biodiversa+
European Biodiversity Partnership

Making Research Data FAIR: An overview of (Meta)data Harmonisation by

Senem Önen Tarantini &
Martina Pulieri

Capacity Building Workshop on Darwin Core
07-06-2024

FAIRness in biodiversity concept

- **Findable:** Detailed metadata; descriptive information about the data and registering datasets in public repositories with search functionalities. Such as species base, area base, habitat base etc.
  That facilitates findability of the data both human and computers.

- **Accessible**: Using open data formats whenever possible, and providing clear instructions on how to access and download the data . Moreover data should be stored -long term- so it could be easily accessed

- **Interoperable**: Biodiversity data comes from diverse sources, using different methods and formats. Interoperability ensures that data from various studies can be integrated and compared. Using standardized vocabularies, units, and data structures guaranty interoperability.

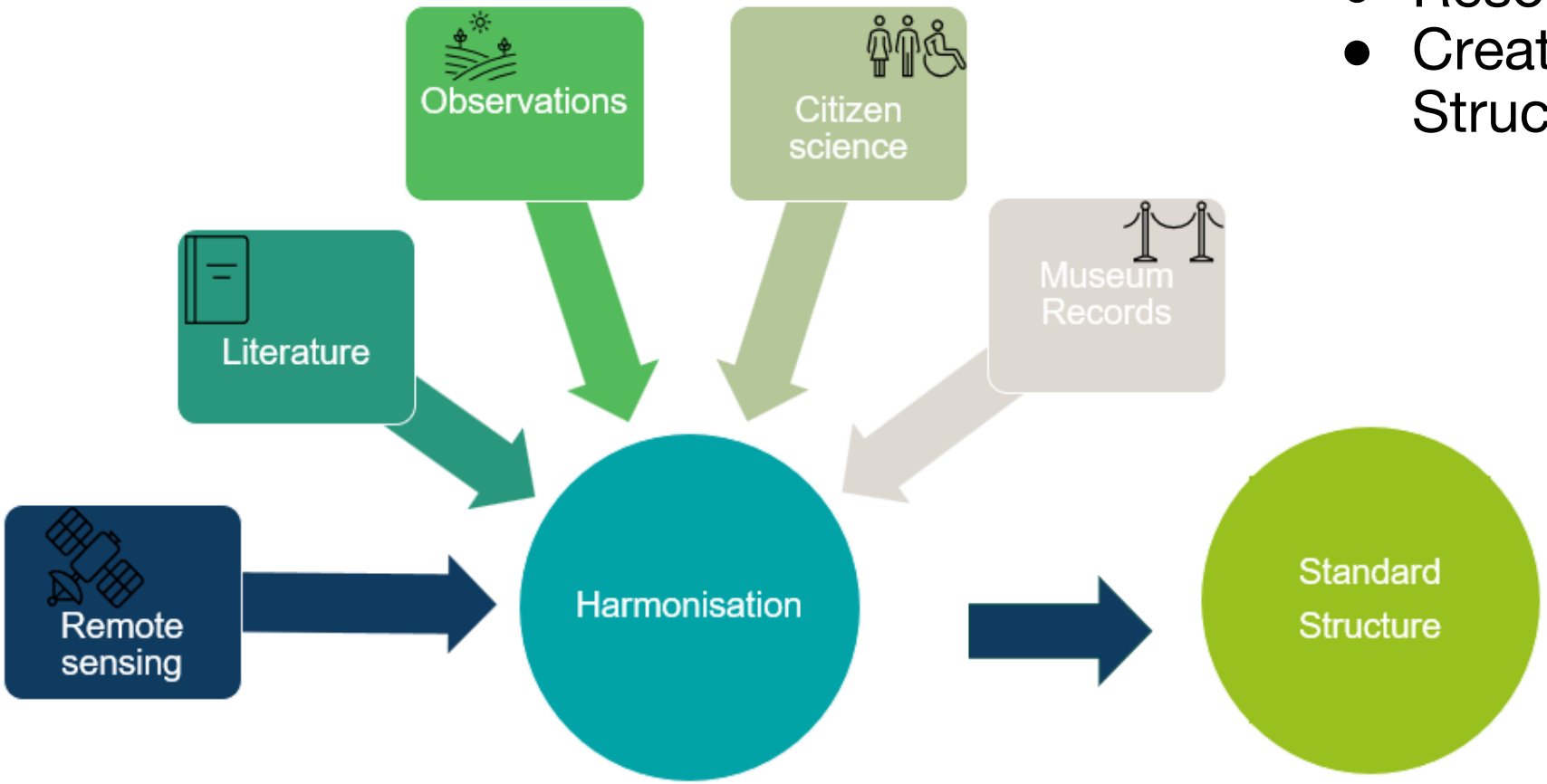
- **Reusable:** Data should be usable in new research. This means providing comprehensive documentation about the data collection process, using standard methods, units and also having appropriate licence .

(Meta)data harmonisation

- **Data harmonisation** is the process of bringing together data from various sources and transforming them into a consistent format.

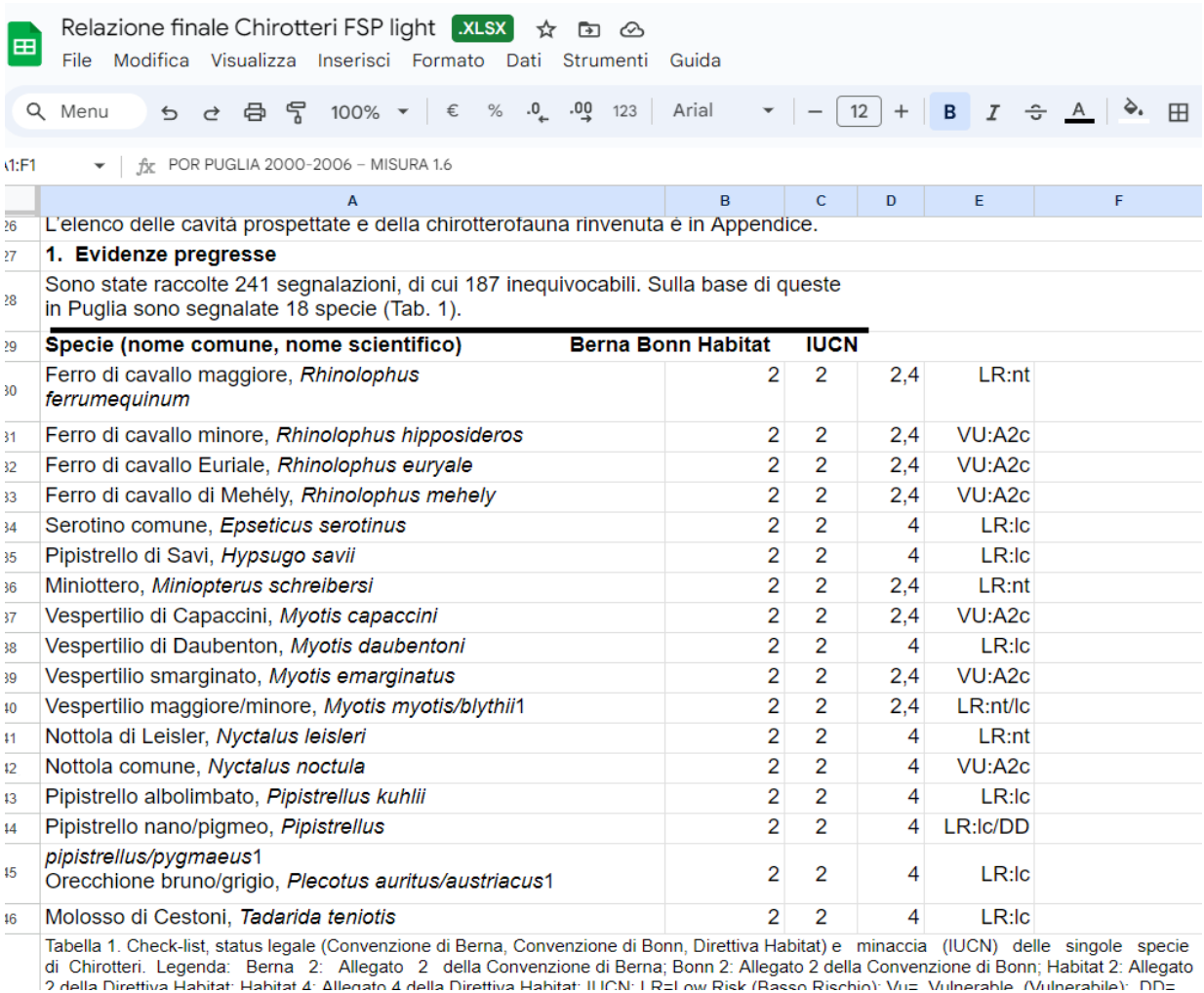
- **Metadata harmonisation** ensures the descriptive information (metadata) about the data is consistent across different sources. The aim of metadata harmonisation is creating a universal instruction manual for biodiversity data.

Data harmonisation



- Unifying formats
- Resolving incoherency
- Creating a Standardize Structure

Data harmonisation: Creating a standard structure



Relazione finale Chirotteri FSP light .XLSX

File Modifica Visualizza Inserisci Formato Dati Strumenti Guida

Menu 100% € % .0 .00 123 Arial 12 B I A

A1:F1 POR PUGLIA 2000-2006 – MISURA 1.6

L'elenco delle cavità prospettate e della chirotterofauna rinvenuta è in Appendice.

1. Evidenze pregresse

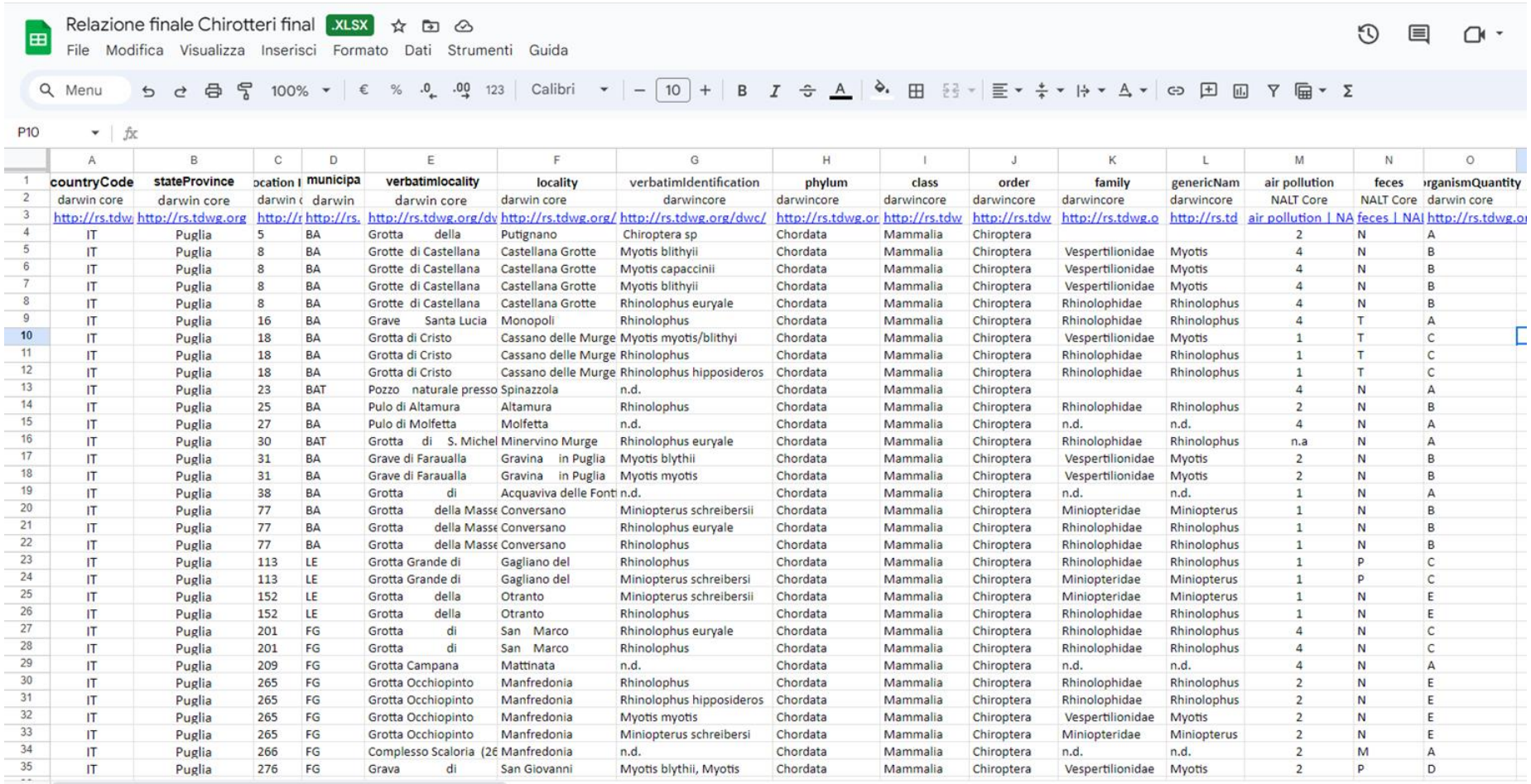
Sono state raccolte 241 segnalazioni, di cui 187 inequivocabili. Sulla base di queste in Puglia sono segnalate 18 specie (Tab. 1).

Specie (nome comune, nome scientifico)	Berna	Bonn	Habitat	IUCN
Ferro di cavallo maggiore, *Rhinolophus ferrumequinum*	2	2	2,4	LR:nt
Ferro di cavallo minore, *Rhinolophus hipposideros*	2	2	2,4	VU:A2c
Ferro di cavallo Euriale, *Rhinolophus euryale*	2	2	2,4	VU:A2c
Ferro di cavallo di Mehély, *Rhinolophus mehely*	2	2	2,4	VU:A2c
Serotino comune, *Epseticus serotinus*	2	2	4	LR:lc
Pipistrello di Savi, *Hypsugo savii*	2	2	4	LR:lc
Miniottero, *Miniopterus schreibersi*	2	2	2,4	LR:nt
Vespertilio di Capaccini, *Myotis capaccini*	2	2	2,4	VU:A2c
Vespertilio di Daubenton, *Myotis daubentoni*	2	2	4	LR:lc
Vespertilio smarginato, *Myotis emarginatus*	2	2	2,4	VU:A2c
Vespertilio maggiore/minore, *Myotis myotis/blythii*1	2	2	2,4	LR:nt/lc
Nottola di Leisler, *Nyctalus leisleri*	2	2	4	LR:nt
Nottola comune, *Nyctalus noctula*	2	2	4	VU:A2c
Pipistrello albolimbato, *Pipistrellus kuhlii*	2	2	4	LR:lc
Pipistrello nano/pigmeo, *Pipistrellus*	2	2	4	LR:lc/DD
*pipistrellus/pygmaeus*1 Orecchione bruno/grigio, *Plecotus auritus/austriacus*1	2	2	4	LR:lc
Molosso di Cestoni, *Tadarida teniotis*	2	2	4	LR:lc

Tabella 1. Check-list, status legale (Convenzione di Berna, Convenzione di Bonn, Direttiva Habitat) e minaccia (IUCN) delle singole specie di Chirotteri. Legenda: Berna 2: Allegato 2 della Convenzione di Berna; Bonn 2: Allegato 2 della Convenzione di Bonn; Habitat 2: Allegato 2 della Direttiva Habitat; Habitat 4: Allegato 4 della Direttiva Habitat; IUCN: LR=Low Risk (Basso Rischio); Vu= Vulnerable (Vulnerabile); DD=

- In Italian
- Limited geographical information
- Sampling protocol within pdf text
- Limited taxonomic information

Data harmonisation: Creating a standard structure



Relazione finale Chirotteri final .XLSX

File Modifica Visualizza Inserisci Formato Dati Strumenti Guida

	countryCode	stateProvince	ocation I	municipa	verbatimlocality	locality	verbatimIdentification	phylum	class	order	family	genericNam	air pollution	feces	rganismQuantity
2	darwin core	darwin core	darwin (	darwin	darwin core	darwin core	darwincore	darwincore	darwincore	darwincore	darwincore	darwincore	NALT Core	NALT Core	darwin core
3	http://rs.tdw	http://rs.tdwg.org	http://r	http://rs.	http://rs.tdwg.org/dv	http://rs.tdwg.org/	http://rs.tdwg.org/dwc/	http://rs.tdwg.or	http://rs.tdw	http://rs.tdw	http://rs.tdwg.o	http://rs.td	air pollution I NA	feces I NAI	http://rs.tdwg.o
4	IT	Puglia	5	BA	Grotta della	Putignano	Chiroptera sp	Chordata	Mammalia	Chiroptera			2	N	A
5	IT	Puglia	8	BA	Grotte di Castellana	Castellana Grotte	Myotis blithyii	Chordata	Mammalia	Chiroptera	Vespertilionidae	Myotis	4	N	B
6	IT	Puglia	8	BA	Grotte di Castellana	Castellana Grotte	Myotis capaccinii	Chordata	Mammalia	Chiroptera	Vespertilionidae	Myotis	4	N	B
7	IT	Puglia	8	BA	Grotte di Castellana	Castellana Grotte	Myotis blithyii	Chordata	Mammalia	Chiroptera	Vespertilionidae	Myotis	4	N	B
8	IT	Puglia	8	BA	Grotte di Castellana	Castellana Grotte	Rhinolophus euryale	Chordata	Mammalia	Chiroptera	Rhinolophidae	Rhinolophus	4	N	B
9	IT	Puglia	16	BA	Grave Santa Lucia	Monopoli	Rhinolophus	Chordata	Mammalia	Chiroptera	Rhinolophidae	Rhinolophus	4	T	A
10	IT	Puglia	18	BA	Grotta di Cristo	Cassano delle Murge	Myotis myotis/blithyi	Chordata	Mammalia	Chiroptera	Vespertilionidae	Myotis	1	T	C
11	IT	Puglia	18	BA	Grotta di Cristo	Cassano delle Murge	Rhinolophus	Chordata	Mammalia	Chiroptera	Rhinolophidae	Rhinolophus	1	T	C
12	IT	Puglia	18	BA	Grotta di Cristo	Cassano delle Murge	Rhinolophus hipposideros	Chordata	Mammalia	Chiroptera	Rhinolophidae	Rhinolophus	1	T	C
13	IT	Puglia	23	BAT	Pozzo naturale presso	Spinazzola	n.d.	Chordata	Mammalia	Chiroptera			4	N	A
14	IT	Puglia	25	BA	Pulo di Altamura	Altamura	Rhinolophus	Chordata	Mammalia	Chiroptera	Rhinolophidae	Rhinolophus	2	N	B
15	IT	Puglia	27	BA	Pulo di Molfetta	Molfetta	n.d.	Chordata	Mammalia	Chiroptera	n.d.	n.d.	4	N	A
16	IT	Puglia	30	BAT	Grotta di S. Michel	Minervino Murge	Rhinolophus euryale	Chordata	Mammalia	Chiroptera	Rhinolophidae	Rhinolophus	n.a	N	A
17	IT	Puglia	31	BA	Grave di Faraualla	Gravina in Puglia	Myotis blythii	Chordata	Mammalia	Chiroptera	Vespertilionidae	Myotis	2	N	B
18	IT	Puglia	31	BA	Grave di Faraualla	Gravina in Puglia	Myotis myotis	Chordata	Mammalia	Chiroptera	Vespertilionidae	Myotis	2	N	B
19	IT	Puglia	38	BA	Grotta di	Acquaviva delle Fonti	n.d.	Chordata	Mammalia	Chiroptera	n.d.	n.d.	1	N	A
20	IT	Puglia	77	BA	Grotta della Masse	Conversano	Miniopterus schreibersii	Chordata	Mammalia	Chiroptera	Miniopteridae	Miniopterus	1	N	B
21	IT	Puglia	77	BA	Grotta della Masse	Conversano	Rhinolophus euryale	Chordata	Mammalia	Chiroptera	Rhinolophidae	Rhinolophus	1	N	B
22	IT	Puglia	77	BA	Grotta della Masse	Conversano	Rhinolophus	Chordata	Mammalia	Chiroptera	Rhinolophidae	Rhinolophus	1	N	B
23	IT	Puglia	113	LE	Grotta Grande di	Gagliano del	Rhinolophus	Chordata	Mammalia	Chiroptera	Rhinolophidae	Rhinolophus	1	P	C
24	IT	Puglia	113	LE	Grotta Grande di	Gagliano del	Miniopterus schreibersi	Chordata	Mammalia	Chiroptera	Miniopteridae	Miniopterus	1	P	C
25	IT	Puglia	152	LE	Grotta della	Otranto	Miniopterus schreibersii	Chordata	Mammalia	Chiroptera	Miniopteridae	Miniopterus	1	N	E
26	IT	Puglia	152	LE	Grotta della	Otranto	Rhinolophus	Chordata	Mammalia	Chiroptera	Rhinolophidae	Rhinolophus	1	N	E
27	IT	Puglia	201	FG	Grotta di	San Marco	Rhinolophus euryale	Chordata	Mammalia	Chiroptera	Rhinolophidae	Rhinolophus	4	N	C
28	IT	Puglia	201	FG	Grotta di	San Marco	Rhinolophus	Chordata	Mammalia	Chiroptera	Rhinolophidae	Rhinolophus	4	N	C
29	IT	Puglia	209	FG	Grotta Campana	Mattinata	n.d.	Chordata	Mammalia	Chiroptera	n.d.	n.d.	4	N	A
30	IT	Puglia	265	FG	Grotta Occhiopinto	Manfredonia	Rhinolophus	Chordata	Mammalia	Chiroptera	Rhinolophidae	Rhinolophus	2	N	E
31	IT	Puglia	265	FG	Grotta Occhiopinto	Manfredonia	Rhinolophus hipposideros	Chordata	Mammalia	Chiroptera	Rhinolophidae	Rhinolophus	2	N	E
32	IT	Puglia	265	FG	Grotta Occhiopinto	Manfredonia	Myotis myotis	Chordata	Mammalia	Chiroptera	Vespertilionidae	Myotis	2	N	E
33	IT	Puglia	265	FG	Grotta Occhiopinto	Manfredonia	Miniopterus schreibersi	Chordata	Mammalia	Chiroptera	Miniopteridae	Miniopterus	2	N	E
34	IT	Puglia	266	FG	Complesso Scaloria (26	Manfredonia	n.d.	Chordata	Mammalia	Chiroptera	n.d.	n.d.	2	M	A
35	IT	Puglia	276	FG	Grava di	San Giovanni	Myotis blythii, Myotis	Chordata	Mammalia	Chiroptera	Vespertilionidae	Myotis	2	P	D

