata to support the creation of search engines which can crawl the metadata content of different data repositories, access to data might be regulated e.g. the access to personal information or information regarding protected species. Restriction of access can be done in different ways but often require manual interventions. Access policies are often provided as a free-text document. To harmonise the access policies descriptions and improve the machine actionability of the access policies, CDIF proposes to use the Open Digital Right Language (ODRL) model/ontology to cover cases where data access should be controlled (see figure 9). ODRL is an RDF based model that can be serialised in different formats proposed by the standard organisation W3C. It is widely adopted for expression permissions, conditions and obligations.

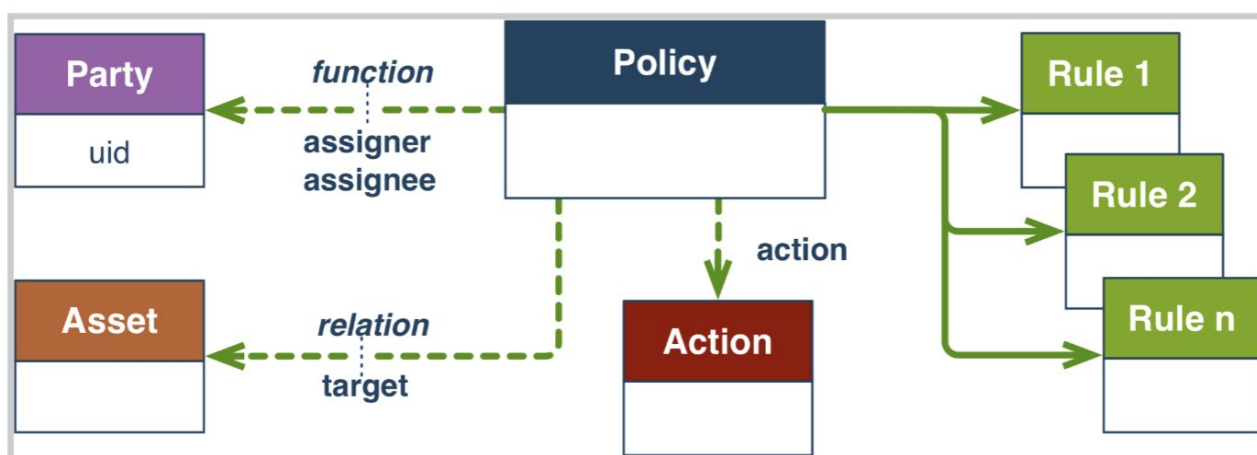


Figure 9. Schematic diagram showing the relations between the classes (from CDIF Report, Gregory et al., 2024)

The core classes of ODRL are:

- **Party**: actors or agents;
- **Rule**: defines what a party can do (permissions), must do(duties) and must not do (prohibitions);

- **Action:** operations that can be performed on digital assets;
- **Assets:** digital objects in a CDIF context;
- **Policy:** a structured artefact which encapsulates the above information.

The simplified diagram above (figure 9), extracted from the CDIF report (Gregory et al., 2024)⁶⁵, shows the relationship between the different classes.

Although ODRL does not cover natively all the access use-cases, it offers different solutions to extend its coverage. For more information and more practical examples, we invite the reader to explore section 6.4 (p50) of the CDIF report. This section looks at different access cases, including simple direct access for which they derived the brokering process for data access requests (see figure 10 below).

ODRL Brokerage: Simple Resource Request

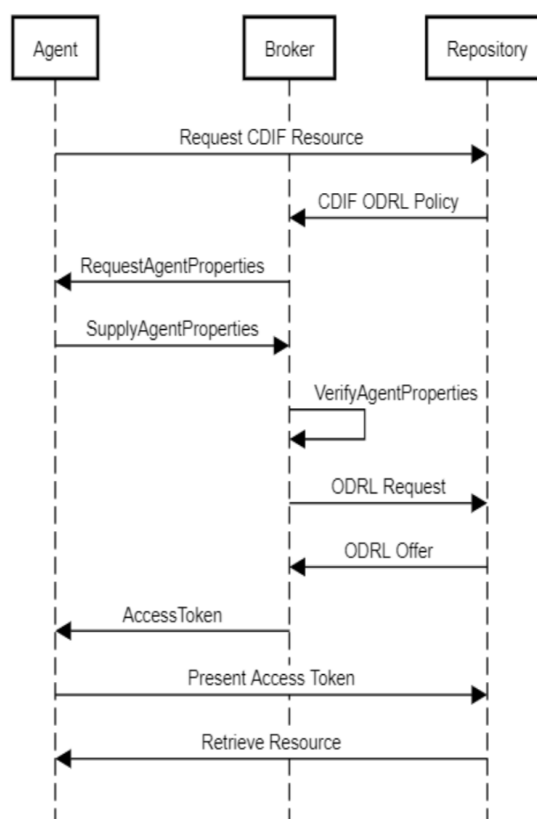


Figure 10. Flow diagram of the information brokering for simple access request (extracted from CDIF Report, Gregory et al., 2024)

Generic Recommendation 16

Whenever possible, describe data access procedures using ODRL as proposed by CDIF.

⁶⁵ <https://doi.org/10.5281/zenodo.11236871>

3. Putting the Framework in Practice

The initial goals of this framework are (1) to improve findability/discoverability of the datasets collected by the various biodiversity monitoring and surveillance organisations across Europe and (2) to ease integration of datasets from various sources. In this section, we propose practical recommendations to address these goals in line with the general recommendations of the interoperability framework.

3.1. Findability/discoverability

Most of the generic recommendations of this interoperability framework are linked to findability. Finding the right datasets is the starting point for any data aggregation and integration activities. Data discoverability poses a real challenge requiring for the biodiversity monitoring community to agree on common standards for describing the data (metadata schemas) and for harmonising the semantics of the metadata layer. Converging towards common metadata schemas will ease the process of either harvesting the metadata for creating a centralised data catalogue or performing federated queries on distributed and heterogeneous information systems. Either way, it is necessary to harmonise the semantics of the metadata. For this, semantic artefacts can be used for annotating metadata and data, providing a way to harmonise domain expert terminologies and knowledge models.

To enable the capacity of finding datasets on distributed and heterogeneous data sources, several approaches already exist. The most appropriate solution to interconnect the content of the distributed data sources is to leverage the standards in relation with the concept of Linked Data on the web. The implementation of the Linked Data is based on W3C recommendations such as RDF, RDFS, OWL, Linked Data Platform⁶⁶, SHACL⁶⁷, or JSON-LD⁶⁸. The whole approach is centred on RDF which provides a model for connecting two data elements together as a triple where one data element would be the subject, the second data element is the object and the relation that links them together is the predicate. Both the subject and the object can be related to any other subjects and objects. Adopting this representation would enable biodiversity monitoring organisations to align and harmonise their content to be leveraged for more efficient searches.

One of the generic recommendations concerning metadata is to use EML and DCAT as standards for publishing the metadata (Generic recommendation 9). DCAT is a data model proposed by W3C for sharing information about collections of datasets. It has been designed as an RDF model therefore it is already compliant with our recommendation to use JSON-LD (Generic recommendation 7). Several biodiversity monitoring organisations already publish their metadata using the DCAT model and the associated profiles which have been adopted by the European Commission and the Member states to publish the metadata associated with open governmental data⁶⁹. EML, on the contrary, is not directly compatible with our recommendation of using RDF/JSON-LD. Therefore, it is necessary to consider developing a converter to transform EML into a JSON-LD structure.

⁶⁶ <https://www.w3.org/TR/ldp/>

⁶⁷ <https://www.w3.org/TR/shacl/>

⁶⁸ <https://www.w3.org/TR/json-ld11/>

⁶⁹ <https://interoperable-europe.ec.europa.eu/collection/semic-support-centre>

Practical Recommendation 1

Build a converter from EML to JSON-LD that can be used by the whole biodiversity monitoring community.

During this process, core fields will be aligned with standard semantic artefacts (general or from environmental sciences) and additional metadata fields with community-specific semantic artefacts, making the meaning of the metadata more explicit for the machines and the humans. The alignment of some of the core fields and additional metadata fields with standard semantic artefacts can be extracted from triples encoded in EML semantic annotations^{70, 71} (see Generic recommendation 10). Finally, although different in many ways, EML and DCAT share a common set of metadata elements. Several initiatives have proposed mappings between EML and DCAT^{72, 73}.

Practical Recommendation 2

Use EML semantic annotations to enrich the metadata with explicit semantics provided by semantic artefacts.

Having an official semantic artefact for Darwin Core in the form of a formal ontology would be of great help to enrich metadata and for describing occurrence data in RDF with ease. The recent development of the DwC Conceptual Model could serve as a solid foundation for formalising an ontology in RDFS or OWL, according to the level of expressivity aimed. New DwC terms will likely have to be created for object properties so as to guarantee a good RDF experience for describing actual occurrence data, as demonstrated by a previous proposal (Baskauf et al., 2021) addressing the limitations of the current RDF guide⁷⁴ (Darwin Core and RDF/OWL Task Groups, 2021).

Practical Recommendation 3

Provide the DwC Conceptual Model as a semantic artefact.

Users gain a more powerful search experience by combining metadata schemas with semantic artefacts through semantic annotations. This integration supports features⁷⁵ such as:

- **Synonym handling:** searching for "carbon dioxide flux" should yield results that include "CO2 flux";

⁷⁰ <https://eml.ecoinformatics.org/semantic-annotation-primer>

⁷¹ As a side remark, we recommend giving preference to SKOS taxonomies over formal ontologies in semantic annotations as they are safer to use, especially for users unfamiliar with logical systems. Formal ontologies involve logical rules that may lead to unintended conclusions if not used carefully.

⁷² <https://github.com/inbo/ipt-dcat>

⁷³ https://gitlab.in2p3.fr/aurelien.schmitt/eml_dcatap

⁷⁴ <https://dwc.tdwg.org/rdf/>

⁷⁵ Examples taken from <https://eml.ecoinformatics.org/semantic-annotation-primer>

- **Term disambiguation:** replacing the string "litter" with an IRI clarifies whether it refers to "plant litter," "garbage," or a "group of animals born together";
- **Hierarchical searches:** results related to carbon flux should encompass information on both carbon dioxide and carbon monoxide fluxes.

Furthermore, semantic artefacts can be mapped to one another, enriching and broadening search capabilities across diverse data sources. It is essential to use FAIR semantic artefacts to enable search engines to deliver such advanced experiences.

In addition to EML and DCAT, which are recommended by this interoperability framework, schema.org may also be considered to ensure that metadata is indexed by search engines that focus on datasets, such as Google Dataset Search⁷⁶. Such an extra effort is foreseen to be lightweight as mappings from DCAT to schema.org exist⁷⁷.

Semantic annotations, semantic artefacts, DCAT and [Schema.org](https://schema.org) are all based on RDF. To implement a search engine over this data, it is common to involve databases specialised in storing RDF data, which are called triplestores, which are nowadays efficient at managing billions of triples. Note that metadata tends to be many orders of magnitude smaller than data, so scalability in terms of performance is rarely a concern here. All this metadata can be merged into the same *metadata knowledge graph* and queried together with the same SPARQL query. Finally, since the stack is domain-agnostic, it can easily accommodate metadata from other research domains and facilitate their integration with biodiversity metadata. This is particularly attractive as many related communities such as SeaDataNet⁷⁸ have already migrated to the RDF stack.