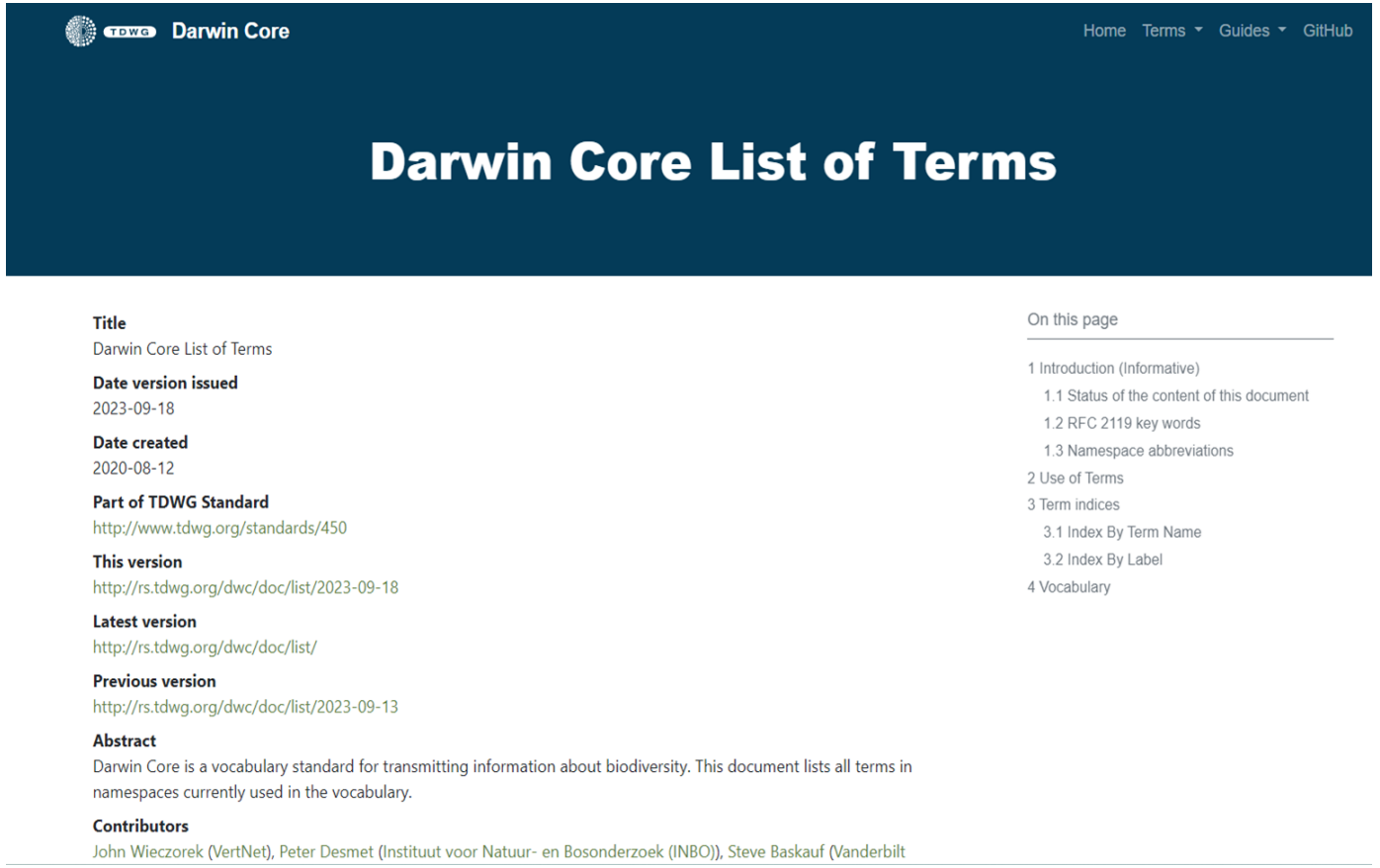
- In English
- Detailed geographical information
- Detailed taxonomic information
- Detailed sampling findings

biodiversa+
European Biodiversity Partnership

How (meta)data harmonisation works in biodiversity research

Data harmonisation focuses on bringing the data itself into a consistent format, while metadata harmonisation specifically tackles ensuring the descriptive information (metadata) about the data is consistent across different sources .

- Using standardised vocabularies: Darwin Core (DwC) is a widely used vocabulary for sharing information on biodiversity.
- Standard meta(data) structure
- Shared definition for describing the (meta)data



biodiversa+
European Biodiversity Partnership

https://dwc.tdwg.org/terms/

Metadata harmonisation: Creating a standard structure

End Date	A single time stamp indicating the end of a time period.	1			
Taxonomic Coverage		0..∞		Yes	
Taxon Rank Name	The name of the taxonomic rank for which the Taxon rank value is provided (e.g., kingdom, class, order, family, genus, etc.).	Phylum, Class, Order ,Family , Genus , Species			
Taxon Rank Value	The taxonomic rank name being described (e.g., Acer).	1			
Common Name	Specification of applicable common names (e.g., insects, vertebrate, grasses, etc.).	Reptilia	Amphibia		
Taxon ID	The identifier for this taxon from an authority, such as ITIS (https://www.itis.gov) or USDA Plant Database (https://plants.usda.gov).	https://www.itis.gov/servlet/SingleRpt/SingleRpt?search_topic=TSN&search_value=173747#null	https://www.itis.gov/servlet/SingleRpt/SingleRpt#null		
License Information					
Intellectual Rights	Intellectual property rights regarding usage and licensing of the resource. E.g., This data package is released to the "public domain" under Creative Commons CC0 1.0 "No Rights Reserved" (see: https://creativecommons.org/publicdomain/zero/1.0/).	LifeWatch Metadata			
License Name	The official name of the license applied to the data and metadata described in this metadata record. The name should match the name of a well-known license from the SPDX license vocabulary or a similar persistent vocabulary.	1	Yes		
		02-3888-7900			

biodiversa+
European Biodiversity Partnership

(Meta)data harmonisation and FAIR (interoperability)

Interoperable data means it can be integrated with different datasets, application and workflows.

It has different layers as

- **Organisational**: Governance, policies
- **Legal**: Licences, copyright, data protection
- **Technical**: Infrastructure, services, technologies
- **Semantic**: Community standards, ontologies & metadata

Towards interoperability

Data interoperability

Implies **semantic interoperability**, which ensures that the *precise format and meaning of exchanged data and information* is preserved and understood throughout exchanges between parties.

https://joinup.ec.europa.eu/collection/nifo-national-interoperability-framework-observatory/glossary/term/semantic-interoperability

Semantic interoperability: problems

- Lack of common explicit definitions

Among different vocabularies there are different definitions about the same concepts.

- Lack of common semantic artefacts across communities

More than 540 semantic artefacts were found with numerous overlapping in topics. A great number were created *ad hoc* for projects and never “connected” to the existing semantic artefacts, impeding interoperability.

- Lack of expertise and skills related to semantics

Negatively influences the use of common definitions.

https://data.europa.eu/doi/10.2777/620649

Data and semantic interoperability: recommendations

- Clear and precise definitions for the concepts

Use of commonly agreed vocabularies to allow an easier data exchange and integration.

- Thoroughly document your data

What, Who, Where, When, How to not lose information and allow the reuse of data.

https://data.europa.eu/doi/10.2777/620649

biodiversa+
European Biodiversity Partnership

EUROPEAN PARTNERSHIP

Co-funded by the European Union

Thank you!

www.biodiversa.eu

contact@biodiversa.eu

BiodiversaPlus

NFDI 4 BIODIVERSITY

Semantic technologies and their usage for biodiversity data: opportunities and challenges

Naouel Karam

Institute for Applied Informatics (InfAI), Leipzig University

Biodiversa+ Capacity building workshop on Darwin Core standard

7th of June 2024

Agenda

NFDI 4 BIODIVERSITY

Why **... do we need Semantic technologies?**

- Ambiguity of Natural Language
- Keywords vs Concepts

What **... are terminologies? ... terminologies do we need?**

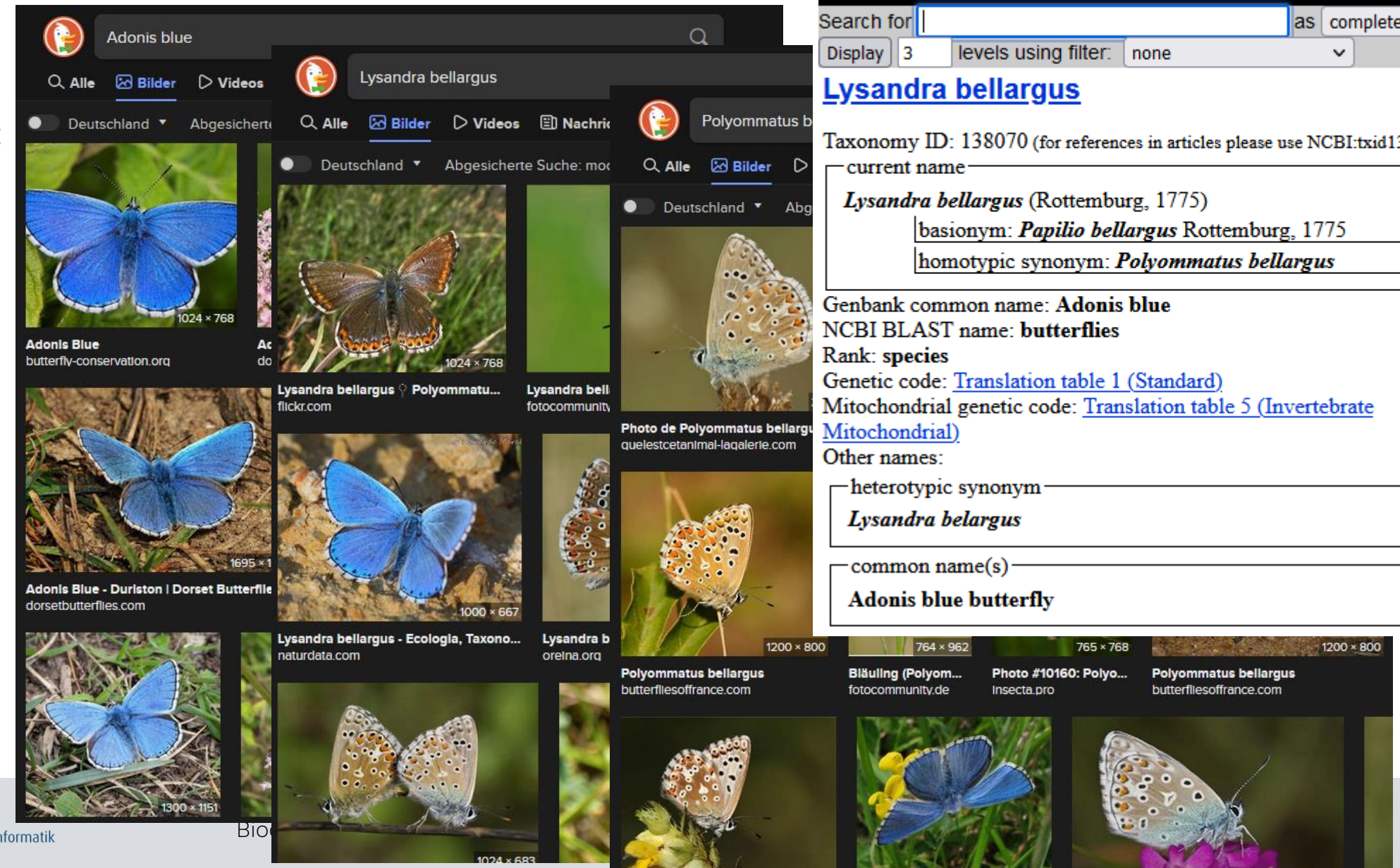
- Terminology?
- Terminologies for Semantic Search
- Terminologies for Metadata Harmonization
- Terminologies for Data Interoperability

How **... do we use terminologies?**

- Semantic Services for Biodiversity
- Requirements on a Terminology Service
- Ongoing Work

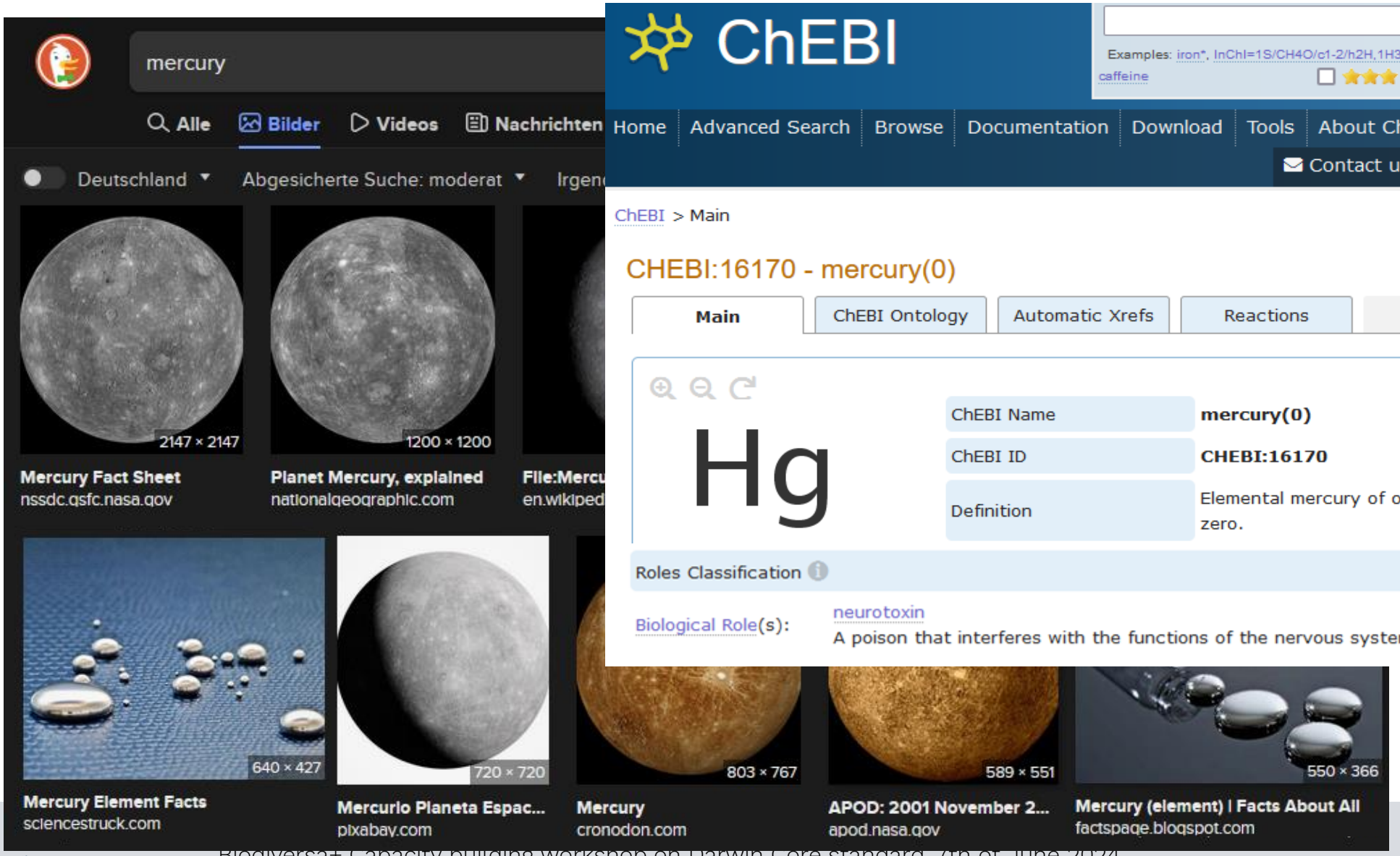
Why Ambiguity of Natural Langua

> Different words for the same concept (Synonymy)

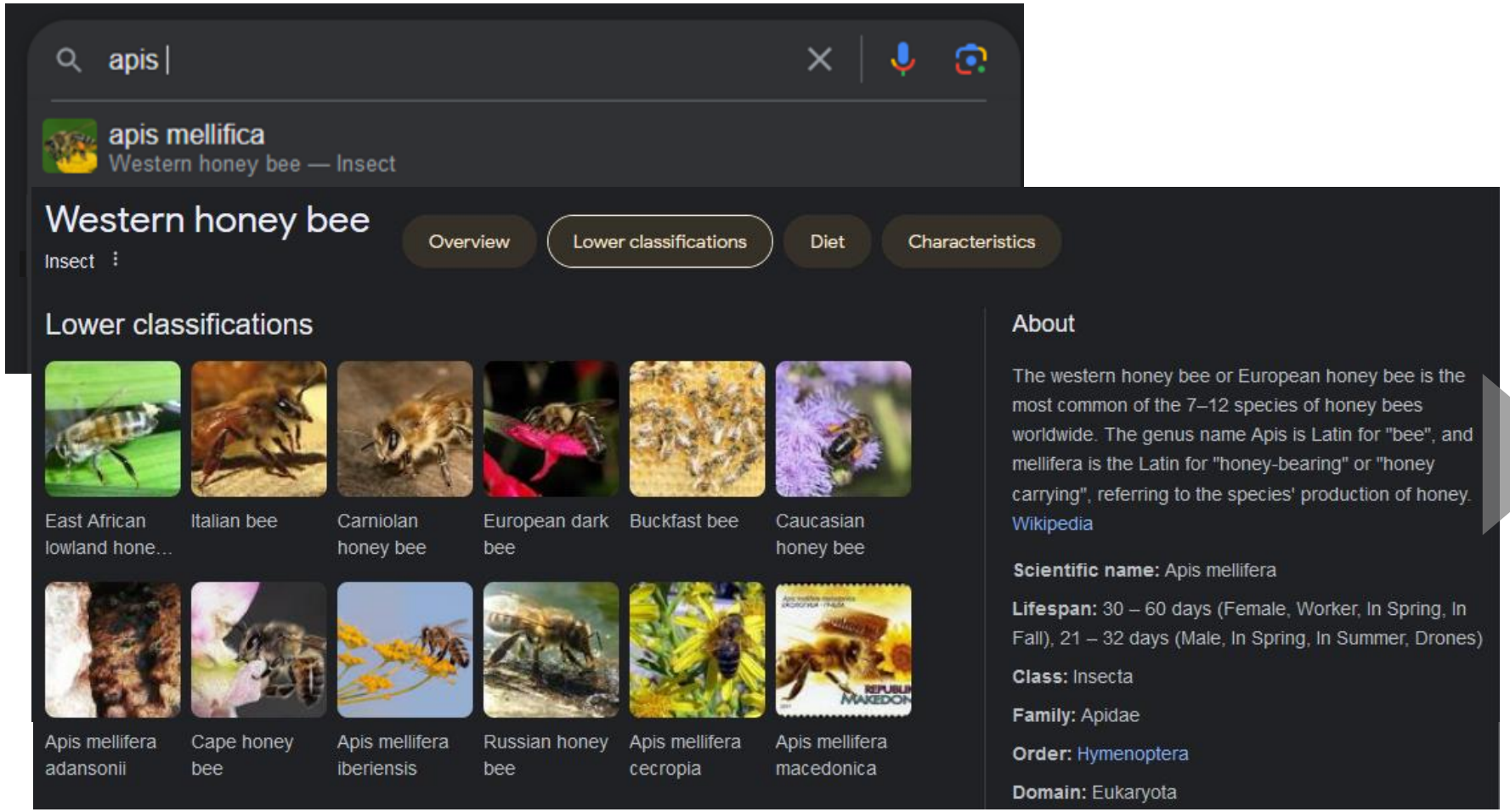


Why Ambiguity of Natural Language

> Same word for different concepts (**Polysemy**)



Why Keywords vs Concepts



Meaning (**Semantics**) is expressed with the help of knowledge representation (**Terminologies**)

What is a Terminology?