Integrating semantics with the metadata requires to easily find and reuse the relevant semantic artefacts and to have access to their content through User Interfaces for humans and through APIs for machines. Therefore, it is essential to publish and share biodiversity related semantic artefacts into a dedicated catalogue, called Semantic Artefact Catalogues (SAC, Corcho et al., 2024). These services make semantic artefacts content available through APIs and offer the possibility to explore them both manually or programmatically, offering complex search capabilities. Using these services is also in line with recommendations from EOSC-IF and the recommendations of the EOSC Semantic Interoperability Task Force.

Practical Recommendation 4

Publish biodiversity monitoring semantic artefacts in a Semantic Artefact Catalogues such as EcoPortal⁷⁹, provided by LifeWatch ERIC, or BiodivPortal⁸⁰, provided by GFBio and NFDI4Biodiversity.

Semantic artefact catalogues also offer search functionalities for retrieving relevant ontologies. For this purpose, ontologies should be also documented with proper metadata. Several metadata schemas and

⁷⁶ <https://datasetsearch.research.google.com/>

⁷⁷ <https://ec-jrc.github.io/dcat-ap-to-schema-org/>

⁷⁸ <https://www.seadatanet.org/>

⁷⁹ <https://ecoportal.lifewatch.eu/>

⁸⁰ <https://biodivportal.gfbio.org/>

ontologies have been developed over the years to describe ontologies but most of them were incomplete or are not maintained anymore, therefore creating a gap. The MOD ontology⁸¹ has been developed to fill this gap and to provide many metadata elements to describe ontologies, which have been mapped with the legacy ontologies. In the context of the FAIRsFAIR project and the RDA Vocabulary and Semantic Service Interest Group⁸², MOD has been aligned with DCAT to support the publication of collections of semantic artefacts and a subset of MOD concepts have been identified to create a minimum set of common metadata for FAIR Semantic Artefacts (Jonquet et al., 2023). This extension of MOD has been consolidated in the FAIR-IMPACT project (Gonzalez Beltram et al., 2024) and has been used to define a common API for Semantic Artefact Catalogues (Wilson et al., 2024).

Practical Recommendation 5
Provide metadata to document semantic artefacts using the MOD Ontology.

A large diversity of semantic artefacts exists in the biodiversity domain with various levels of complexity, granularity and expressivity. It would be unreasonable to consider building only one representation of a particular domain which would be adopted by the whole community. Indeed, semantic artefacts are created with a particular and limited perspective. To overcome this and cope with the diversity of representations, it is possible to build bridges between the models to make them work together. Aligning semantic artefacts together can be resource intensive and may be useful in other contexts. Therefore, it is essential to promote these mappings by publishing them in a dedicated repository which will provide access to the mappings both through a user interface or through an API.

Practical Recommendation 6
Mappings should be published into a mapping repository such as the Metadata Schema and Crosswalk Registry (MSCR) ⁸³ .

The problem of findability is not uniquely related to metadata in the context of data integration. Part of the important information for deciding which datasets to integrate relies on the data structure itself and its relation to the semantic description of the variable measured. In addition to transforming the way the biodiversity monitoring community handles metadata, it is also necessary to consider how to handle the data structure description (e.g. wide, long, key-value pairs, ...). To retrieve datasets based on their content, it is necessary to make them accessible to the search. In practice, this means that the metadata in the “meta.xml” file provided in the DwC-A or in the “datapackage.json” file in DwC-DP should be accessible for search. The evolution of the DwC model towards the Data Package structure is an important step to ease this capability. Indeed, the data structure is now described in a JSON structure which can be aligned with terms from DwC. Thus, this eases the possibility to convert the

⁸¹ <https://github.com/FAIR-IMPACT/MOD?tab=readme-ov-file>
⁸² <https://www.rd-alliance.org/groups/vocabulary-services-interest-group/activity/>
⁸³ <https://mscr-release.2.rahtiapp.fi/>

“datapackage.json” file to JSON-LD and therefore enables the integration of this information into the knowledge graph for the discovery. An example of such conversion is shown below in Figure 11, demonstrating the minimal adaptation required for this transformation and the resulting Turtle representation (a compact and human-readable format for expressing RDF triples).

Figure 11. Example of JSON-LD conversion of DwC-DP JSON metadata

Adding

```
"@context": {
  "dc": "http://purl.org/dc/terms/",
  "schema": "https://schema.org/",
  "@vocab": "http://example.org/dwc-dp/voc/",
  "voc": "http://example.org/dwc-dp/voc/",
  "url": {
    "@id": "schema:url",
    "@type": "@id"
  },
  "name": "schema:name",
  "version": "schema:version",
  "title": "dc:title",
  "description": "dc:description",
  "issued": {
    "@id": "dc:issued",
    "@type": "xsd:date"
  },
  "isLatest": {
    "@type": "xsd:boolean"
  },
  "tableSchemas": {
    "@container": "@set"
  },
  "xsd": "http://www.w3.org/2001/XMLSchema#",
  "identifier": "@id"
}
```

to

```
{
  "identifier": "http://rs.tdwg.org/dwc/dwc-dp",
  "url": "https://rs.gbif.org/sandbox/experimental/data-packages/dwc-dp/0.1",
  "name": "dwc-dp",
  "version": "0.1",
  "tableSchemas": [
    {
      "identifier": "http://rs.tdwg.org/dwc/dwc-dp/agent",
      "url": "https://rs.gbif.org/sandbox/experimental/data-packages/dwc-dp/0.1/table-
schemas/agent.json",
      "name": "agent",
      "title": "Agent",
      "description": "A person, group, organization, machine, software or other entity
that can act.",
      "comments": "",
      "examples": "`Carl Linnaeus`; `The Terra Nova Expedition`; `The National Science
Foundation`; `The El Yunque National Forest ARBIMON System`; `ChatGPT`",
      "namespace": "dwcp",
      "iri": ""
    }
  ],
  "title": "Darwin Core Data Package",
  "shortTitle": "dwc-dp",
  "description": "A data package for sharing biodiversity data using Darwin Core.",
}
```

```
"issued": "2025-06-10",  
"isLatest": true  
}
```