NFDI 4 BIODIVERSITY

informal ← → formal

- Controlled Vocabulary: List of terms
- Glossary: List of terms + **informal definitions** of their meaning
- Taxonomy: List of terms organised in a **hierarchical** structure
- Thesaurus: Controlled vocabulary connected via **relations** between terms (e.g. "narrower/broader", "related term", "synonym")
- Ontology: Formal representation of a set of concepts and the relationships between them using **logical axioms**

gfbio RecordBasis

- **Absence Observation**
- DrawingOrPhotograph
- Earth Science Specimen
- Fossil
- Human Observation
- Literature
- Living Specimen
- Machine Observation
- MaterialSample
- Mineral Specimen
- MultimediaObject
- OtherSpecimen
- Preserved Specimen
- Unknown

AMMOD IOC Bird List

- Aves
  - ACCIPITRIFORMES
  - OTIDIFORMES
  - PASSERIFORMES
    - Acanthisittidae
    - Acanthizidae
    - Thraupidae
      - Acanthidops
      - Anisognathus
      - Bangsia
    - Nesospiza
      - Nesospiza acunhae
      - **Nesospiza questi**
      - Nesospiza wilkinsi

EuroVoc

- physical environment
  - aquatic environment
  - atmosphere
  - biosphere
    - **biodiversity**

broader	biosphere
related	animal resources
	plant resources
	protected species

ChEBI

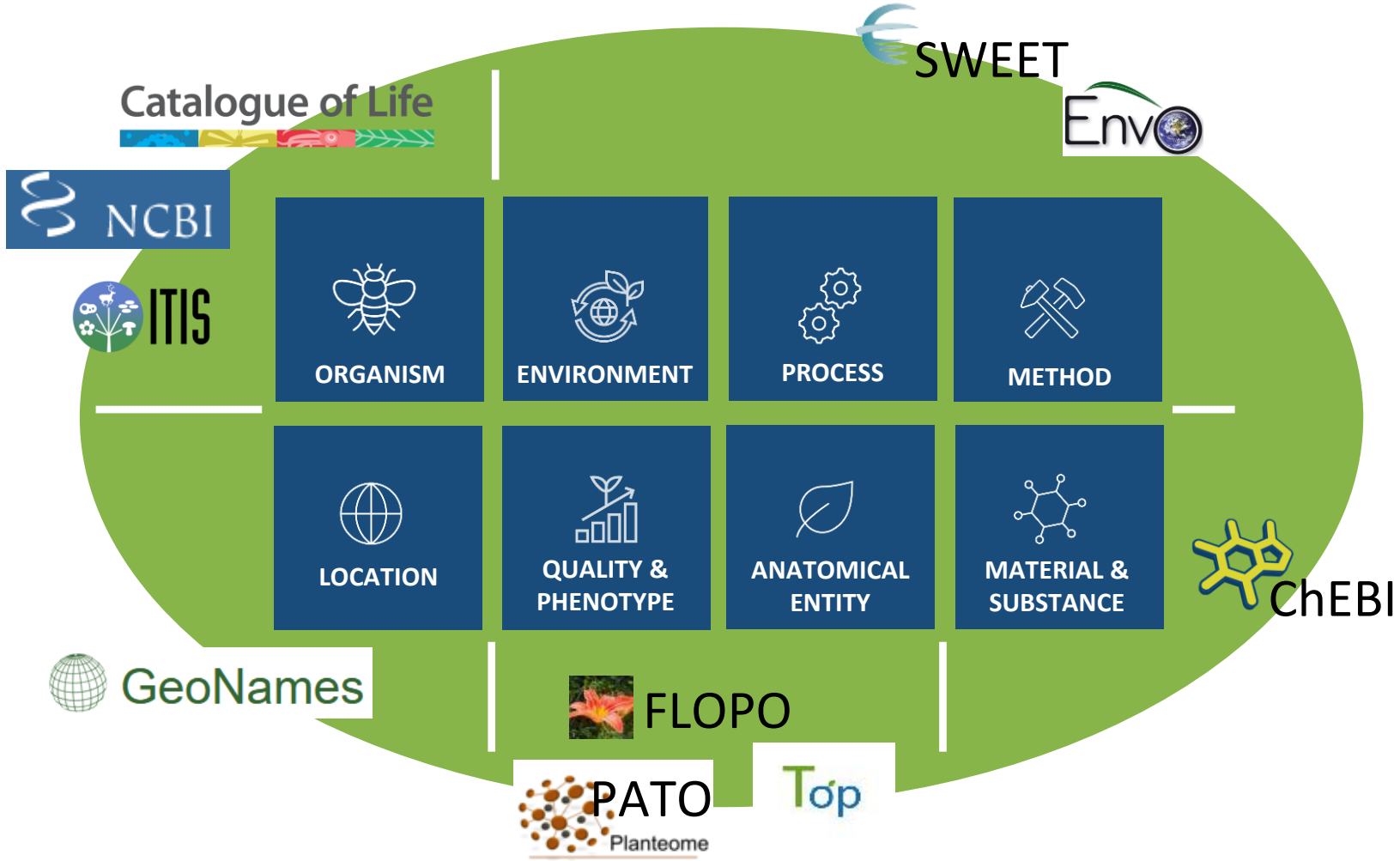
- mercury molecular entity
  - elemental mercury
    - mercury cation
    - **mercury(0)**

Preferred Name	mercury(0)
Synonyms	mercury Hg Hgn mercury(0) Elemental mercury metallic mercury Mercury Hg(0)
monoisotopicmass	201.97064
subClassOf	elemental mercury
	has role *some* neurotoxin

What Terminologies for Semantic Search

NFDI 4 BIODIVERSITY

Do butterflies occur in calcareous grassland?
How does agriculture affect the ground water composition?
Is there data on the influence of geographic elevation on the growth rate of Zea mays? ….



F. Löffler, C. Pfaff, N. Karam, D. Fichtmüller, F. Klan: What do Biodiversity Scholars Search for? Identifying High-Level Entities for Biological Metadata. S4BioDiv@ISWC 2017

What Terminologies for Metadata Harmonization

NFDI 4 BIODIVERSITY

ABCD Access to Biological Collection Data

ABCD OWL

Schema.org OWL


```
    ],
    "http://rs.tdwg.org/abcd/terms/name": [
      {
        "@value": "Botanic Garden and Botanical Museum Berlin-Dahlem"
      }
    ],
    "http://rs.tdwg.org/abcd/terms/abbreviation": [
      {
        "@value": "BGBM"
      }
    ],
    "@type": [
      "http://rs.tdwg.org/abcd/terms/Organization"
    ]
  },
  {
    "@id": "http://www.bgbm.org/Herbarium_Berolinense",
    "http://rs.tdwg.org/abcd/terms/hasContact": [
      {
        "@id": "http://www.bgbm.org"
      }
    ],
    "http://rs.tdwg.org/abcd/terms/details": [
      {
        "@value": "The herbarium of the Botanic Garden and Botanical Museum Berlin-Dahlem (
collection of more than 3.5 million preserved specimens. All plant groups – flowering plant
and lichens – are represented in the collections which are worldwide in scope. Associated w
fruits and seeds, wood samples, and specimens preserved in alcohol. The collections of the
conducted by staff, and through gifts, acquisitions, and exchanges of specimens from other
      }
    ],
    "@type": [
      "http://rs.tdwg.org/abcd/terms/DataSet"
    ],
    "http://rs.tdwg.org/abcd/terms/title": [
      {
        "@value": "Herbarium Berolinense"
      }
```



```
{
  "@context": "http://schema.org/",
  "@id": "https://doi.org/10.1594/PANGAEA.718130",
  "@type": "Dataset",
  "identifier": "https://doi.org/10.1594/PANGAEA.718130",
  "url": "https://doi.pangaea.de/10.1594/PANGAEA.718130",
  "creator": [
    {
      "@id": "https://orcid.org/0000-0001-8807-4597",
      "@type": "Person",
      "name": "Frédéric Gazeau",
      "familyName": "Gazeau",
      "givenName": "Frédéric",
      "identifier": "https://orcid.org/0000-0001-8807-4597",
      "email": "gazeau@obs-vlfr.fr"
    },
    {
      "@type": "Person",
      "name": "Christophe Quiblier",
      "familyName": "Quiblier",
      "givenName": "Christophe"
    },
    {
      "@type": "Person",
      "name": "Jeroen M Jansen",
      "familyName": "Jansen",
      "givenName": "Jeroen M"
    },
    {
```


abstract	Text	An abstract is a short description that summarizes a CreativeWork.
accessMode	Text	The human sensory perceptual system or cognitive faculty through which a person may process or perceive information. Values should be drawn from the approved vocabulary.

What Terminologies for Data Integration

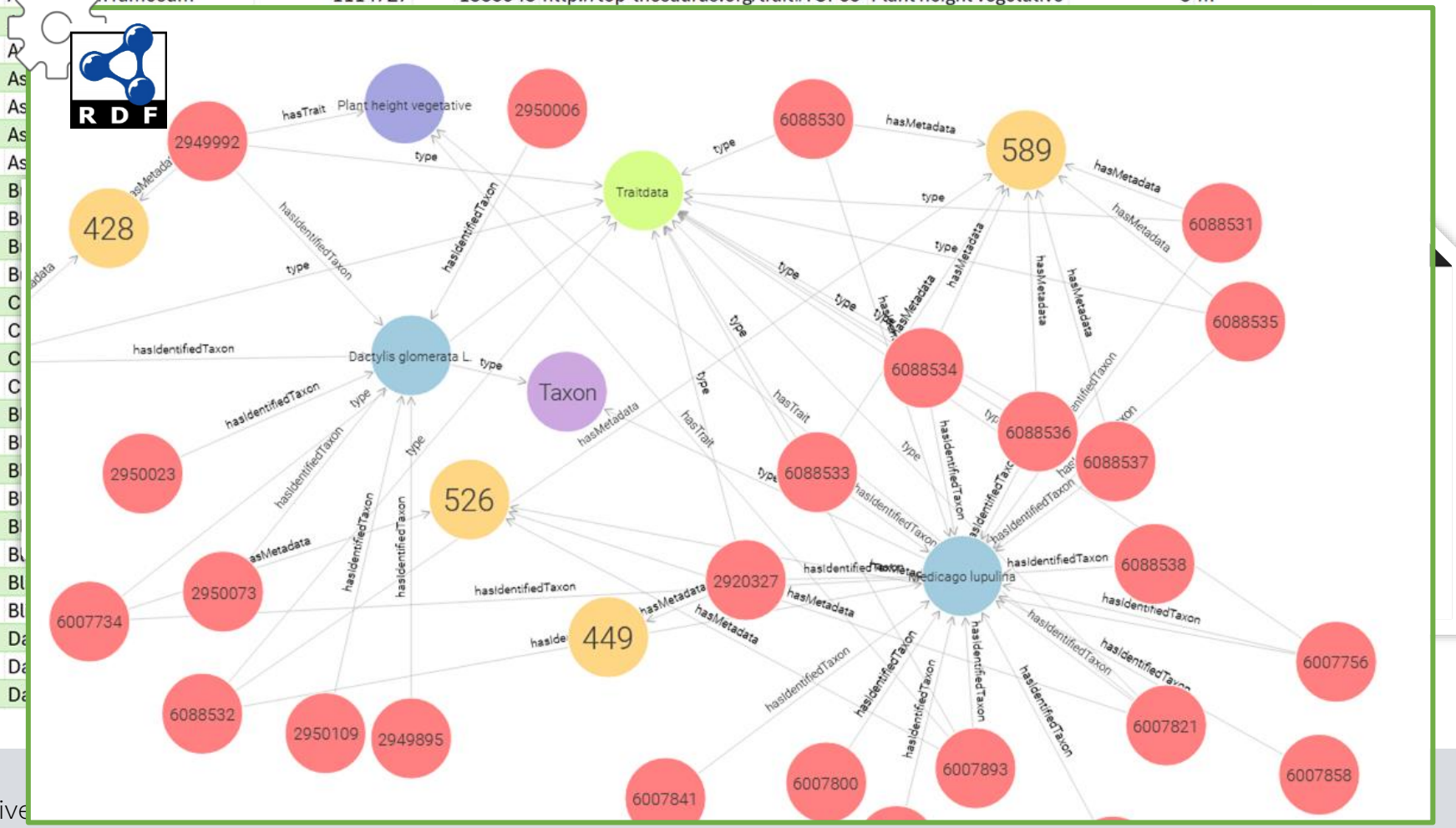
NFDI 4 BIODIVERSITY

Catalogue of Life

Top

QUDT

LastNam	FirstNa	DatasetID	Dataset	AccSpeciesName	ObservationID	ObsDataID	TraitID	TraitName	StdValue	UnitName
Kleyer	Michael	25	The LEDA Traitbase	Anthericum ramosum	173636	1835943	http://top-thesaurus.org/trait#TOP69	Plant height vegetative	3	m
Kleyer	Michael	25	The LEDA Traitbase	A m ramosum	1114727	1835948	http://top-thesaurus.org/trait#TOP69	Plant height vegetative	3	m
Kleyer	Michael	25	The LEDA Traitbase							
Kleyer	Michael	25	The LEDA Traitbase	A						
Kleyer	Michael	25	The LEDA Traitbase	As						
Kleyer	Michael	25	The LEDA Traitbase	As						
Kleyer	Michael	25	The LEDA Traitbase	As						
Kleyer	Michael	25	The LEDA Traitbase	As						
Kleyer	Michael	25	The LEDA Traitbase	B						
Kleyer	Michael	25	The LEDA Traitbase	B						
Kleyer	Michael	25	The LEDA Traitbase	B						
Kleyer	Michael	25	The LEDA Traitbase	B						
Kleyer	Michael	25	The LEDA Traitbase	C						
Kleyer	Michael	25	The LEDA Traitbase	C						
Kleyer	Michael	25	The LEDA Traitbase	C						
Kleyer	Michael	25	The LEDA Traitbase	C						
Kleyer	Michael	25	The LEDA Traitbase	B						
Kleyer	Michael	25	The LEDA Traitbase	B						
Kleyer	Michael	25	The LEDA Traitbase	B						
Kleyer	Michael	25	The LEDA Traitbase	B						
Kleyer	Michael	25	The LEDA Traitbase	B						
Kleyer	Michael	25	The LEDA Traitbase	B						
Kleyer	Michael	25	The LEDA Traitbase	Bl						
Kleyer	Michael	25	The LEDA Traitbase	Bl						
Kleyer	Michael	25	The LEDA Traitbase	se						
Kleyer	Michael	25	The LEDA Traitbase	se						
Kleyer	Michael	25	The LEDA Traitbase	se						



TRY Plant Trait Database

How Semantic Services for Biodiversity

NFDI 4 BIODIVERSITY

Goal: Development of **semantic standards**, **services** and **tools** for the discovery, integration, transformation and harmonization of heterogeneous data

2014 2015 2020 2022 2023

gfbio GERMAN FEDERATION FOR BIOLOGICAL DATA

gfbio Terminology Service

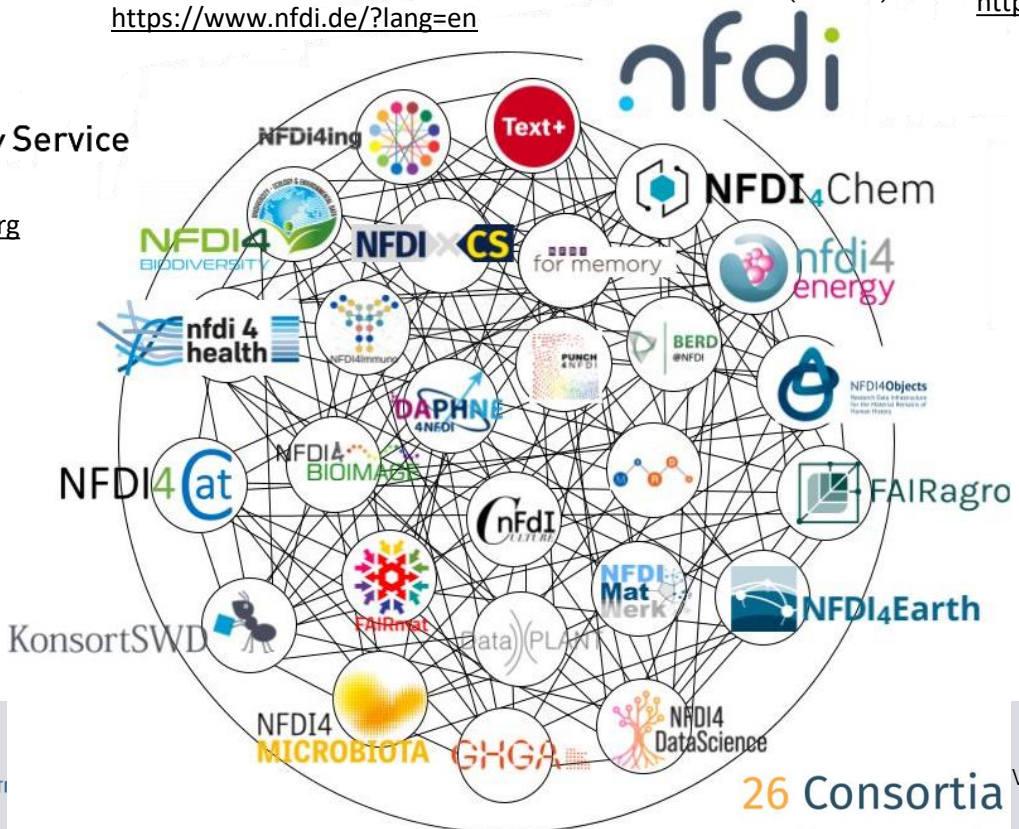
Up and running

https://terminologies.gfbio.org

NFDI 4 BIODIVERSITY

National Research Data Infrastructure (NFDI)

https://www.nfdi.de/?lang=en



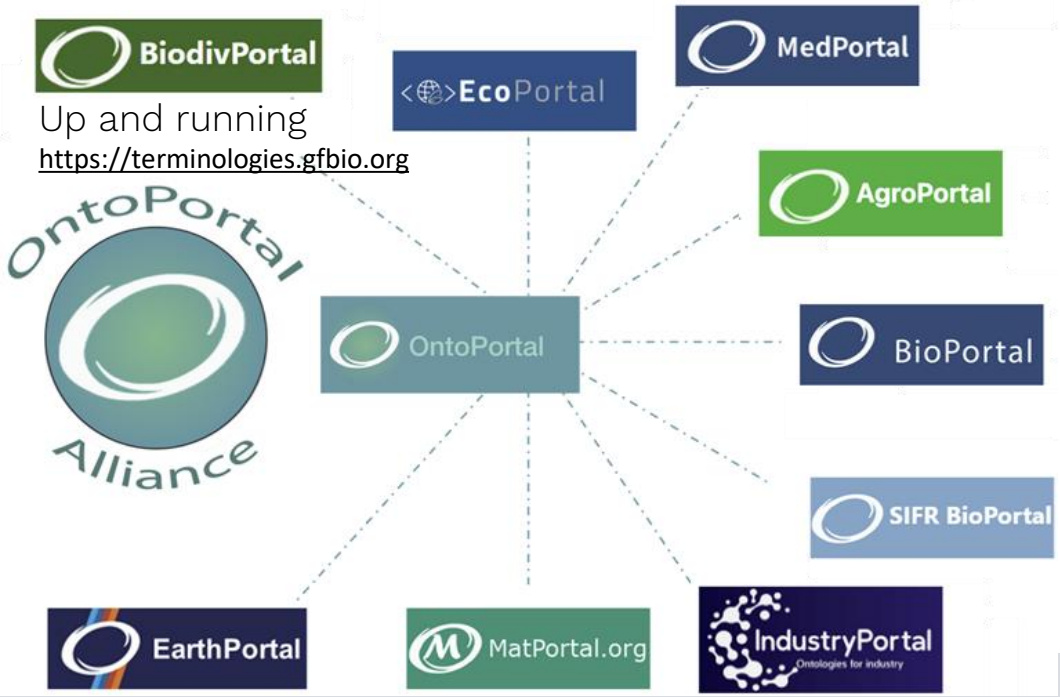
26 Consortia

Joined the OntoPortal Alliance

https://ontoportal.org/about/

Up and running

https://terminologies.gfbio.org



How Requirements on a Terminology Service

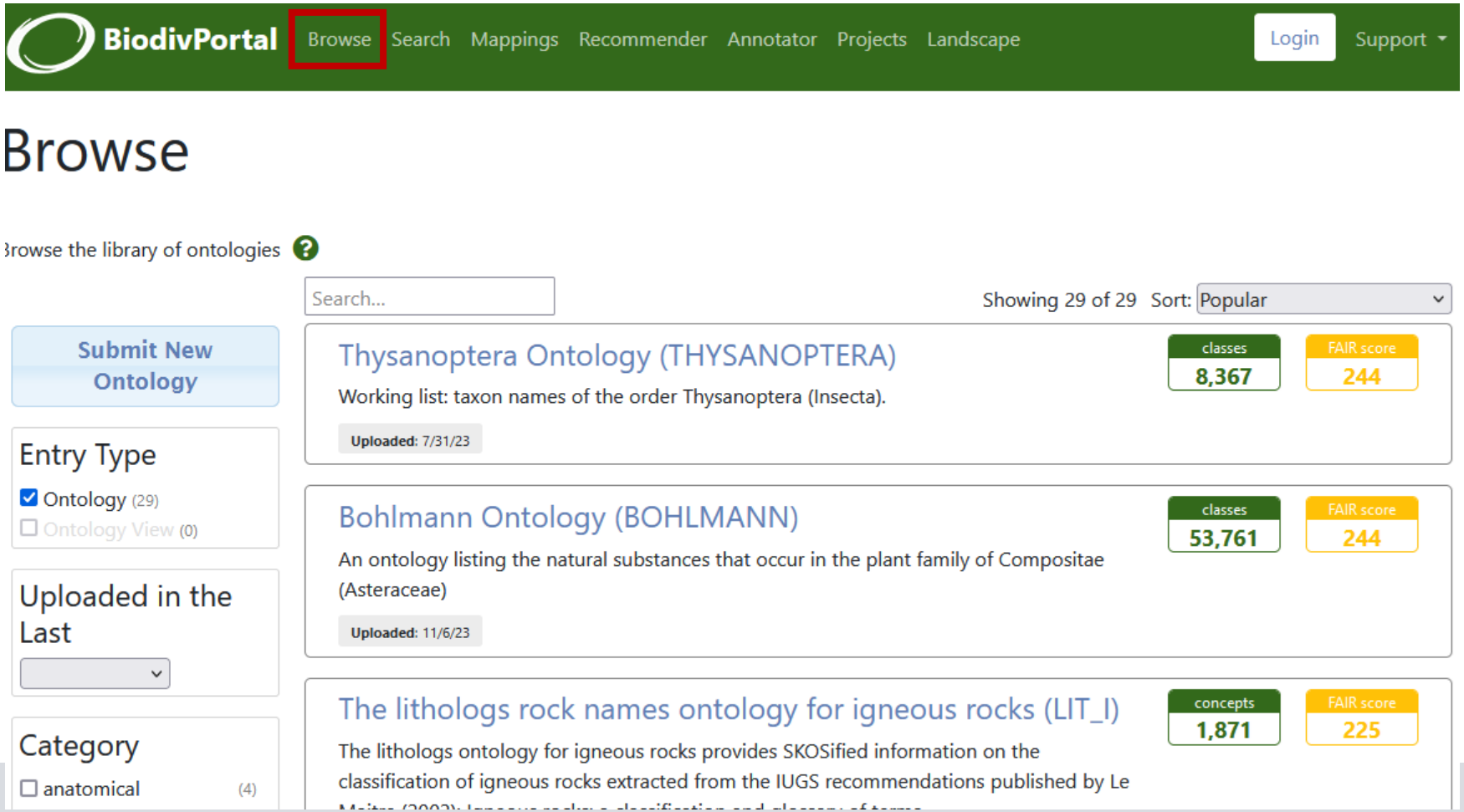
NFDI 4 BIODIVERSITY

- Repository to store, search, browse and visualize terminologies
- Offer efficient semantic annotation for text and tabular data
- Automatically generate and store mappings between terminologies
- Provide a single access point (API) to heterogenous terminological resources
- Offer an environment for the development, curation, and publication of project terminologies
- Provide efficient terminology versioning and evolution mechanisms
- Offer Community feedback mechanisms