Produces in Turtle

```
@prefix schema: <http://schema.org/> .  
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .  
@prefix dcterms: <http://purl.org/dc/terms/> .  
@prefix voc: <http://example.org/dwc-dp/voc/> .  
@prefix name: <schema:name> .  
@prefix version: <schema:version> .  
@prefix title: <dc:title> .  
@prefix description: <dc:description> .
```

```
<http://rs.tdwg.org/dwc/dwc-dp> <https://schema.org/url> <https://rs.gbif.org/sandbox/experimental/data-  
packages/dwc-dp/0.1> ;  
  <https://schema.org/name> "dwc-dp" ;  
  <https://schema.org/version> "0.1" ;  
  voc:tableSchemas <http://rs.tdwg.org/dwc/dwc-dp/agent> ;  
  dcterms:title "Darwin Core Data Package" ;  
  voc:shortTitle "dwc-dp" ;  
  dcterms:description "A data package for sharing biodiversity data using Darwin Core." ;  
  dcterms:issued "2025-06-10"^^xsd:date ;  
  voc:isLatest true .
```

```
<http://rs.tdwg.org/dwc/dwc-dp/agent> <https://schema.org/url>  
<https://rs.gbif.org/sandbox/experimental/data-packages/dwc-dp/0.1/table-schemas/agent.json> ;  
  <https://schema.org/name> "agent" ;  
  dcterms:title "Agent" ;  
  dcterms:description "A person, group, organization, machine, software or other entity that can act." ;  
  voc:comments "" ;  
  voc:examples "`Carl Linnaeus`"; "`The Terra Nova Expedition`"; "`The National Science Foundation`";  
  "`The El Yunque National Forest ARBIMON System`"; "`ChatGPT`" ;  
  voc:namespace "dwcp" ;  
  voc:iri "" .
```

Practical Recommendation 7

Convert datapackage.json into JSON-LD and publish the description of the data structure.

In practice, the implementation of this framework will require serious internal changes and adaptations of the existing information systems. To reduce the cost and accelerate the increase in interoperability, we recommend leveraging existing services.

To publish metadata in DCAT, organisations can leverage a FAIR Data Point (FDP, Bonino da Silva Santos et al., 2022) either by aligning their APIs with the required specifications or by using a third-party service. This process requires a transformer to map internal data models to the DCAT standard. One possible solution is to use FAIRCAt (Le Franc et al., 2022; Jonquet et al., 2023), a tool developed within the FAIRsFAIR project⁸⁴. This service connects to the API, collects the metadata and enables the

⁸⁴ <https://www.fairsfair.eu/>

mapping between the data source metadata schema and DCAT. When data is available in JSON format, it can be easily converted to RDF by adding a JSON-LD context⁸⁵ and making minor adjustments to the JSON structure. By using the FAIR Data Point as an additional database to publish the metadata in another format, it prevents the organisation from redeveloping their platform completely. One of the key functionalities of the FAIR Data Point is the possibility to create a federation of FDP enabling federated queries across the different FDP.

3.2. Data integration

The RDF stack can also facilitate the integration of heterogeneous datasets together by merging them into *data knowledge graphs*. These data knowledge graphs can be built on demand from the datasets selected during the search phase for performing studies on specific topics. Various mapping languages for mapping non-RDF data models to RDF exist, such as the W3C standard R2RML⁸⁶ for relational data and RML⁸⁷ for arbitrary data sources.

For instance, once an ontology would have been developed for Darwin Core, one could integrate DwC-DP datasets into data knowledge graphs the following way:

- Load the DwC-DP datasets into a relational database⁸⁸ using IPT⁸⁹;
- Map the tables and columns in the relational database to the ontology using R2RML;
- Materialise the data into a triplestore or virtualise it as a virtual knowledge graph with a framework like Ontop⁹⁰. With virtualisation, no additional copy of the data is created, and no extra database is needed.

Both data and metadata knowledge graphs can be queried together through SPARQL queries, enabling the possibility to perform advanced data analyses.

For more advanced automated data integration and for the integration of complex datasets, we propose to investigate at term the use of the DDI-CDI model which provides a more granular description of the data structure, including the meaning of the data element with the use of semantic artefacts. Such granular description enables the possibility to automate data integration processes. Therefore, we recommend using the DDI-CDI model to publish a machine actionable description of the data structure. The Data Package approach should be used for data exchange while the DDI-CDI would enable the automation of data integration from distributed sources. The use of DDI-CDI could be also considered as a good way to represent EBVs datasets either raster or not. DDI-CDI has been developed to support the description and the integration of complex statistical data.

⁸⁵ <https://www.w3.org/TR/json-ld11/#the-context>

⁸⁶ <https://www.w3.org/TR/r2rml/>

⁸⁷ <https://kg-construct.github.io/rml-core/spec/docs/>

⁸⁸ https://github.com/gbif/dwc-dp-examples/blob/master/gbif/dwc_dp_schema.sql

⁸⁹ <https://www.gbif.org/ipt>

⁹⁰ <https://ontop-vkg.org/>

Practical Recommendation 8

Publish DDI-CDI data description to ease complex data integration and automation.