How Requirements on a Terminology Service

NFDI 4 BIODIVERSITY

Repository to store, search, browse and visualize terminologies

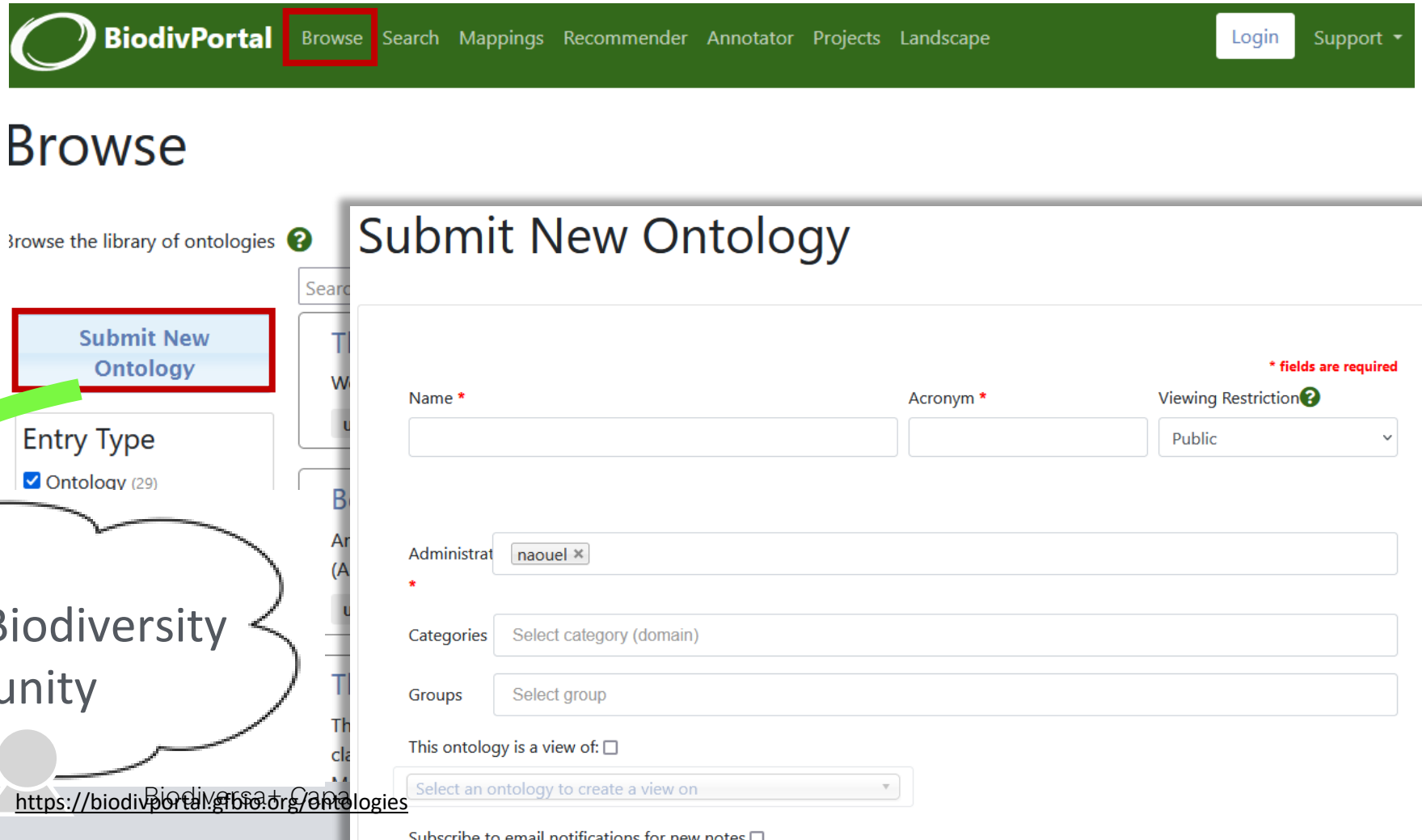


https://biodivportal.gfbio.org/ontologies

How Requirements on a Terminology Service

NFDI 4 BIODIVERSITY

Repository to store, search, browse and visualize terminologies



https://biodivportal.gfbio.org/ontologies

How Requirements on a Terminology Service

NFDI 4 BIODIVERSITY

Offer efficient semantic annotation for text and tabular data

BiodivPortal | Browse | Search | Mappings | Recommender | Annotator | Projects | Landscape | Login | Support

Annotator

Get annotations for biomedical text with classes from the ontologies

This dataset provides palaeoecological data for the glacial and Holocene sediment sequence retrieved from Lake Uddelermeer (The Netherlands) in 2012. Counts of pollen and chironomids are presented against both depth (cm) and age (cal yr. BP), and a loss-on ignition record is presented against depth. A total of 110 samples were analysed for their pollen content; a subset of 32 samples was analysed for sterol and stanol compounds. A total of 1412 samples were analysed for their organic content, approximated through loss-on-ignition. The data provides information on changes in the local and regional vegetation, changes in the within-lake ecosystem and changes in minerogenic input into the lake.

Get annotations

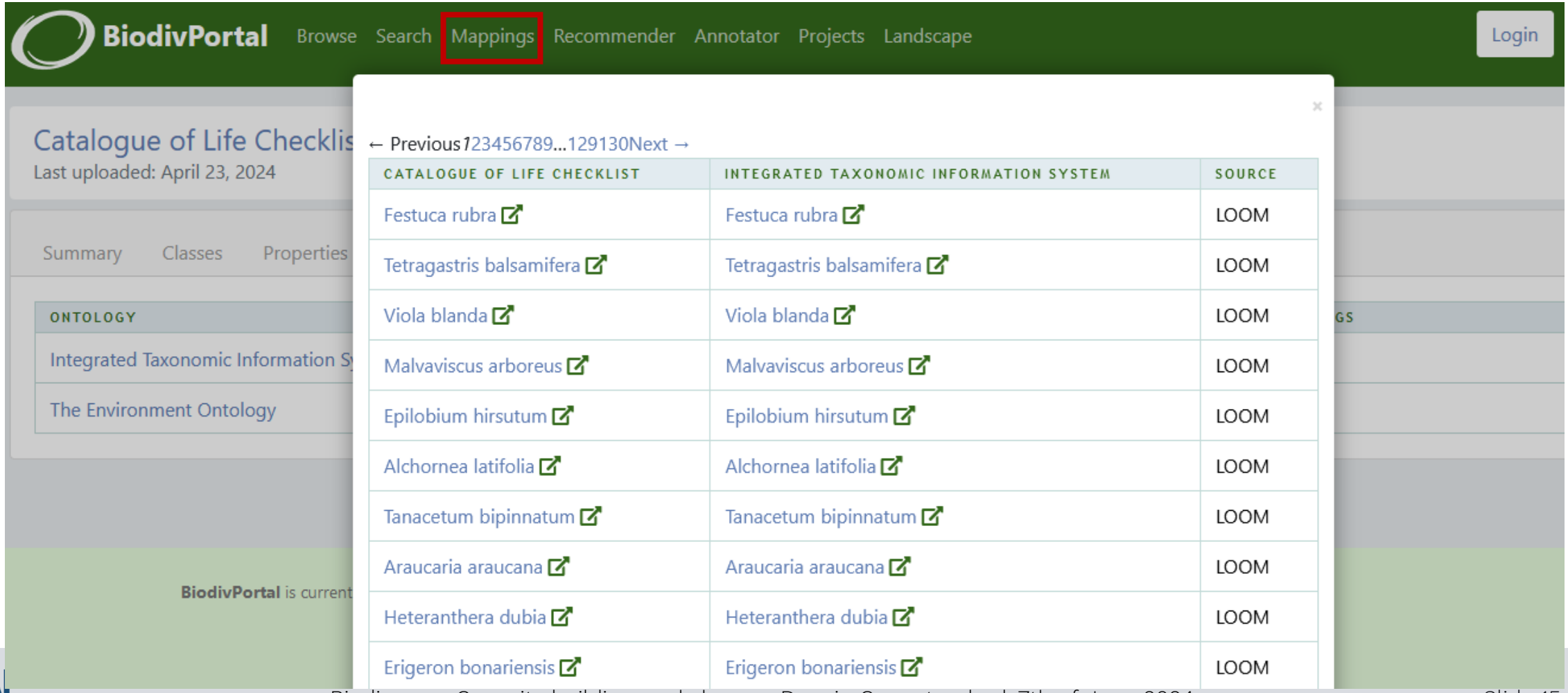
CLASS filter	ONTOLOGY filter	TYPE filter	UMLS SEM TYPE	CONTEXT	MATCHED CLASS filter	MATCHED ONTOLOGY filter
Dataset	Schema.org Vocabulary	direct		This **dataset** provides palaeoecological data ...	Dataset	Schema.org Vocabulary
glaciation	The Environment Ontology	direct		... for the **glacial** and Holocene sediment ...	glaciation	The Environment Ontology
sediment	The Environment Ontology	direct		... and Holocene **sediment** sequence retrieved from ...	sediment	The Environment Ontology
Sequence	ABCD Base Ontology	direct		... Holocene sediment **sequence** retrieved from Lake ...	Sequence	ABCD Base Ontology
saline evaporation pond	The Environment Ontology	direct		... retrieved from **Lake** Uddelermeer (The Netherlands) ...	saline evaporation pond	The Environment Ontology

InfAI Institut für Angew

How Requirements on a Terminology Service

NFDI 4 BIODIVERSITY

Automatically generate and store mappings between terminologies



How Requirements on a Terminology Service

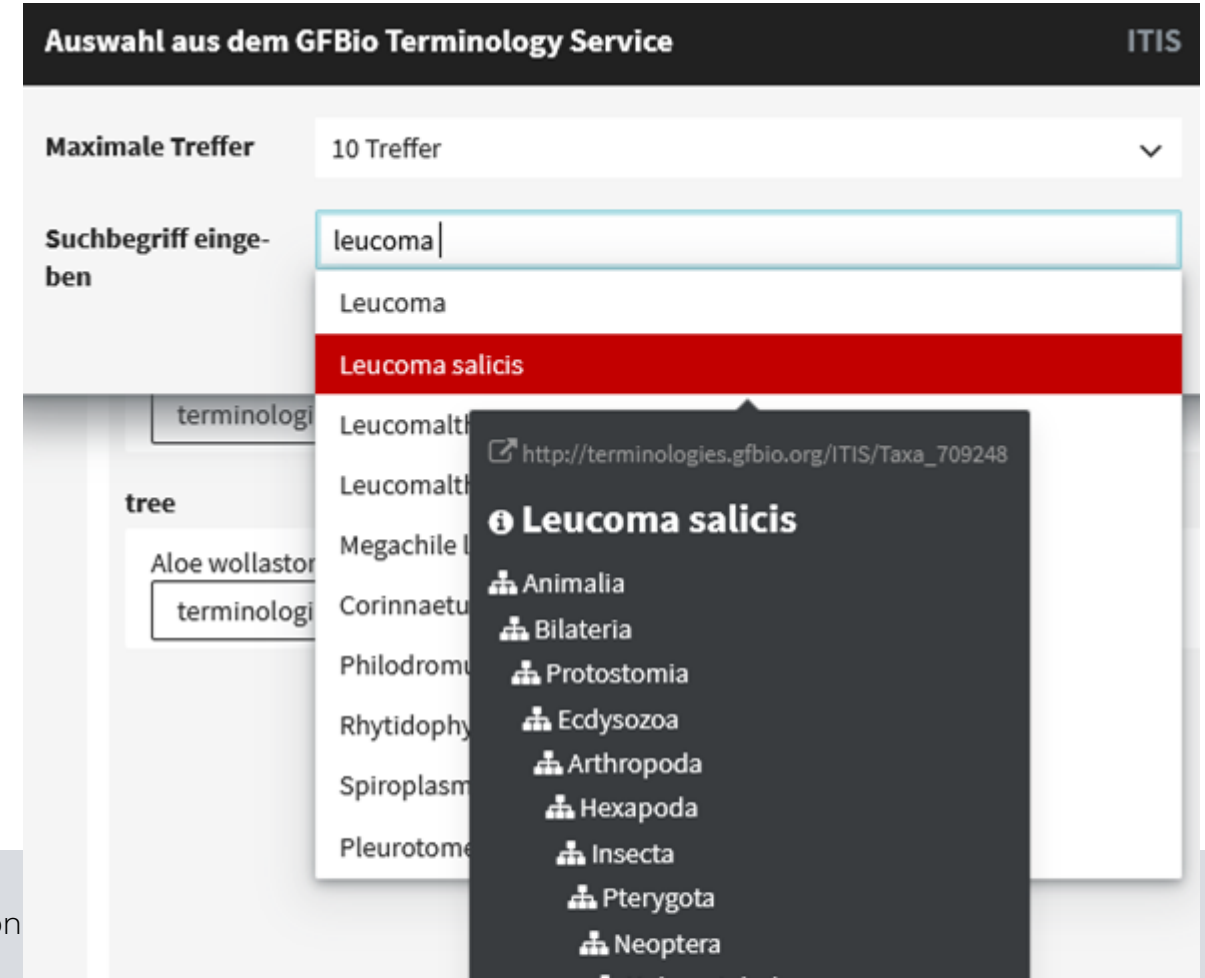
NFDI 4 BIODIVERSITY

Provide a single access point (API) to heterogenous terminological resources

```
{
  - links: {
        annotator: https://data.biodivportal.gfbio.org/annotator,
        categories: https://data.biodivportal.gfbio.org/categories,
        groups: https://data.biodivportal.gfbio.org/groups,
        documentation: https://data.biodivportal.gfbio.org/documentation,
        mappings: https://data.biodivportal.gfbio.org/mappings,
        metrics: https://data.biodivportal.gfbio.org/metrics,
        notes: https://data.biodivportal.gfbio.org/notes,
        ontologies: https://data.biodivportal.gfbio.org/ontologies,
        ontologies_full: https://data.biodivportal.gfbio.org/ontologies_full,
        analytics: https://data.biodivportal.gfbio.org/analytics,
        submissions: https://data.biodivportal.gfbio.org/submissions,
        projects: https://data.biodivportal.gfbio.org/projects,
        property_search: https://data.biodivportal.gfbio.org/property_search,
        provisional_classes: https://data.biodivportal.gfbio.org/provisional_classes,
        provisional_relations: https://data.biodivportal.gfbio.org/provisional_relations,
        recommender: https://data.biodivportal.gfbio.org/recommender,
        recommender_v1: https://data.biodivportal.gfbio.org/recommender_v1,
        replies: https://data.biodivportal.gfbio.org/replies,
        reviews: https://data.biodivportal.gfbio.org/reviews,
        search: https://data.biodivportal.gfbio.org/search,
        slices: https://data.biodivportal.gfbio.org/slices,
        submission_metadata: https://data.biodivportal.gfbio.org/submission_metadata,
        ontology_metadata: https://data.biodivportal.gfbio.org/ontology_metadata,
        users: https://data.biodivportal.gfbio.org/users,
```


https://data.biodivportal.gfbio.org

API integration in GBV NaviWeb Naturwissenschaften Niedersachsen



How Ongoing Work

- Offer an environment for the development, curation, and publication of project terminologies
- Provide efficient terminology versioning and evolution mechanisms
- Offer Community feedback mechanisms

NFDI 4
BIODIVERSITY

Contact:
karam@infai.org

naouelkaram

www.nfdi4biodiversity.org

Approaches for an Open and FAIR Research Lifecycle

Ilaria Rosati, Mariantonietta La Marra, Alexandra Muresan, Martina Pulieri, Andrea Tarallo

biodiversa+ European Biodiversity Partnership

Data management and capacity building workshop on Darwin Core standard – 7 June 2024

Open Science & FAIR Principles: what?

Data and other research outputs are available in the public domain or under copyright and licensed under an open licence that allows access, re-use, repurpose, adaptation and distribution under specific conditions.

UNESCO Recommendation on Open Science 2022 version 1.
https://doi.org/10.54677/UTCD9302

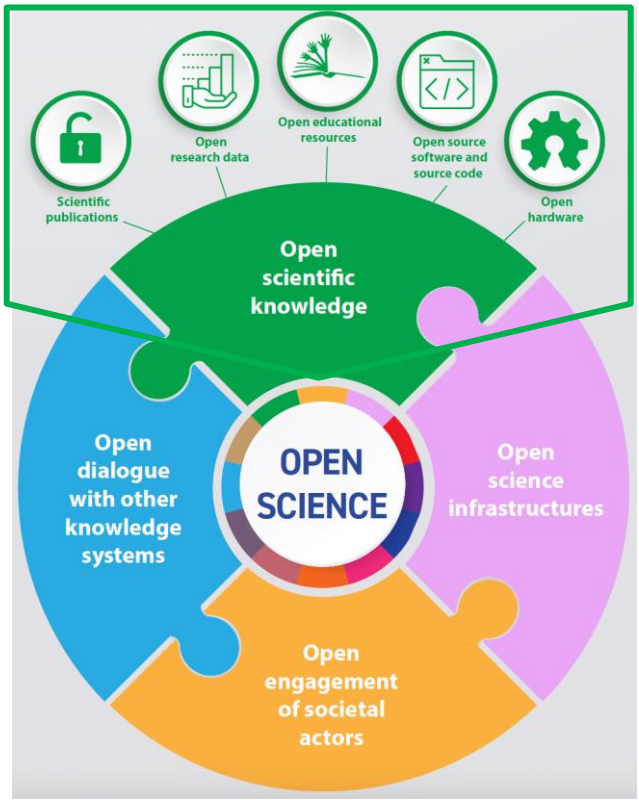


Image: UNESCO Open Science brochure

Open Science & FAIR Principles: what?

Image: Illustrations from the Turing Way book dashes. Zenodo. http://doi.org/10.5281/zenodo.3695300

Image: Icons by Freepik from www.flaticon.com and ARDC https://conference.eresearch.edu.au/fair-go-new-resources-to-support-fair-data/

Open Science & FAIR Principles: what?

FAIR ≠ Open

"FAIR is not the equivalent of open, but open needs to be FAIR to be useful"

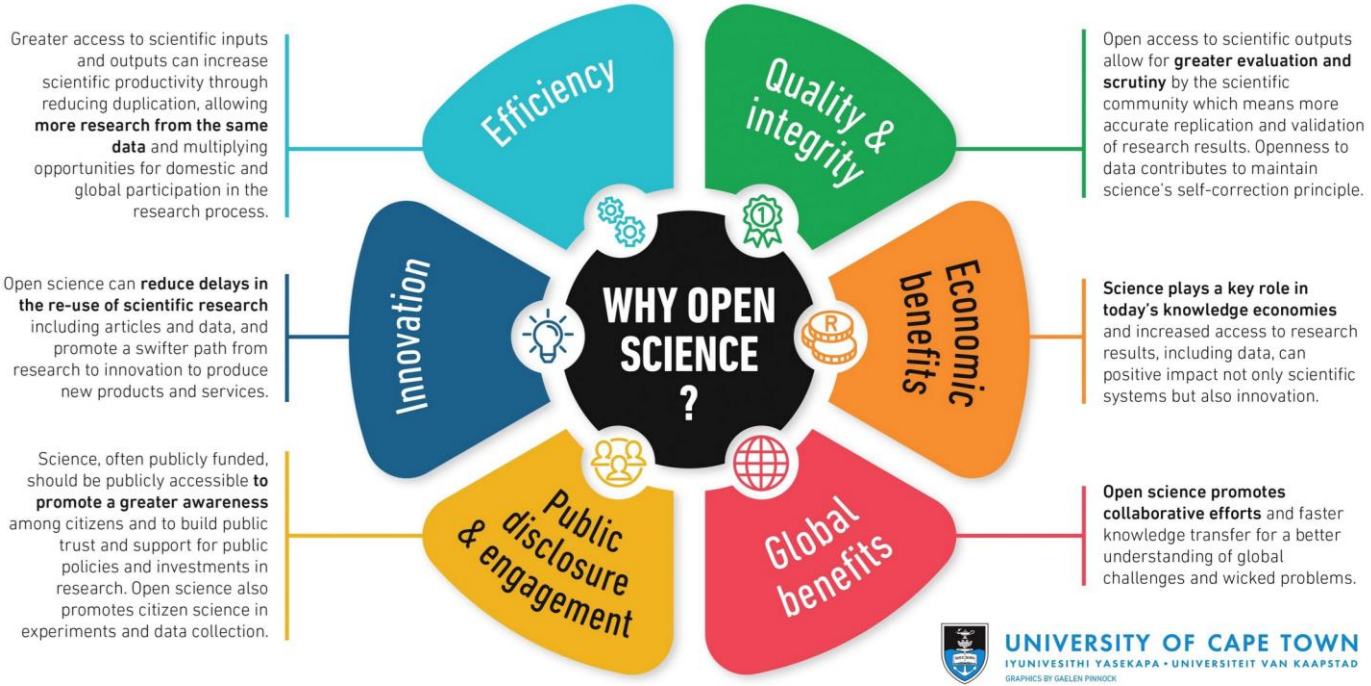
Making your data/digital resources openly available does not translate to it being reusable!

To do so, we need clear and detailed information and data description.

Data can be FAIR but not Open! FAIR data maxim is: “as open as possible, as closed as necessary”

Ideally we want FAIR data/digital resources openly!

Open Science & FAIR Principles: why?



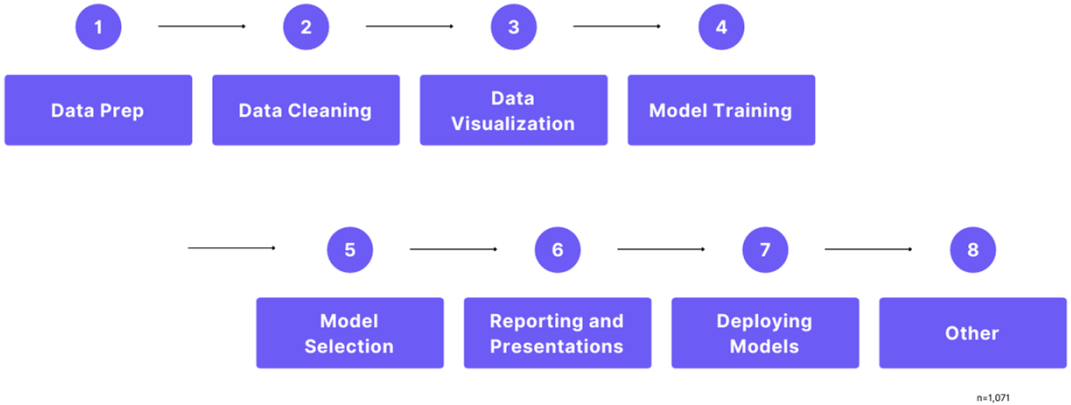
Bildquelle: University of Cape Town, Research Support Hub CC-BY, Creative Commons Attribution 4.0 International License

Open Science & FAIR Principles: who?

Gomes DGE et al. 2022 *Why don't we share data and code? Perceived barriers and benefits to public archiving practices*. Proc. R. Soc. B 289: 20221113. https://doi.org/10.1098/rspb.2022.1113

Open Science & FAIR Principles: who?

Thinking about your current role, what tasks are most time consuming?



Anaconda's “The State of Data Science 2023” report

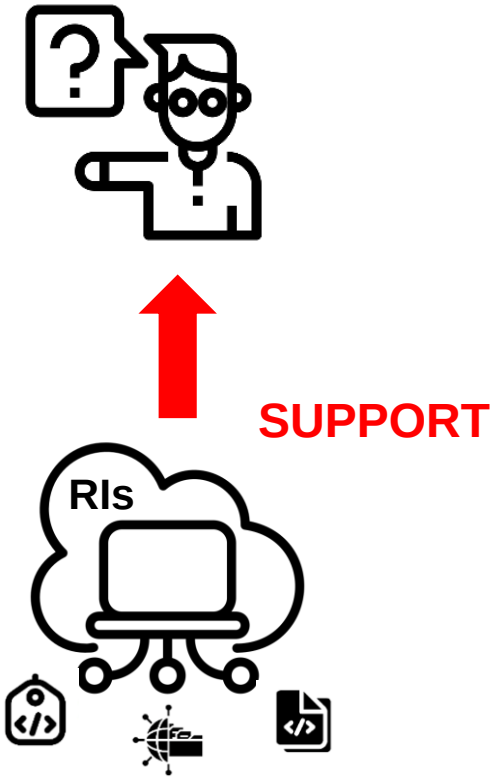
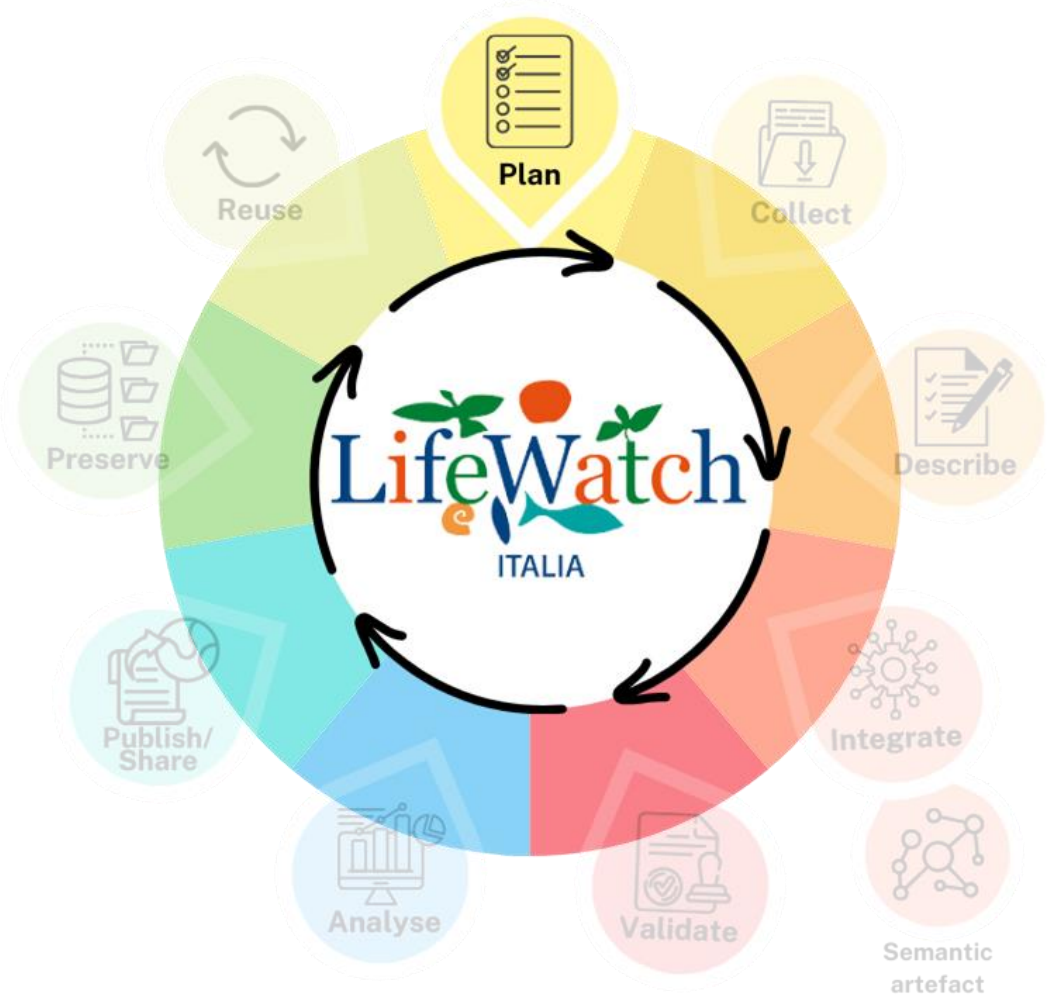


Image by LUTFI GANI AL ACHMAD from Noun Project (CC BY 3.0)

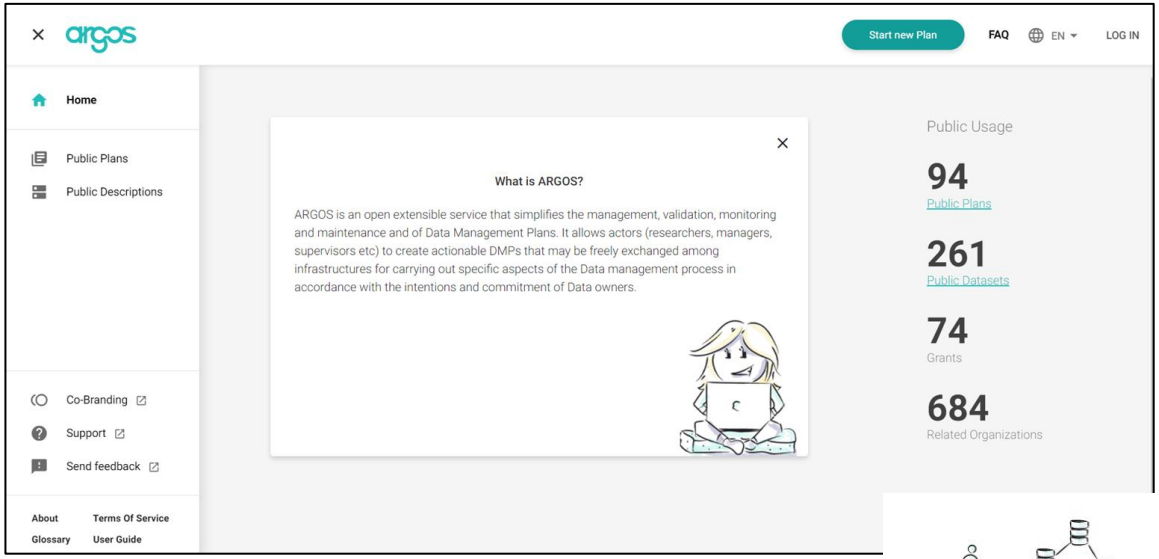
Open & FAIR Research Lifecycle

- Findable
- Accessible
- Interoperable
- Reusable
- Open

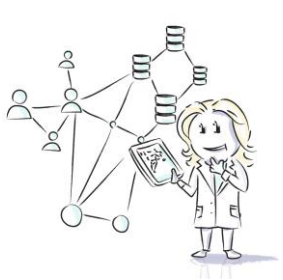
Data Management Plan



Plan the management of data and other research products



https://argos.openaire.eu/home

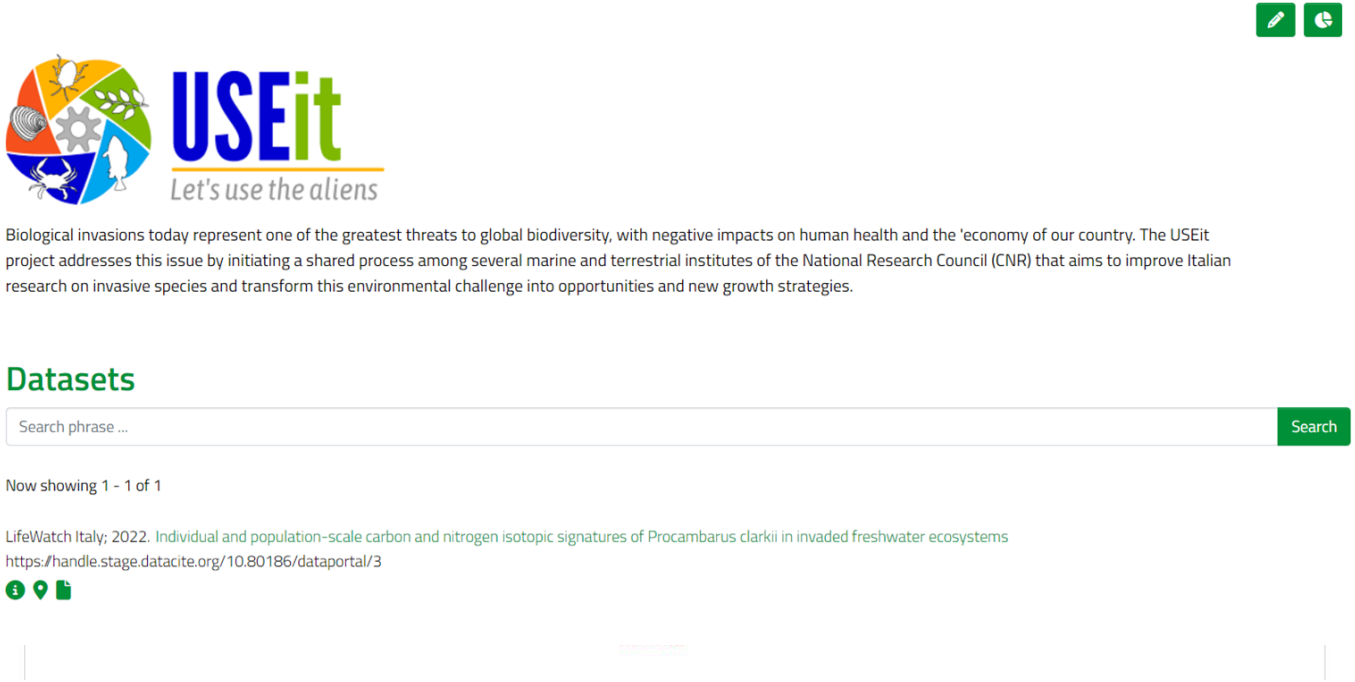


- Create machine actionable DMPs
- Configure to best fit your discipline
- Link to EOSC components out of the box
- Share easily in your repository

Data Portal

Datasets

Collect data



Creation and customimisation of sub-portals for gathering datasets produced of the same project or initiative.

Describe your data

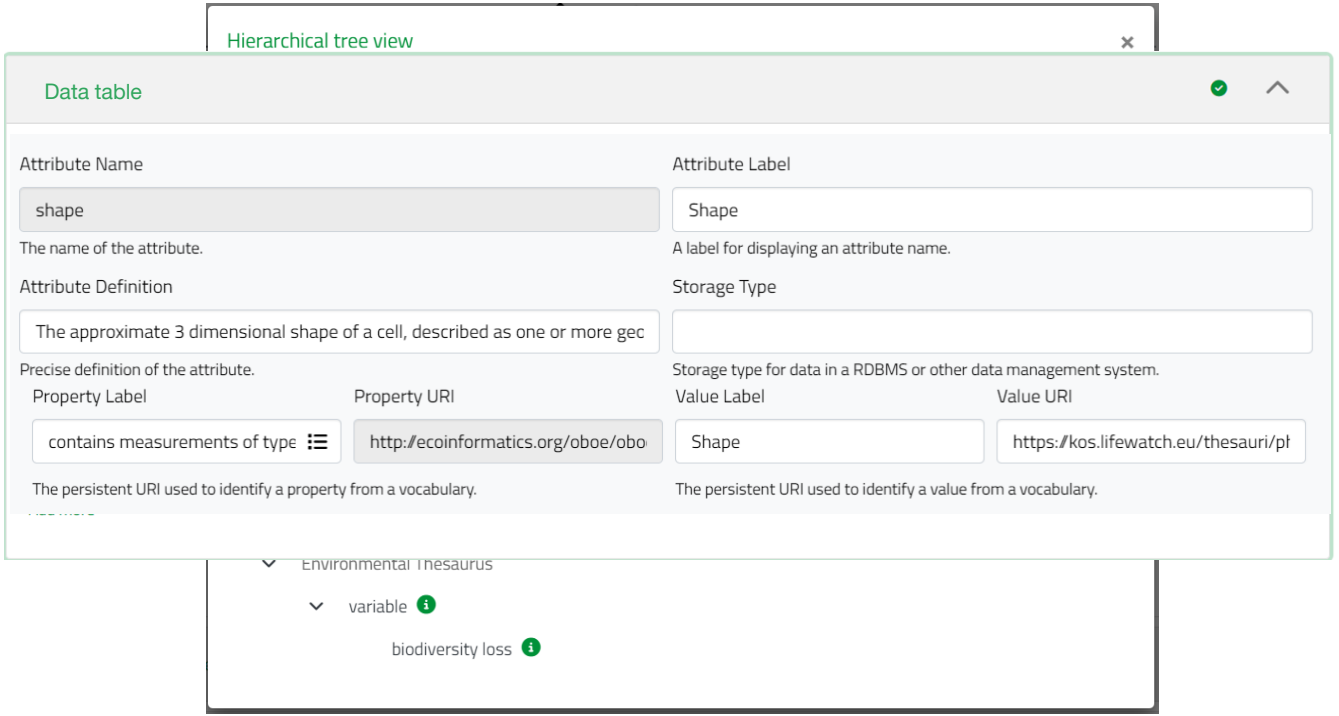
https://eml.ecoinformatics.org/eml-schema

Ecological Metadata Language 2.2.0 as metadata schema for describing ecological data.

Every section contains a set of attributes, e.g. Title, Publication date, Temporal coverage, Taxonomic coverage, ect, to thoroughly describe your data.

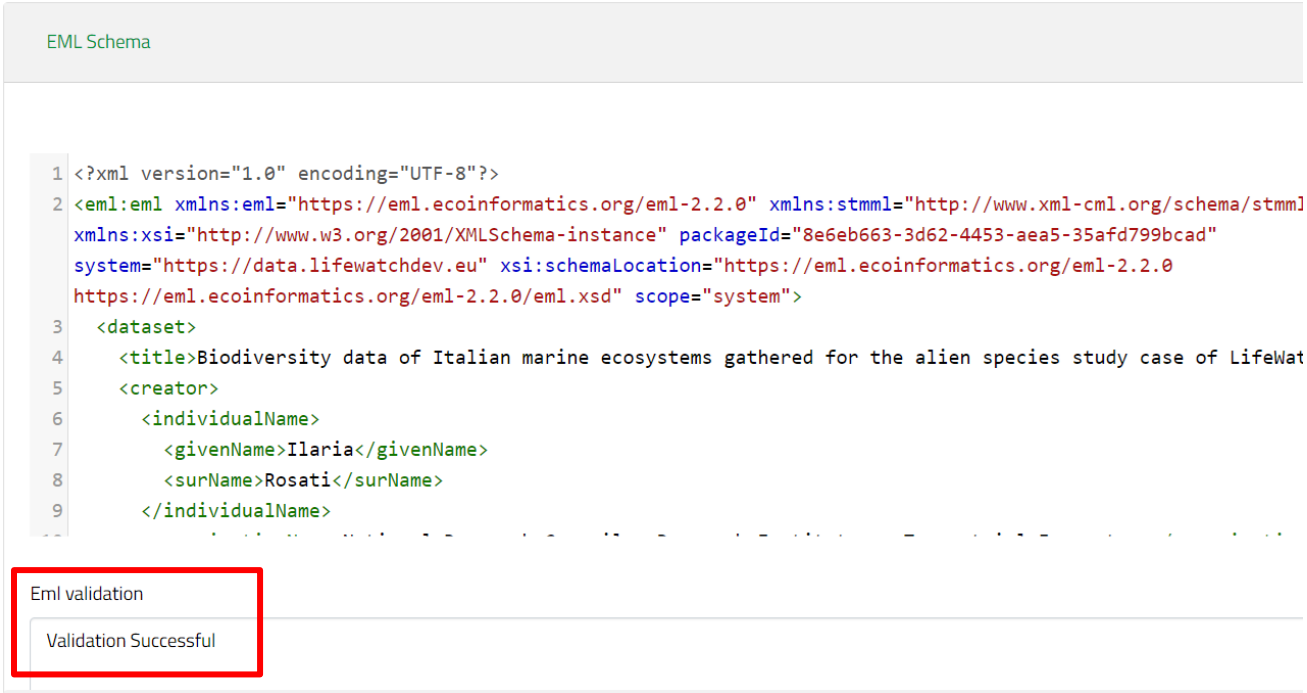
Integration with semantic artefacts

EcoPortal integration to annotate (meta)data and to semantically enrich them.



https://ecoportal.lifewatch.eu

Metadata validation



The xml file is checked against the EML schema validator (.xsd file)