Another important aspect to consider is the traceability of information and this is of particular importance when performing data integration to be able to identify the sources of data, document the transformation that happened on the data to support the integration.

This aspect is one of the major topics of work for CDIF. A first simple recommendation would be to consider the use of the W3C recommendation PROV which provides a simple model to capture and represent provenance. We also recommend following and contributing to the ongoing work of CDIF.

4. Conclusions

The biodiversity community is an (interacting) system of initiatives and organisations of different scales (subnational, national, and European) and different purposes (research, monitoring, policymaker, etc.) (Basille et al., 2024). These organisations collect vast amounts of data that are useful for the development, implementation and evaluation of biodiversity related policies, informed private decision-making processes and research purposes. To be fully exploited, this data must be shared and aggregated from multiple sources. Currently, most organisations manage their data on their own infrastructure and services, developed in silos.

This represents a significant barrier to retrieving and integrating datasets from multiple distributed and heterogeneous data sources. To address this issue, the community established various standards to harmonise the metadata and the data structure. Although such standards have been established, it is still impossible to retrieve easily and efficiently datasets from various biodiversity monitoring organisations in Europe and to integrate them together due to the lack of interoperability between the different information systems. Research organisations in biodiversity are facing similar problems but are constantly improving the situation with various drivers such as the increasing need from their community, the FAIR principles and the emergence of the European Open Science Cloud which aims at becoming a web of FAIR data and services. In this context, many research organisations and infrastructures develop solutions to integrate distributed and heterogeneous data sources, using sets of common rules and technical approaches between the different organisations, i.e. interoperability frameworks.

This document presented a first proposition for an interoperability framework for biodiversity monitoring, in line with existing interoperability frameworks developed in various contexts and technical approaches developed by the research communities. This interoperability framework aims at supporting data integration between the different biodiversity monitoring organisations for reporting purposes, research purposes, etc. The Interoperability Framework provides 16 generic and 8 practical recommendations and is strongly inspired by the Cross-Disciplinary Interoperability Framework (CDIF). It can be considered as a CDIF profile specific for biodiversity monitoring, enriched with recommendations from other interoperability frameworks.

As it was mentioned earlier, reaching interoperability is a journey. The recommendations do not have to be implemented in full in any case, instead even a subset of the recommendations can be implemented to initiate interoperability. The implementation of the recommendations will depend on the resources available to the different organisations and more importantly the cost of the implementation. A step-by-step approach can be devised, starting from low hanging fruits to more resource-demanding recommendations which will gradually ease findability and data integration. The initial focus to start implementing interoperability should be on findability.

The **first step** should be to implement recommendations considered as a low hanging fruit for the biodiversity monitoring community i.e. “*Convert EML to JSON-LD*” (**practical recommendation 1**) and “*use EML semantic annotations to enrich the metadata with explicit semantics provided by semantic artefacts*” (**Practical recommendation 2**). The implementation of these recommendations does not require too many resources for biodiversity monitoring organisations that adopted EML as metadata

schema. A converter from EML to JSON-LD already exists as a R package⁹¹ which could be further developed as a web service to support the different biodiversity monitoring organisations. The enrichment of the EML using semantic annotations would require some tedious work which should be performed once manually and could then be replicated automatically to all the metadata records. This enrichment with the semantic annotations will allow the generation of JSON-LD directly aligned with relevant semantic artefacts. By performing this transformation and providing access to the JSON-LD files, it becomes possible to collect and aggregate these JSON-LD files into a centralised catalogue (metadata knowledge graph) which will allow searching across all the organisations which made the change. The cost of creating this catalogue should be minimal, except for the collection of the metadata from the resources which will require connecting to many different APIs when they exist. In this step, semantic artefacts are essential to enrich the discoverability experience. Therefore, **practical recommendations 4, 5 and 6** should be implemented. This will require investing in the development of missing semantic artefacts, transforming existing standards as semantic artefacts and documenting and publishing them. This can be costly and time consuming. However, the cost can be shared with the existing biodiversity research infrastructures which are providing services and expertise.

The **second step** should be to implement recommendations related to data and the description of the data structure i.e. adopt the DwC Data Package format (DwC-DP) (see **generic recommendation 1**) and convert datapackage.json into JSON-LD and publish the description of the data structure (see **practical recommendation 7**). This change may require a lot of resources if the organisation chose to adopt it for their core model for the database. It would be less if they chose to use it as a reference format for data exchange. It would require a converter from the core model of the database to DwC-DP and from DwC-A to DwC-DP. If the cost of change of the database model is too expensive, it should be sufficient to publish the data structure from the “datapackage.json” file as JSON-LD. This transformation does not cost much as the conversion effort is very limited (see **figure 11**). The main cost will be the creation of the “datapackage.json” file for the different datasets. An additional activity to consider would be to participate in the community review of the DwC-DP proposal, starting in September 2025. Give feedback to make sure it suits Biodiversa+ partners existing use cases. At this stage, it would be also interesting to implement practical recommendation 3, which asks for the DwC conceptual model to be serialised as a semantic artefact. This work can be costly and take a lot of time. However, it is done in collaboration with the different members of the GBIF global collaboration.

The **third step** would be to implement the recommendations about APIs i.e. **generic recommendation 13** (“Provide an API for programmatic access to data and metadata”, **generic recommendation 14** (“Provide a machine-readable description of the API, accessible with a specific API call. “) and **generic recommendation 15** (“Provide access to the metadata schema used by the service to structure the information collected from the API.”). These recommendations require a serious investment for organisations that do not have implemented them. For the other, it would mean updating their API with additional calls to collect the API description in a machine-readable format, the metadata schema and the dataset structure as JSON-LD. The development of the API is essential to interconnect systems together and more importantly to interact with external services which will support the interoperability such as semantic artefact catalogues, mapping and metadata repositories, etc.