Data validation: Taxonomic check

LW-ITA Data Schema based on DarwinCore and controlled vocabularies

scientificName	samplingProtocol	bodyLenght
Idotea balthica	hand operated ekman grab	4.5

scientificName
Idotea balthica

Taxonomic check

Validated data

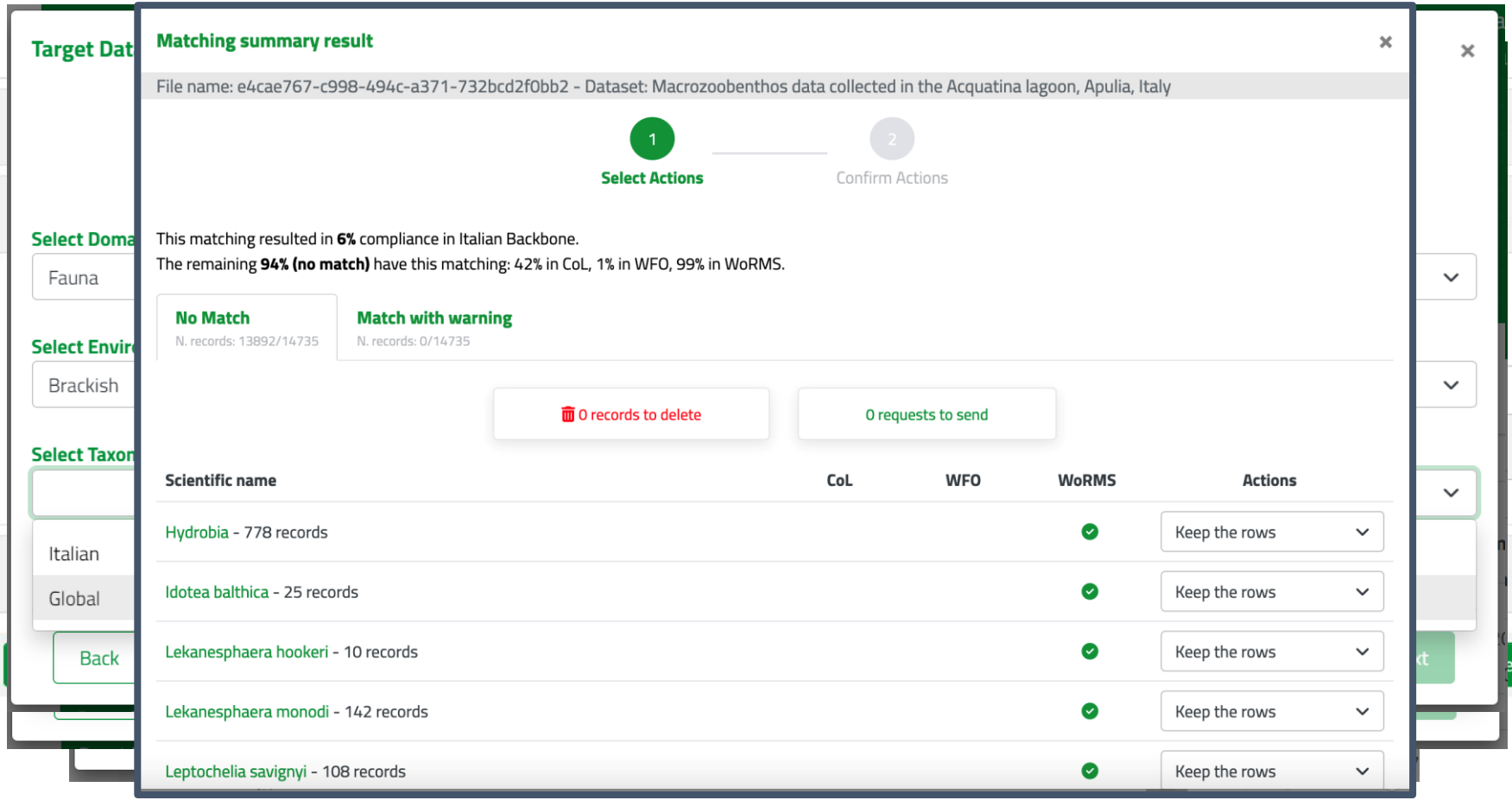
LifeWatch

Italian Taxonomic Backbone

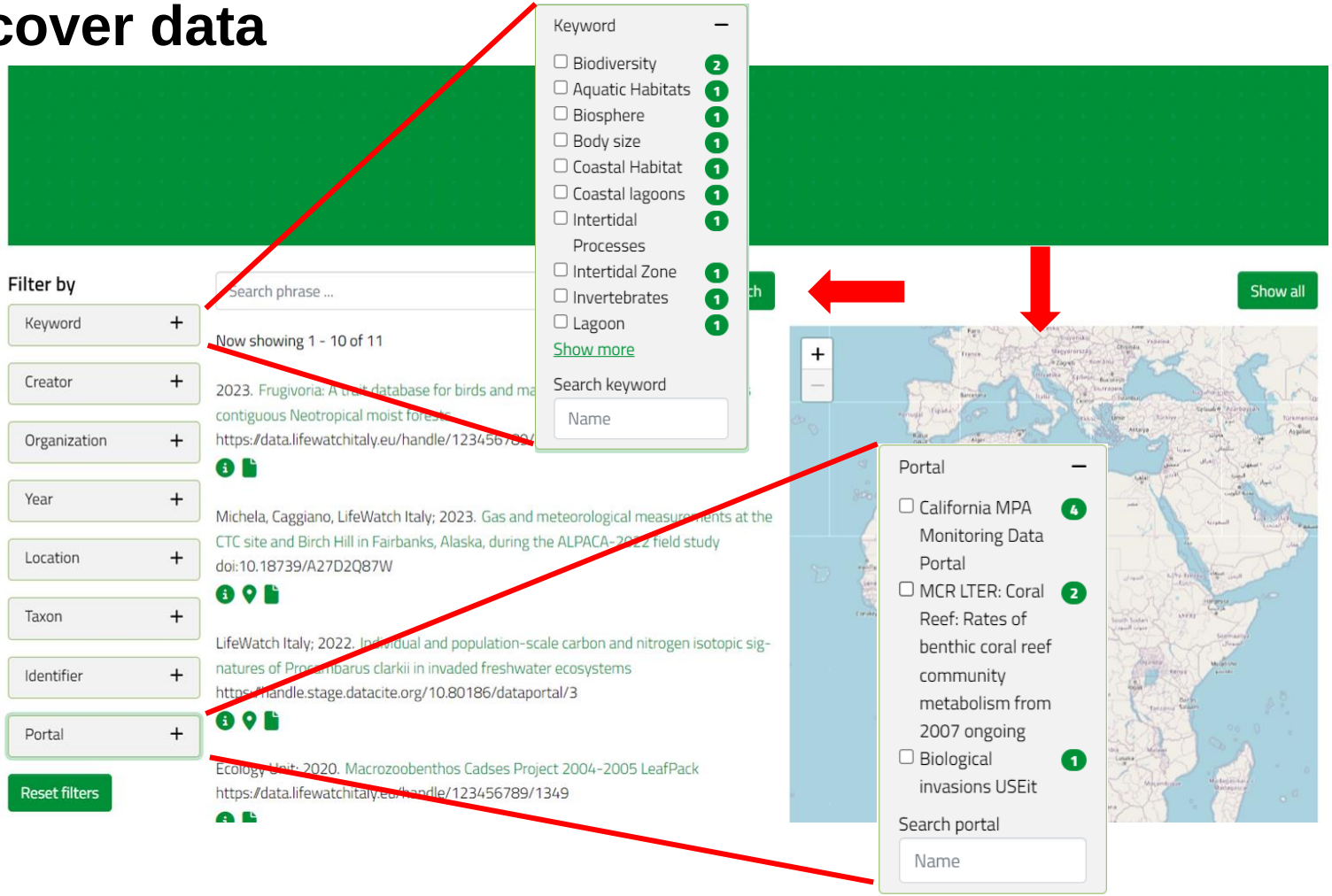
Catalogue of Life

WoRMS World Register of Marine Species

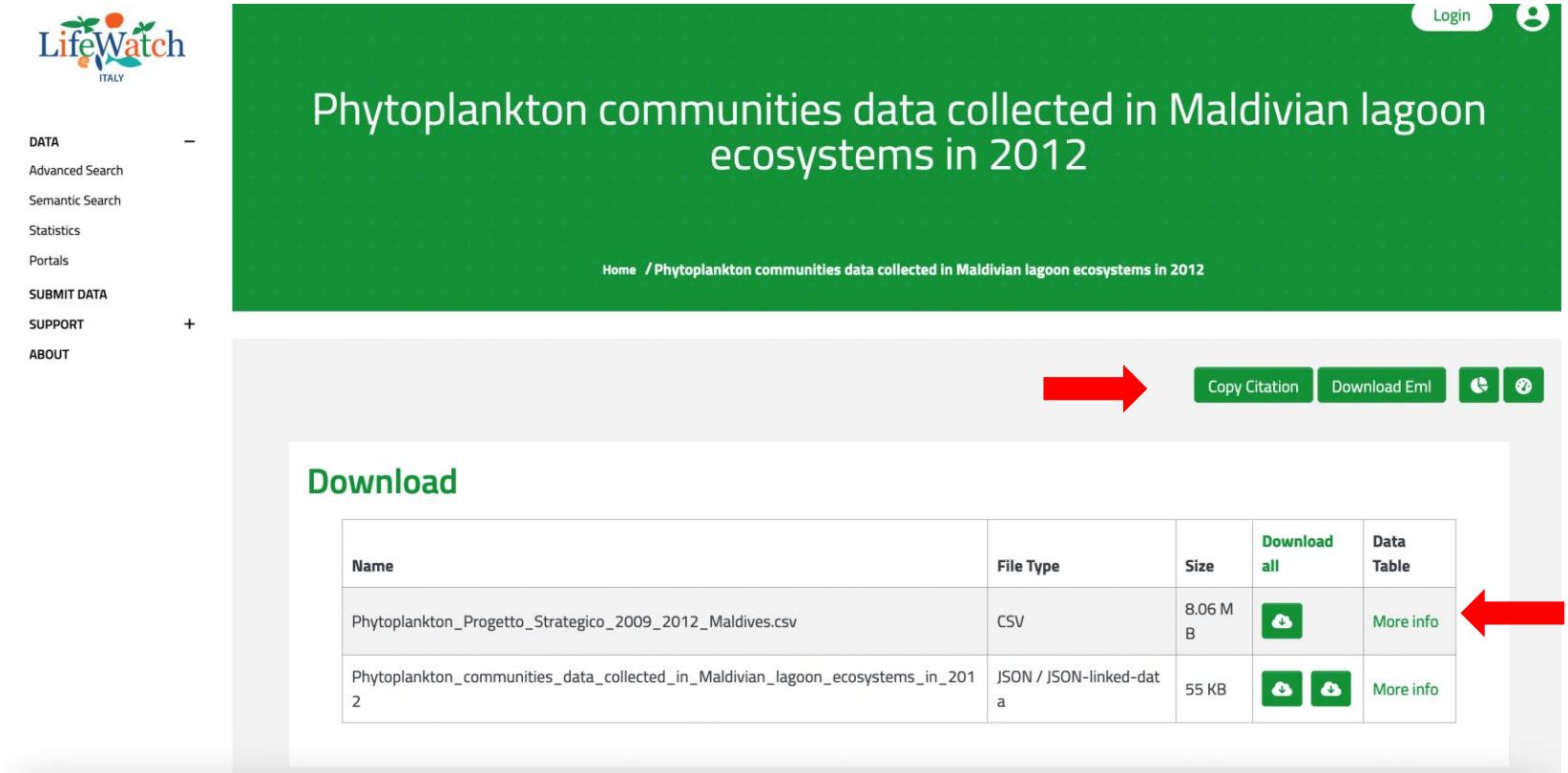
Data validation: Taxonomic check



Discover data



Reuse data


LifeWatch
ITALY
DATA
Advanced Search
Semantic Search
Statistics
Portals
SUBMIT DATA
SUPPORT
ABOUT
Login
Phytoplankton communities data collected in Maldivian lagoon ecosystems in 2012
Home / Phytoplankton communities data collected in Maldivian lagoon ecosystems in 2012
Copy Citation
Download Eml
Download
Name
File Type
Size
Download all
Data Table
Phytoplankton_Progetto_Strategico_2009_2012_Maldives.csv
CSV
8.06 MB
More info
Phytoplankton_communities_data_collected_in_Maldivian_lagoon_ecosystems_in_2012
JSON / JSON-linked-data
55 KB
More info

DataLabs

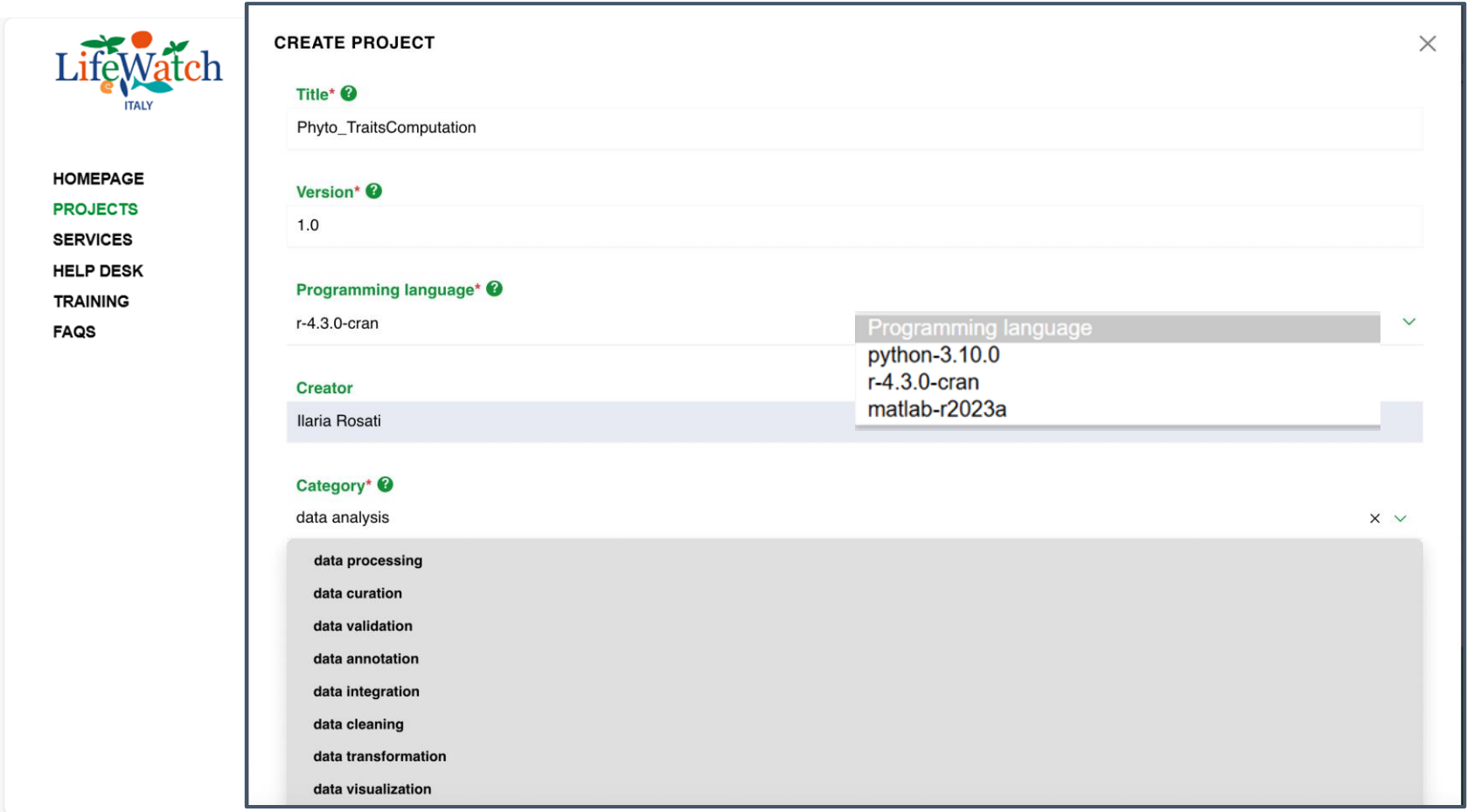
Scripts

Services

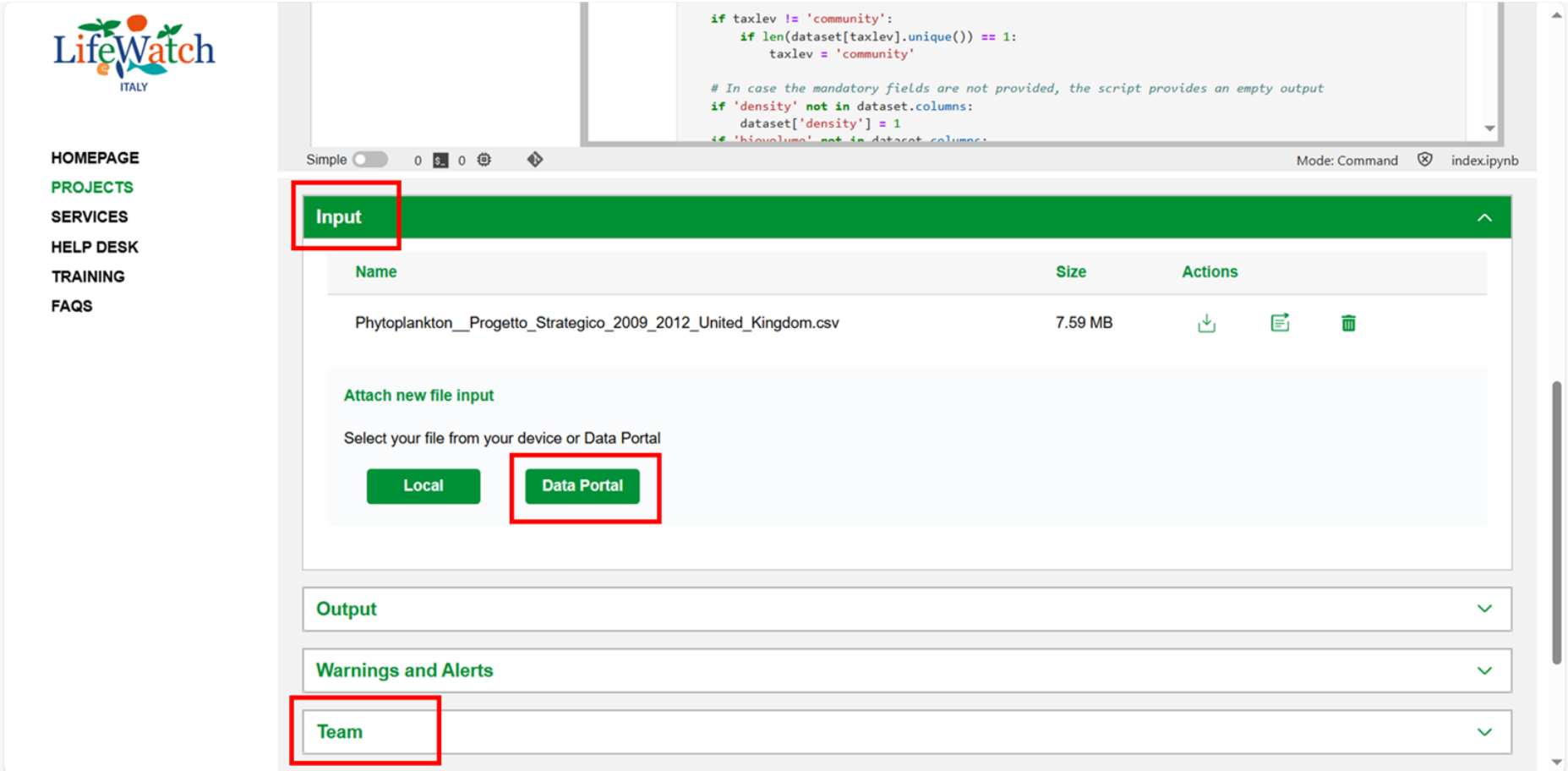
Datasets

https://datalabs.lifewatchitaly.eu

Describe script



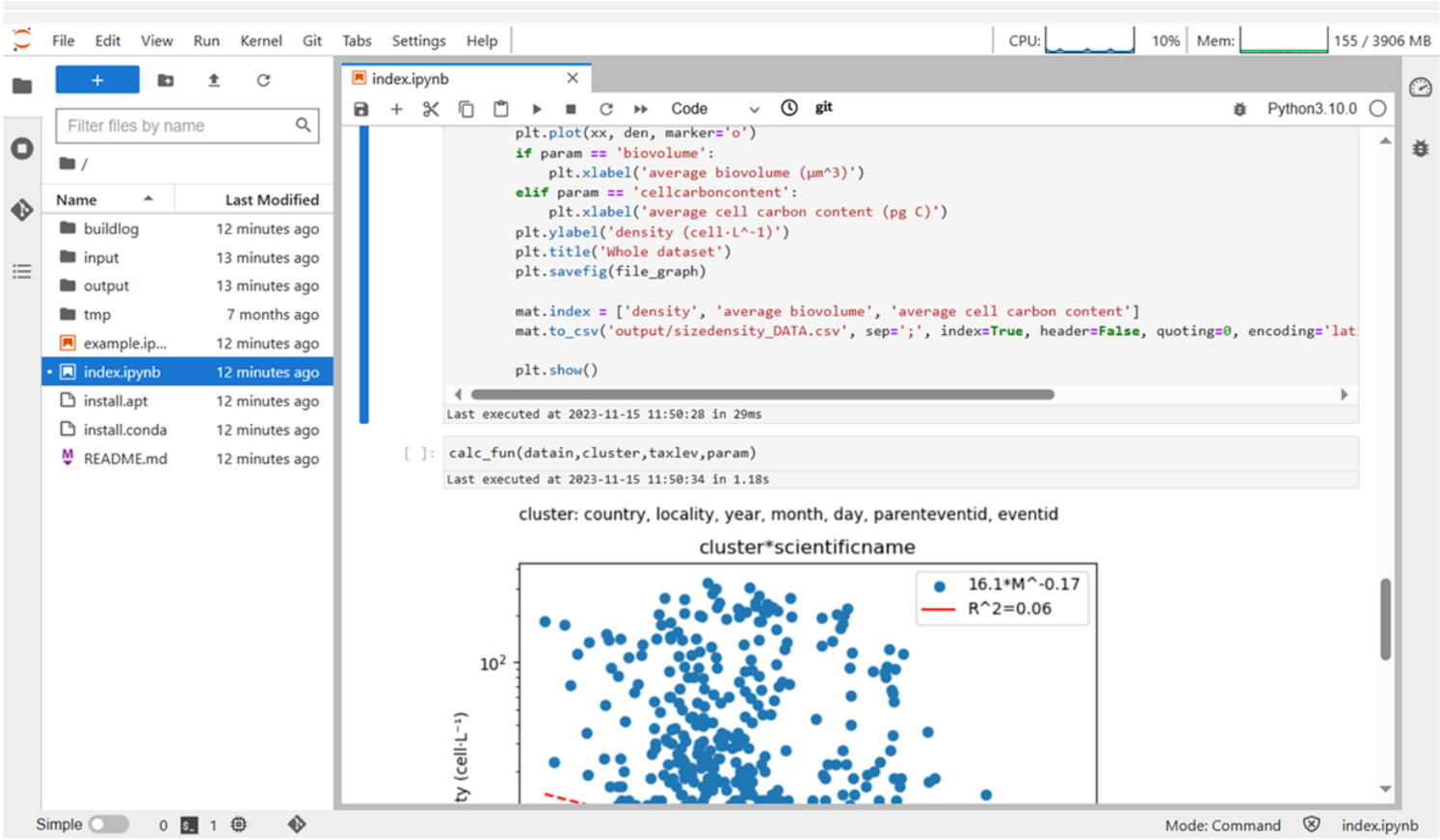
Analyse data



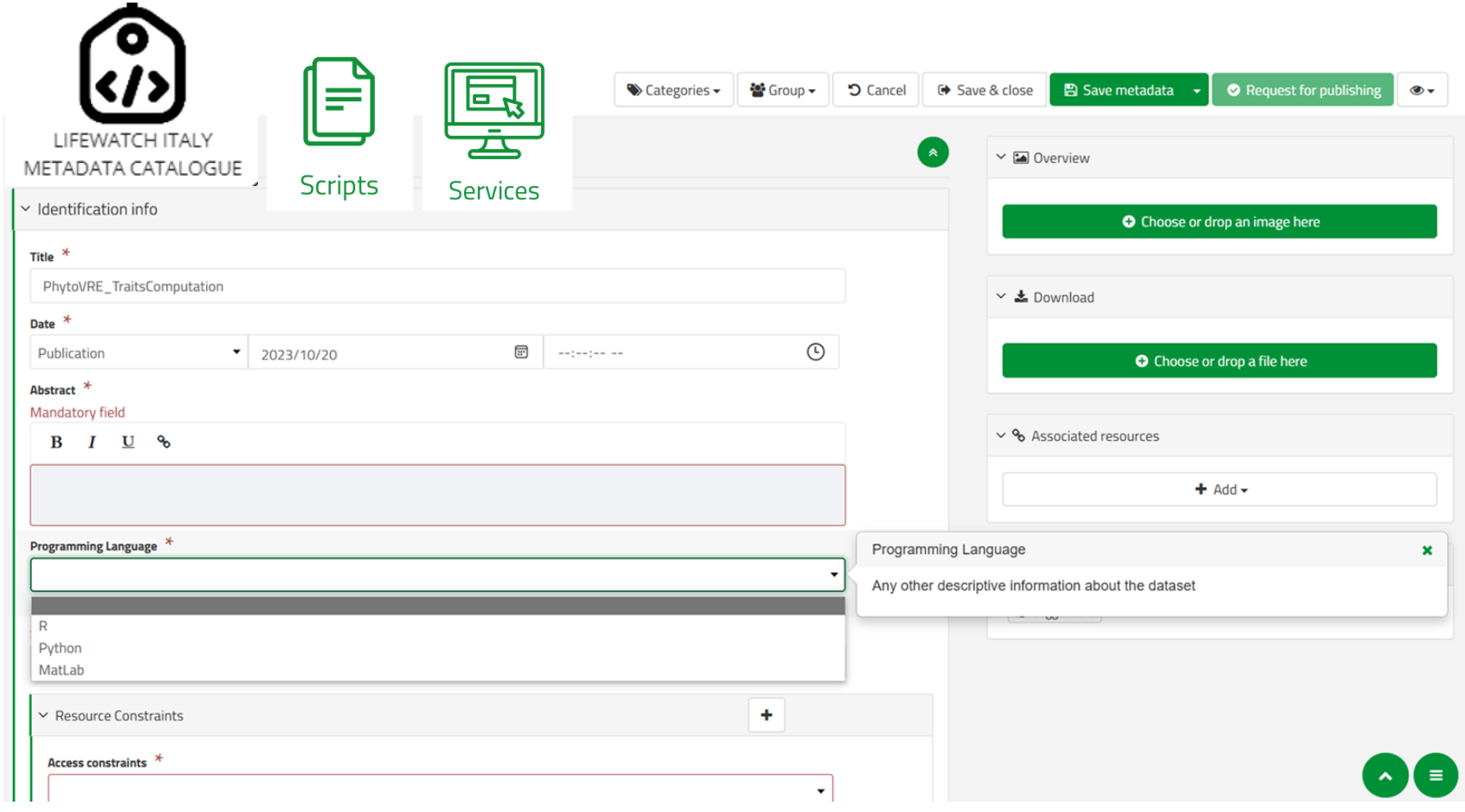
Analyse data

LifeWatch ITALY

- HOMEPAGE
- PROJECTS
- SERVICES
- HELP DESK
- TRAINING
- FAQS

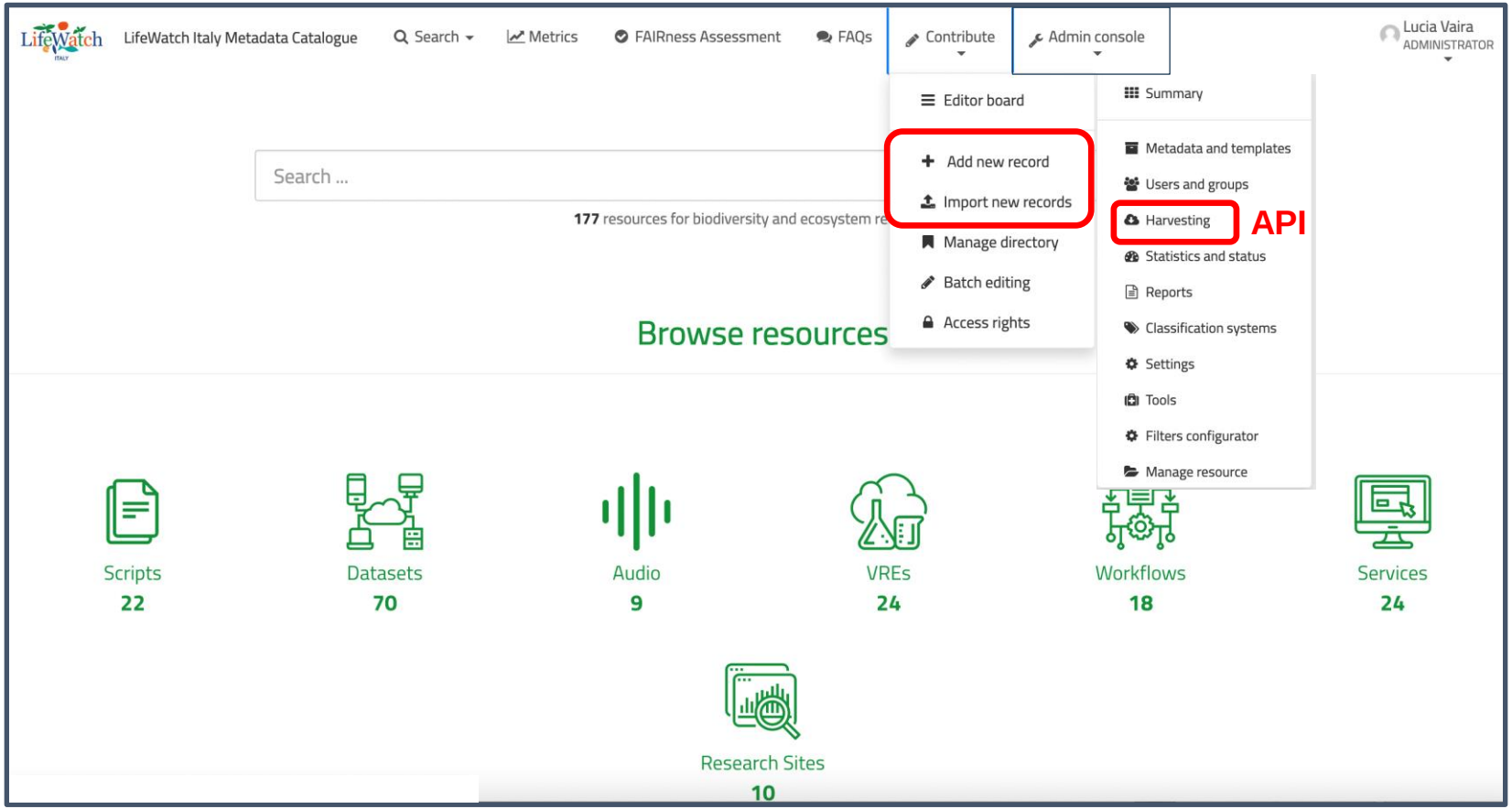


Validate & Publish metadata

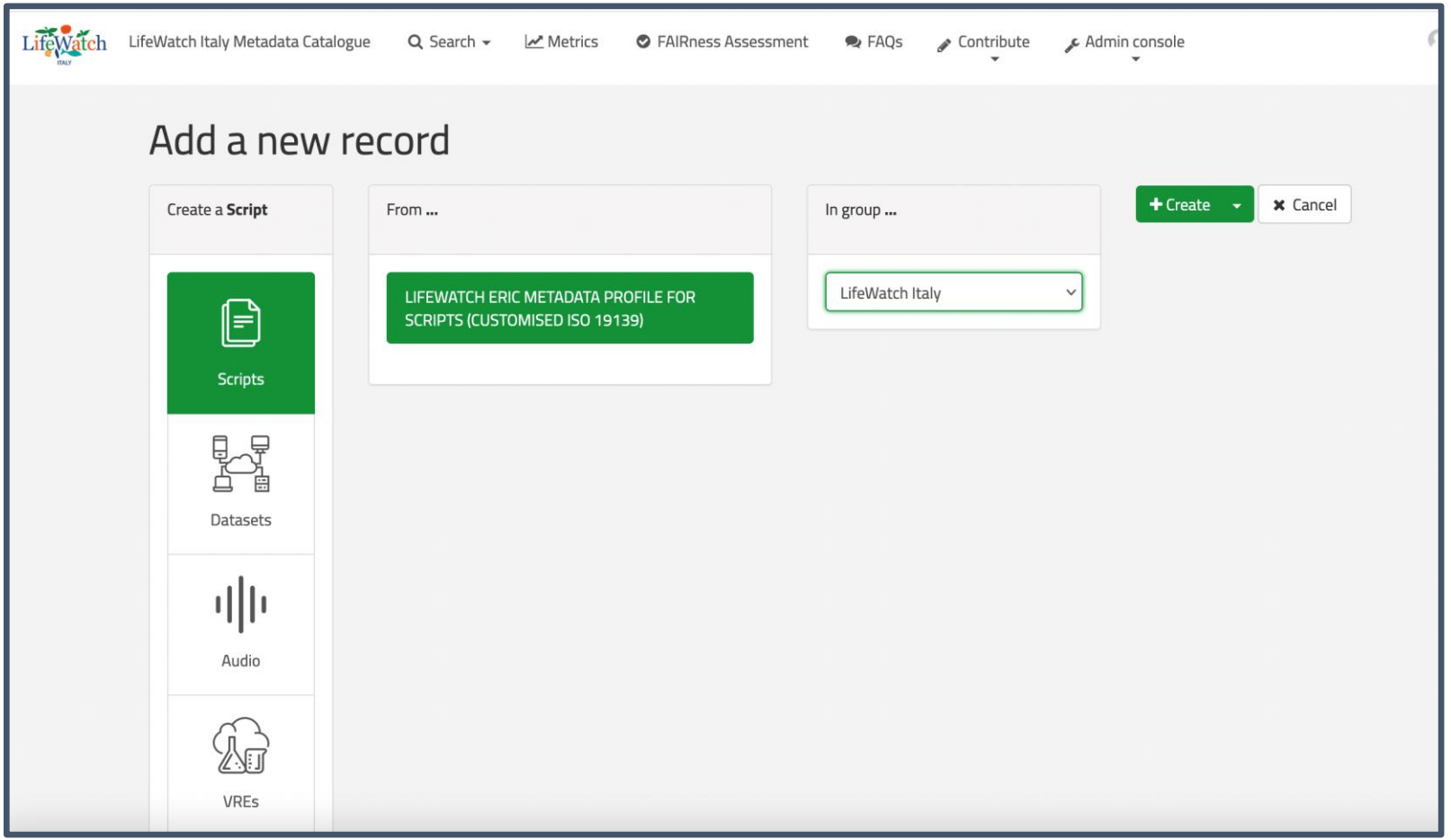


Metadata Catalogue

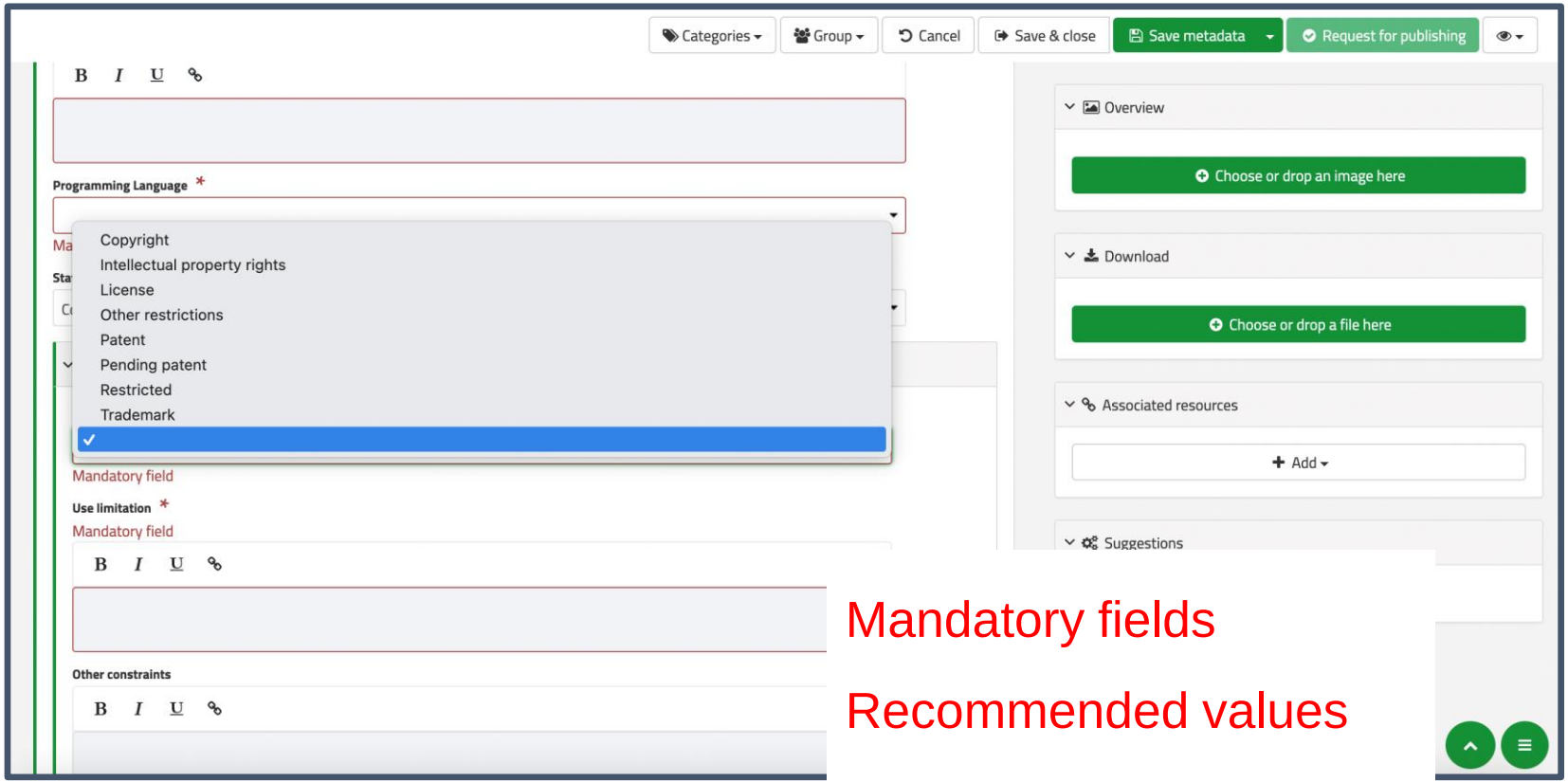
Describe


LifeWatch Italy Metadata Catalogue
Search
Metrics
FAIRness Assessment
FAQs
Contribute
Admin console
Lucia Vaira
ADMINISTRATOR
Editor board
Add new record
Import new records
Manage directory
Batch editing
Access rights
Summary
Metadata and templates
Users and groups
Harvesting
API
Statistics and status
Reports
Classification systems
Settings
Tools
Filters configurator
Manage resource
Search ...
177 resources for biodiversity and ecosystem re
Browse resources
Scripts
22
Datasets
70
Audio
9
VREs
24
Workflows
18
Services
24
Research Sites
10

Describe


LifeWatch Italy Metadata Catalogue
Search
Metrics
FAIRness Assessment
FAQs
Contribute
Admin console
Add a new record
Create a Script
From ...
In group ...
+ Create
Cancel
Scripts
LIFEWATCH ERIC METADATA PROFILE FOR SCRIPTS (CUSTOMISED ISO 19139)
LifeWatch Italy
Datasets
Audio
VREs