⁹¹ <https://docs.ropensci.org/emld/>

To ease the interconnection between the systems, it would be ideal to implement the FAIR Data Point specification, leveraging DCAT and the W3C Linked Data Platform recommendations. This implementation will require for the organisation already proposing APIs a large investment to transform their API while if taken into consideration in the design phase would reduce the cost. To reduce the cost further, it is also possible to leverage existing FAIR Data Point services which will be used as an external metadata store with an API compliant with the FAIR Data Point specification.

While the three first steps were focused on metadata and findability, the **fourth step** should support data integration. Data integration requires not only access. The publication of the DwC-DP data structure as JSON-LD should address most of the simple cases of data integration. For more complex integration problems, it becomes necessary to investigate more expressive models that can be manipulated to semi-automated data integration, and particularly the DDI-CDI model (**practical recommendation 8**). To address this, Biodiversa+ partners should consider investing a serious number of resources for training and implementation. The learning curve is quite steep, but it is compensated by the added value for the data integration effort.

This interoperability framework is not intended to remain static and should be improved by the experience of implementation and new technologies and standards. It proposes “low-cost” recommendations to start the interoperability journey. While this document focuses on technical recommendations, achieving interoperability also depends on organisational and institutional factors, such as incentives, governance arrangements, training, and long-term sustainability. It should be noted that at the data collection level, capacities for data management and handling are often limited. While larger research infrastructures may implement advanced interoperability measures, many smaller or local initiatives still require guidance and support to adopt standards and integrate their data effectively. Without parallel investment in these dimensions, technical measures alone are unlikely to be fully effective. To support these investments, this work highlighted the GORC International Model which covers these diverse aspects which would be beneficial to support the creation of a Biodiversity Monitoring Commons.

5. References

- Alberto Basset, Senem Onen Tarantini, Hilde Eggermont, Cécile Mandon, Petteri Vihervaara. D2.2 - Report on the harmonisation and interoperability of datasets across regions and countries (2023), <https://www.biodiversa.eu/wp-content/uploads/2023/05/D2.2-Report-data-interoperability.pdf>
- Hardisty, Alex, Belbin, Lee, Hobern, Donald, McGeoch, Melodie A., Pirzl, Rebecca, Williams, Kristen J. and Kissling, W. Daniel 2019. Research infrastructure challenges in preparing essential biodiversity variables data products for alien invasive species. *Environmental Research Letters* 14, 025005. [10.1088/1748-9326/aaf5db](https://doi.org/10.1088/1748-9326/aaf5db)
- Matthew B. Jones, Margaret O'Brien, Bryce Mecum, Carl Boettiger, Mark Schildhauer, Mitchell Maier, Timothy Whiteaker, Stevan Earl, Steven Chong. 2019. Ecological Metadata Language version 2.2.0. KNB Data Repository. <https://doi.org/10.5063/F11834T2>
- Wieczorek J, Bloom D, Guralnick R, Blum S, Döring M, et al. (2012) Darwin Core: An Evolving Community-Developed Biodiversity Data Standard. *PLoS ONE* 7(1): e29715. <https://doi.org/10.1371/journal.pone.0029715>
- European Commission: Directorate-General for Research and Innovation, EOSC Executive Board, Corcho, O., Eriksson, M., Kurowski, K., Ojsteršek, M., Choirat, C., Sanden, M. v. d., & Coppens, F. (2021). EOSC interoperability framework: report from the EOSC Executive Board Working Groups FAIR and Architecture, Publications Office. <https://data.europa.eu/doi/10.2777/620649>
- Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* 3, 160018 (2016). <https://doi.org/10.1038/sdata.2016.18>
- European Commission: Directorate-General for Digital Services, New European interoperability framework – Promoting seamless services and data flows for European public administrations, Publications Office, 2017, <https://data.europa.eu/doi/10.2799/78681>
- European Commission: Directorate-General for Research and Innovation, Turning FAIR into reality – Final report and action plan from the European Commission expert group on FAIR data, Publications Office, 2018, <https://data.europa.eu/doi/10.2777/1524>
- Wittenburg, P., Strawn, G., Mons B., Bonino L., Schultes, E. (2019). Digital Objects as Drivers towards Convergence in Data Infrastructures. <https://doi.org/10.23728/b2share.b605d85809ca45679b110719b6c6cb11>
- De Smedt, K., Koureas, D., & Wittenburg, P. (2020). FAIR Digital Objects for Science: From Data Pieces to Actionable Knowledge Units. *Publications*, 8(2), 21. <https://doi.org/10.3390/publications8020021>
- Woodford, C., Treloar, A., Leggott, M., Payne, K., Jones, S., Lopez Albacete, J., Madalli, D., Genova, F., Dharmawardena, K., Chibhira, N., Nyberg Åkerström, W., Macneil, R., Nurnberger, A., Pfeifferberger, H., Tanifuji, M., Zhang, Q., Jones, N., Sesink, L., & Wood-Charlson, E. (2024). The Global Open Research Commons International Model, version 1.1 Model Overview (1.1). Zenodo. <https://doi.org/10.15497/RDA00119>

Jones, S., Leggott, M., Lopez Albacete, J., Pascu, C., Payne, K., Schouppe, M., Treloar, A., & Global Open Research Commons IG. (2023). GORC IG: Typology and Definitions (1.01). Zenodo.

<https://doi.org/10.15497/RDA00087>

Le Franc, Y., Parland-von Essen, J., Bonino, L., Lehväslaiho, H., Coen, G., & Staiger, C. (2020). D2.2 FAIR Semantics: First recommendations (1.0 DRAFT). FAIRsFAIR.

<https://doi.org/10.5281/zenodo.3707985>

Hugo, W., Le Franc, Y., Coen, G., Parland-von Essen, J., & Bonino, L. (2021). D2.5 FAIR Semantics Recommendations Second Iteration (1.0). Zenodo. <https://doi.org/10.5281/zenodo.5362010>

Yann Le Franc, Luiz Bonino, Hanna Koivula, Jessica Parland-von Essen, & Robert Pergl. (2022). D2.8 FAIR Semantics Recommendations Third Iteration (V1.0). Zenodo.

<https://doi.org/10.5281/zenodo.6675295>

Garijo, D.; and Poveda-Villalón, M. (2020) Best Practices for Implementing FAIR Vocabularies and Ontologies on the Web. In Giuseppe Cota, M. D.; and Pozzato, G. L., editor(s), Applications and Practices in Ontology Design, Extraction, and Reasoning. IOS Press, Netherlands, 2020.

<https://doi.org/10.3233/SSW200034>

Cox SJD, Gonzalez-Beltran AN, Magagna B, Marinescu M-C (2021) Ten simple rules for making a vocabulary FAIR. PLoS Comput Biol 17(6): e1009041. <https://doi.org/10.1371/journal.pcbi.1009041>

Xu, F., Juty, N., Goble, C. et al. Features of a FAIR vocabulary. J Biomed Semant 14, 6 (2023).

<https://doi.org/10.1186/s13326-023-00286-8>

Amdouni, E., Bouazzouni, S., & Jonquet, C. (2022). O'FAIRe makes you an offer: metadata-based automatic FAIRness assessment for ontologies and semantic resources. International Journal of Metadata, Semantics and Ontologies, 16(1), 16-46.

FOOPS!: An Ontology Pitfall Scanner for the FAIR Principles. Garijo, D.; Corcho, O.; and Poveda-Villalón, M., 2980. 2021. <http://ceur-ws.org/Vol-2980/paper321.pdf>

Mark Schildhauer, Matthew B. Jones, Shawn Bowers, Joshua Madin, Sergeui Krivov, Deana Pennington, Ferdinando Villa, Benjamin Leinfelder, Christopher Jones, and Margaret O'Brien. 2016. OBOE: the Extensible Observation Ontology, version 1.2. KNB Data Repository.

<https://doi.org/10.5063/F1125R0F>

Di Muri, C., Pulieri, M., Raho, D. et al. Assessing semantic interoperability in environmental sciences: variety of approaches and semantic artefacts. Sci Data 11, 1055 (2024).

<https://doi.org/10.1038/s41597-024-03669-3>

Le Franc, Y., GRAU, N., Juty, N., Mejias, G., Reed, P., Ramezani, P., Poveda-Villalon, M., Garijo, D., Goble, C., & van Horik, R. (2025). D4.5 - Guidelines and methodology to create, document and share mappings and crosswalks (V1.1). Zenodo. <https://doi.org/10.5281/zenodo.15111167>

Juty, N., Le Franc, Y., Goble, C., & Martinková, J. (2025). FAIR-IMPACT Task 4.4 Workshop: Developing a Mapping Process Framework (Version v3). Zenodo.

<https://doi.org/10.5281/zenodo.15310718>

Nicolas Matentzoglou, James P Balhoff, Susan M Bello, Chris Bizon, Matthew Brush, Tiffany J Callahan, Christopher G Chute, William D Duncan, Chris T Evelo, Davera Gabriel, John Graybeal, Alasdair Gray,

Benjamin M Gyori, Melissa Haendel, Henriette Harmse, Nomi L Harris, Ian Harrow, Harshad B Hegde, Amelia L Hoyt, Charles T Hoyt, Dazhi Jiao, Ernesto Jiménez-Ruiz, Simon Jupp, Hyeongsik Kim, Sebastian Koehler, Thomas Liener, Qinqin Long, James Malone, James A McLaughlin, Julie A McMurry, Sierra Moxon, Monica C Munoz-Torres, David Osumi-Sutherland, James A Overton, Bjoern Peters, Tim Putman, Núria Queralt-Rosinach, Kent Shefchek, Harold Solbrig, Anne Thessen, Tania Tudorache, Nicole Vasilevsky, Alex H Wagner, Christopher J Mungall, A Simple Standard for Sharing Ontological Mappings (SSSOM), Database, Volume 2022, 2022, baac035,

<https://doi.org/10.1093/database/baac035>

Gregory, A., Bell, D., Brickley, D., Buttigieg, P. L., Cox, S., Edwards, M., Doug, F., Gonzalez Morales, L. G., Heus, P., Hodson, S., Kanjala, C., Le Franc, Y., Maxwell, L., Molloy, L., Richard, S., Rizzolo, F., Winstanley, P., Wyborn, L., & Burton, A. (2024). WorldFAIR (D2.3) Cross-Domain Interoperability Framework (CDIF) (Report Synthesising Recommendations for Disciplines and Cross-Disciplinary Research Areas) (Version 1). Zenodo. <https://doi.org/10.5281/zenodo.11236871>

Baskauf, S. J.; Webb, C. O. Darwin-SW: Darwin Core-Based Terms for Expressing Biodiversity Data as RDF. SW 2016, 7 (6), 629–643. <https://doi.org/10.3233/SW-150203>.

Darwin Core and RDF/OWL Task Groups. 2021. Darwin Core RDF guide. Biodiversity Information Standards (TDWG). <http://rs.tdwg.org/dwc/terms/guides/rdf/2021-07-15>

Corcho, O., Ekaputra, F.J., Heibi, I. et al. A maturity model for catalogues of semantic artefacts. Sci Data 11, 479 (2024). <https://doi.org/10.1038/s41597-024-03185-4>

Clement Jonquet, Biswanath Dutta, Luiz O. Bonino da Silva Santos, Robert Pergl, Yann Le Franc. Common Minimum Metadata for FAIR Semantic Artefacts. Onto4FAIR 2023 - 2nd Workshop on Ontologies for FAIR and FAIR Ontologies, C. Trojahn; L. O. Bonino da Silva Santos; G. Guizzardi; C. Jonquet, Jul 2023, Sherbrooke, Canada. (hal-04106533v2)

Gonzalez Beltran, A. N., & Wilson, A. (2024). M4.3 - Specification of semantic artefact description. Zenodo. <https://doi.org/10.5281/zenodo.10725304>

Wilson, A., & Jonquet, C. (2024). D4.3 - Specification of shared metadata description of semantic artefacts and their catalogues including common reference API (V1.0 - DRAFT NOT YET APPROVED BY THE EUROPEAN COMMISSION). Zenodo. <https://doi.org/10.5281/zenodo.12579779>

Luiz Olavo Bonino da Silva Santos, Kees Burger, Rajaram Kaliyaperumal, Mark D. Wilkinson; FAIR Data Point: A FAIR-oriented approach for metadata publication. Data Intelligence 2022; https://doi.org/10.1162/dint_a_00160

Basille, M., Body, G., Brotons, L., Gaigr, J., Onen Tarantini, S., Pulieri, M., Rodić, P., Silva del Pozo, M., & Winkler, R. (2024). Concept note on a methodological approach to define priorities for monitoring biodiversity. Biodiversa+. <https://doi.org/10.5281/zenodo.13934670>

6. Appendix

Summary of the generic recommendations for interoperability

Recommendation	Description	Implementation difficulty (1 easy to 5 very challenging)	Reference
Generic Recommendation 1	Adopt the upcoming data model to be proposed by TDWG: DwC-DP. We recommend prioritizing the use of newly available DwC-DP tables over creating custom extensions whenever possible.	2	Page 17
Generic Recommendation 2	Adopt EBVCube for raster data with spatial coverage. For non-spatial datasets investigate other models such as RDF Data Cube as a possible complementary data model.	3	Page 19
Generic Recommendation 3	Use semantic artefacts to align metadata and data to common vocabularies and models.	3	Page 21
Generic Recommendation 4	Semantic artefacts in biodiversity must be FAIR.	2	Page 21
Generic Recommendation 5	Mappings should be shared and exchanged using SSSOM and any proposed extensions from the RDA FAIR Mapping WG.	3	Page 22
Generic Recommendation 6	Mappings whatever their type must be FAIR.	2	Page 22
Generic Recommendation 7	Use JSON-LD as a common format for metadata publication and data structure description.	2	Page 22
Generic Recommendation 8	Provide conceptual models and classifications for biodiversity monitoring aligned with RDFS, SKOS or OWL models and serialised in one of the RDF or OWL-specific formats (e.g. Turtle, RDF/XML, OWL/XML, etc.).	3	Page 23
Generic Recommendation 9	Use EML and DCAT metadata schema to document biodiversity monitoring datasets.	2	Page 24
Generic Recommendation 10	Enrich metadata with semantic annotation leveraging domain specific semantic artefacts.	3	Page 24

An interoperability framework for biodiversity monitoring in Europe

Recommendation	Description	Implementation difficulty (1 easy to 5 very challenging)	Reference
Generic Recommendation 11	Publish metadata schema and associated mappings into a repository providing Globally Unique Persistent and Resolvable Identifiers and access to the metadata schema and the mapping via an API.	3	Page 24
Generic Recommendation 12	Datasets must be identified by a Globally Unique, Persistent and Resolvable Identifier.	2	Page 26
Generic Recommendation 13	Provide an API for programmatic access to data and metadata.	3	Page 26
Generic Recommendation 14	Provide a machine-readable description of the API, accessible with a specific API call.	2	Page 27
Generic Recommendation 15	Provide access to the metadata schema used by the service to structure the information collected from the API.	3	Page 27
Generic Recommendation 16	Whenever possible, describe data access procedures using ODRL as proposed by CDIF.	4	Page 28

Summary of the practical recommendations for interoperability

Recommendation	Description	Implementation difficulty (1 easy to 5 very challenging)	Reference
Practical Recommendation 1	Build a converter from EML to JSON-LD that can be used by the whole biodiversity monitoring community.	1	Page 30
Practical Recommendation 2	Use EML semantic annotations to enrich the metadata with explicit semantics provided by semantic artefacts.	3	Page 30
Practical Recommendation 3	Provide the DwC Conceptual Model as a semantic artefact.	3	Page 30

Recommendation	Description	Implementation difficulty (1 easy to 5 very challenging)	Reference
Practical Recommendation 4	Publish biodiversity monitoring semantic artefacts in a Semantic Artefact Catalogues such as EcoPortal, provided by LifeWatch ERIC, or BiodivPortal, provided by GFBio and NFDI4Biodiversity.	1	Page 31
Practical Recommendation 5	Provide metadata to document semantic artefacts using the MOD Ontology.	3	Page 32
Practical Recommendation 6	Mappings should be published into a mapping repository such as the Metadata Schema and Crosswalk Registry (MSCR).	3	Page 32
Practical Recommendation 7	Convert datapackage.json into JSON-LD and publish the description of the data structure.	2	Page 34
Practical Recommendation 8	Publish DDI-CDI data description to ease complex data integration and automation.	4	Page 36