Validate

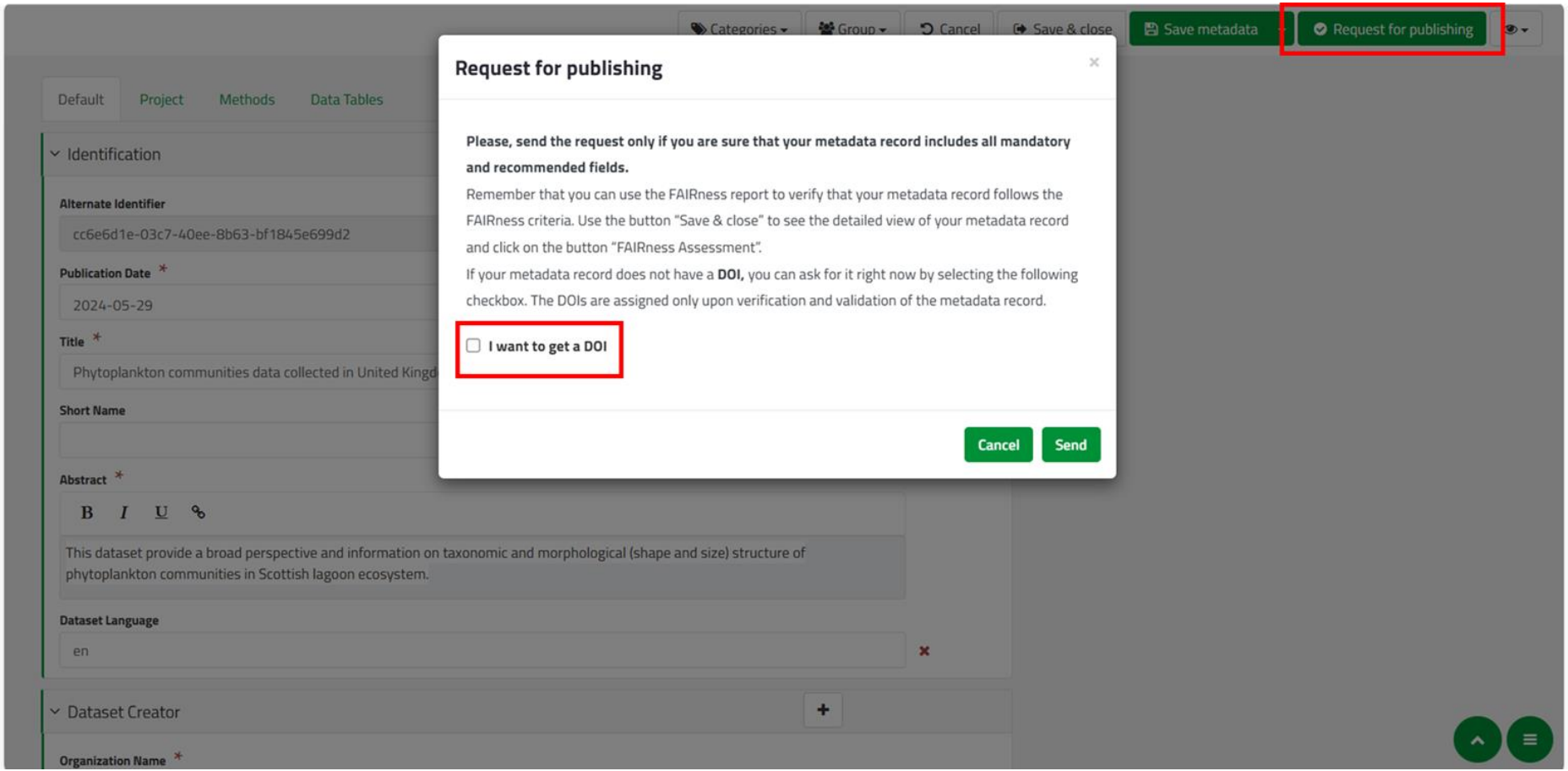


Mandatory fields

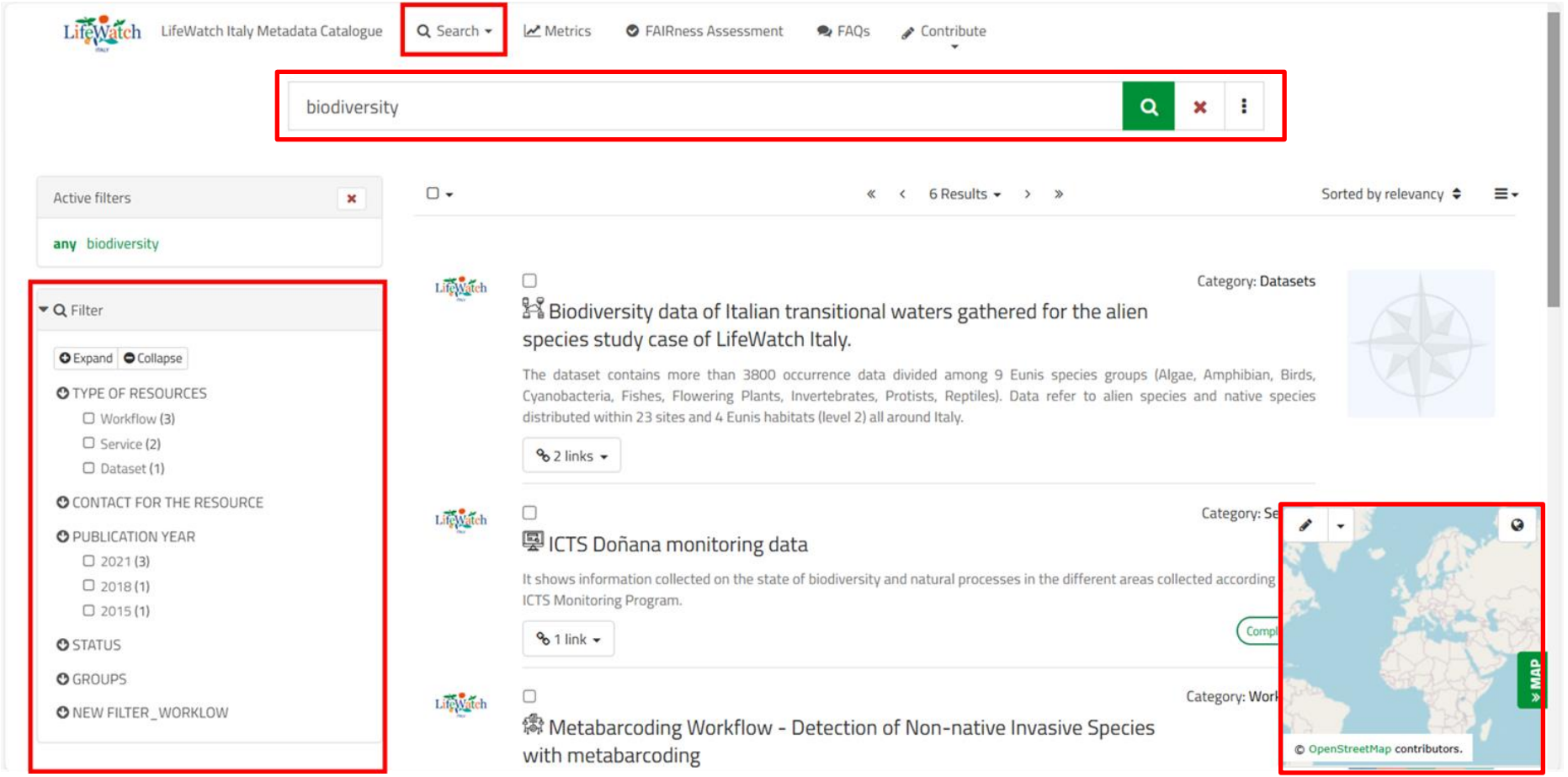
Recommended values

Controlled vocabularies

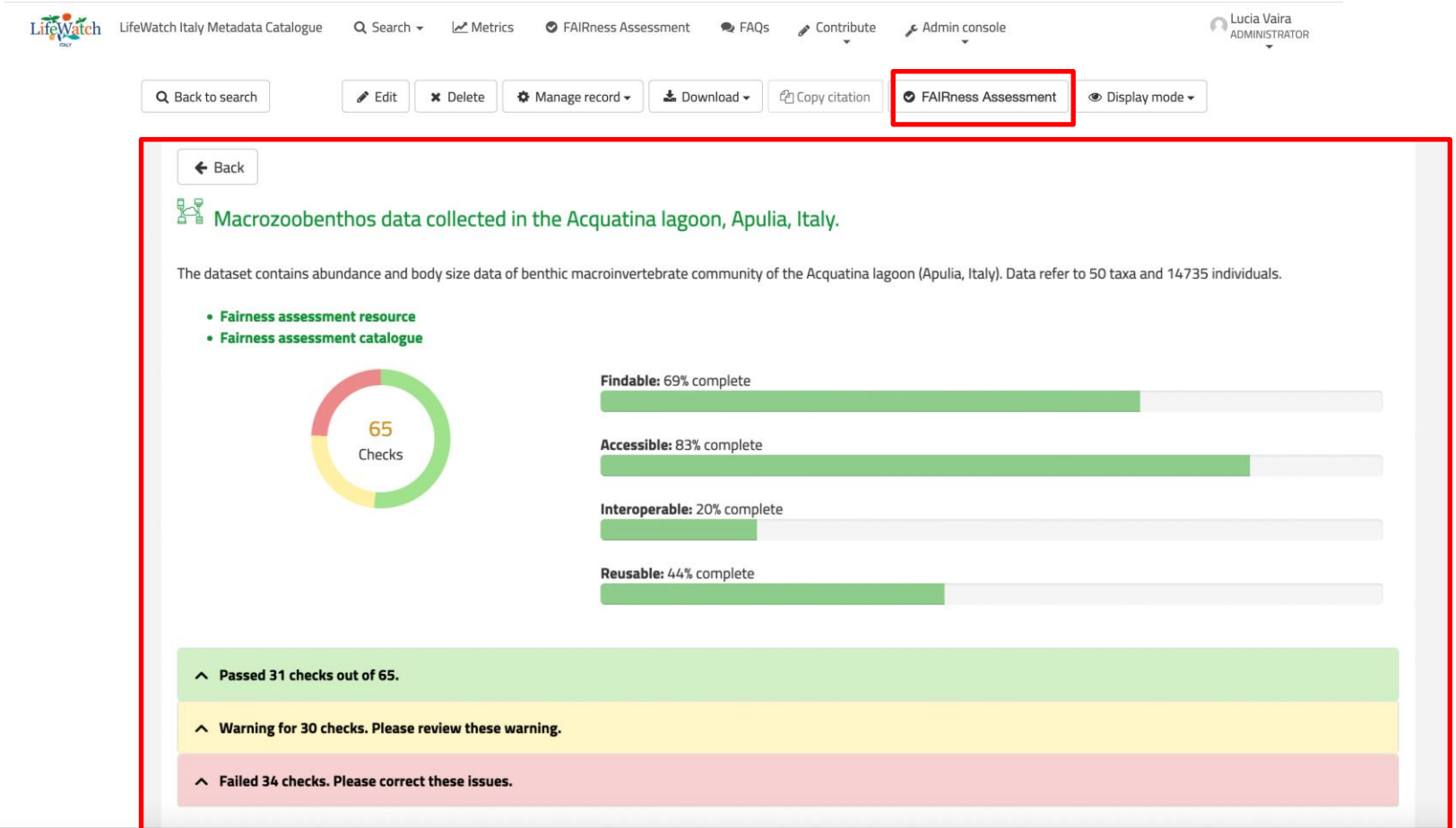
Publish


Categories
Group
Cancel
Save & close
Save metadata
Request for publishing
Default
Project
Methods
Data Tables
Identification
Alternate Identifier
cc6e6d1e-03c7-40ee-8b63-bf1845e699d2
Publication Date *
2024-05-29
Title *
Phytoplankton communities data collected in United Kingd
Short Name
Abstract *
This dataset provide a broad perspective and information on taxonomic and morphological (shape and size) structure of phytoplankton communities in Scottish lagoon ecosystem.
Dataset Language
en
Dataset Creator
Organization Name *
Request for publishing
Please, send the request only if you are sure that your metadata record includes all mandatory and recommended fields.
Remember that you can use the FAIRness report to verify that your metadata record follows the FAIRness criteria. Use the button "Save & close" to see the detailed view of your metadata record and click on the button "FAIRness Assessment".
If your metadata record does not have a DOI, you can ask for it right now by selecting the following checkbox. The DOIs are assigned only upon verification and validation of the metadata record.
I want to get a DOI
Cancel
Send

Discover & Reuse



Discover & Reuse



“To make progress in science, we need to be open and share.”

THANK YOU!

Ilaria Rosati
ilaria.rosati@cnr.it

CNR IRET
Istituto di Ricerca sugli Ecosistemi Terrestri

Consiglio Nazionale delle Ricerche

UNIVERSITÀ DEL SALENTO

LifeWatch ITALIA

French BIF - Introduction to Darwin Core

Guillaume BODY | Sophie PAMERLON

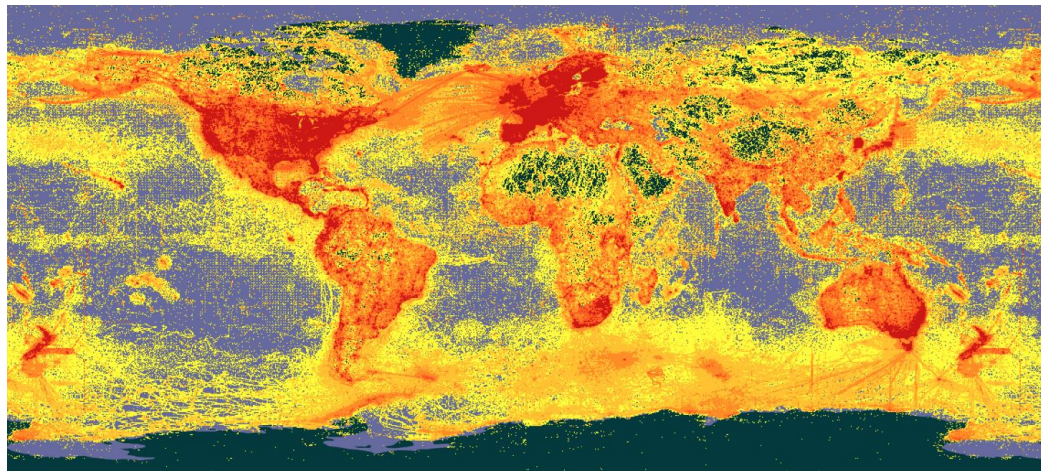
Co-funded by the European Union

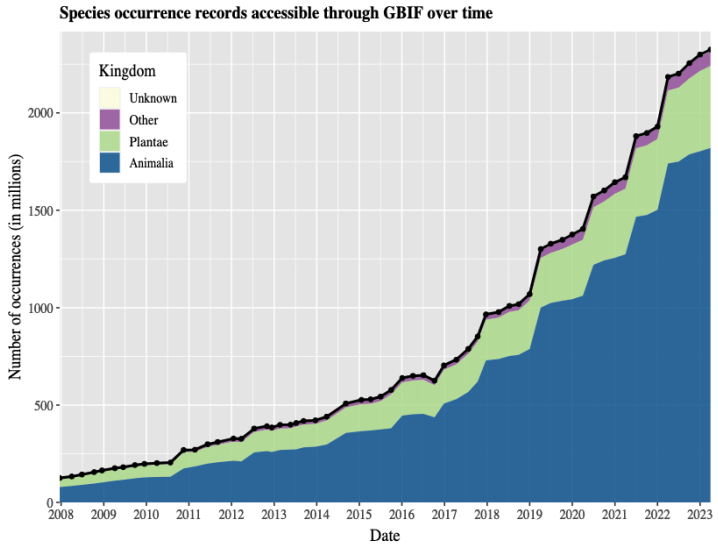
Global Biodiversity Information Facility (GBIF)

- Intergovernmental programme & data infrastructure, created in 2001 by the OECD scientific committee
- Objectives: promote & facilitate the **free and open access to biodiversity data**

More than **2,9 mds data,** cited in **10 229** scientific publications | 173,8 mds data downloaded per month
→ A window on where, when and by whom species were recorded

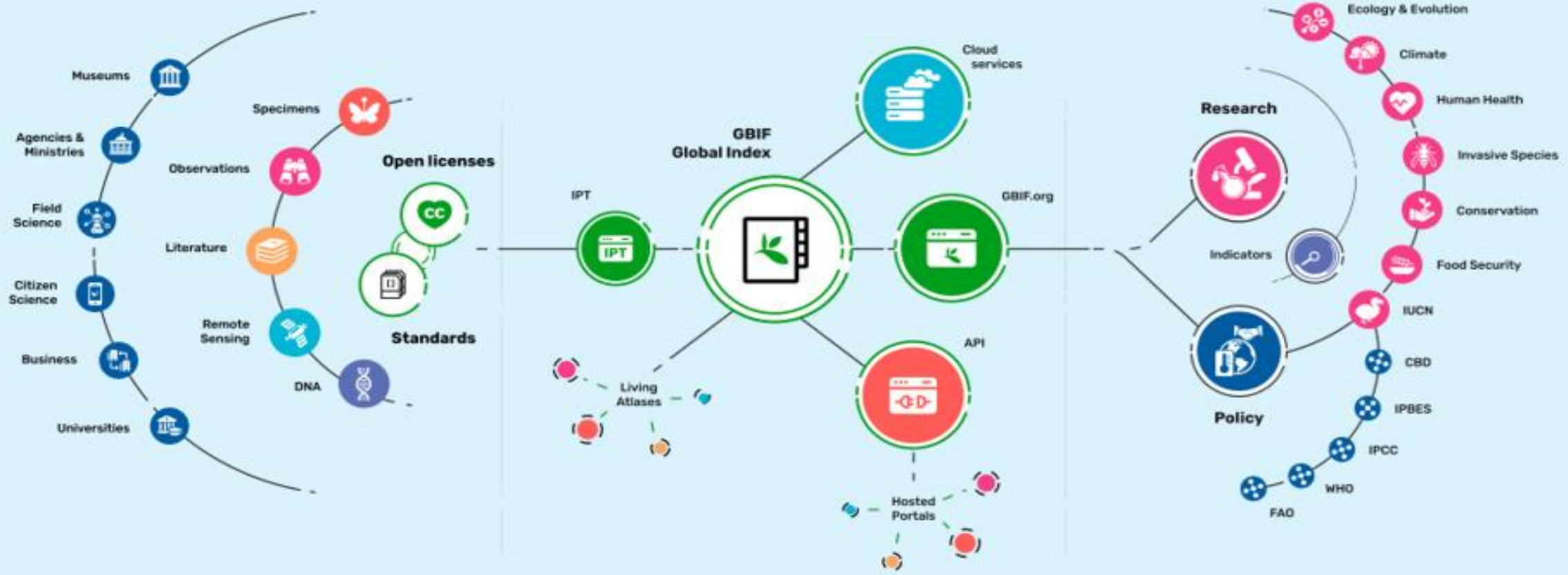
- Collaboration through an agreement protocol
- **105 members** (62 countries, 43 associate organisations)
- **142 contributing countries**
- **2170 data publishers**
- Secretariat based in Copenhagen, Denmark





PatriNat
CENTRE D'EXPERTISE ET DE DONNÉES
OFB·MNHN·CNRS·IRD

PROVIDING BIODIVERSITY EVIDENCE FOR RESEARCH AND POLICY



STANDARDS: LET'S AGREE TO AGREE

> Standardisation does not mean that we all wear the same color and weave of cloth, eat standard sandwiches, or live in standard rooms with standard furnishing. Homes of infinite variety of design are built with a few types of bricks, and with lumber of standard sizes, and with water and heating pipes and fitting of standard dimensions.
>
> W. Edwards Deming

GBIF Secretariat (2021) GBIF Biodiversity Data Mobilization Course. 12th edition. GBIF Secretariat: Copenhagen. https://doi.org/10.35035/ce-c6cr-6w42.

WHAT IS A STANDARD?

An agreed way of doing something

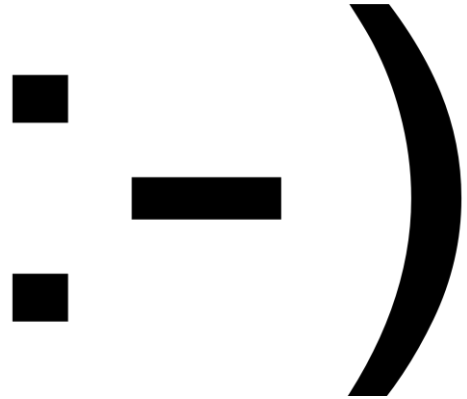
GBIF Secretariat (2021) GBIF Biodiversity Data Mobilization Course. 12th edition. GBIF Secretariat: Copenhagen. https://doi.org/10.35035/ce-c6cr-6w42.

EVERYDAY STANDARDS

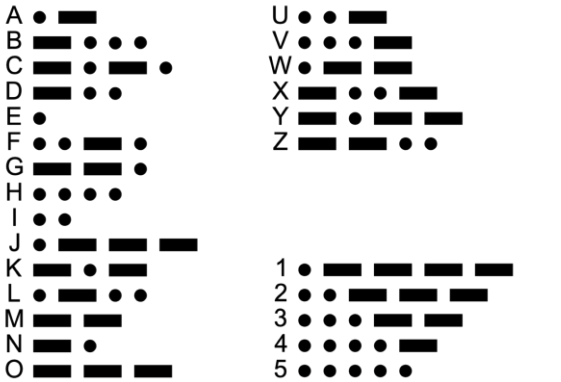
Units of Measurement

(Metric, Imperial)

Alphabets



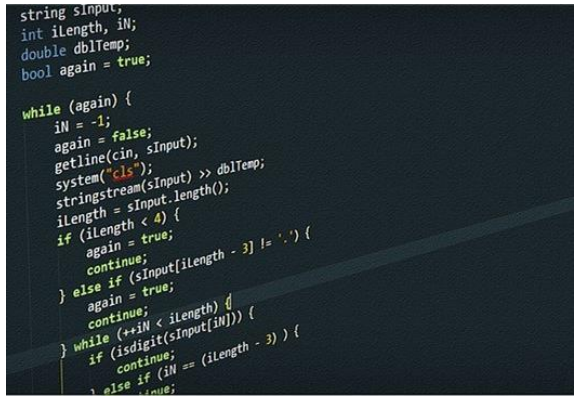
Emojis



Morse Code

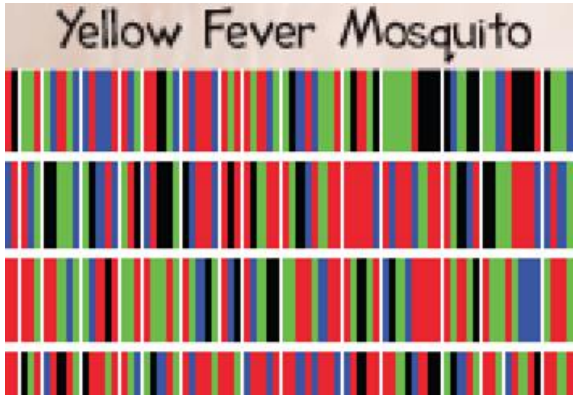
Numeral Systems

(Hindu-Arabic; Roman Numerals)



Languages

Postal addressing



Barcoding

GBIF Secretariat (2021) GBIF Biodiversity Data Mobilization Course. 12th edition. GBIF Secretariat: Copenhagen. https://doi.org/10.35035/ce-c6cr-6w42.

RULES AND RESTRICTIONS

- Type of data
  - restrictions on the category of the field
- Encoding schema
  - restrictions on the range of values in the field
- Format
  - restrictions on the representation of the data
- Character encoding
  - rules for interpreting bytes

Image by Bird Explorers

STANDARDS FOR DATA TRANSFER

- Application schema
  - Specific combinations of data standards for a particular purpose
- Format
  - Restrictions in the dataset structure
- Transfer protocol
  - Where and how to send content

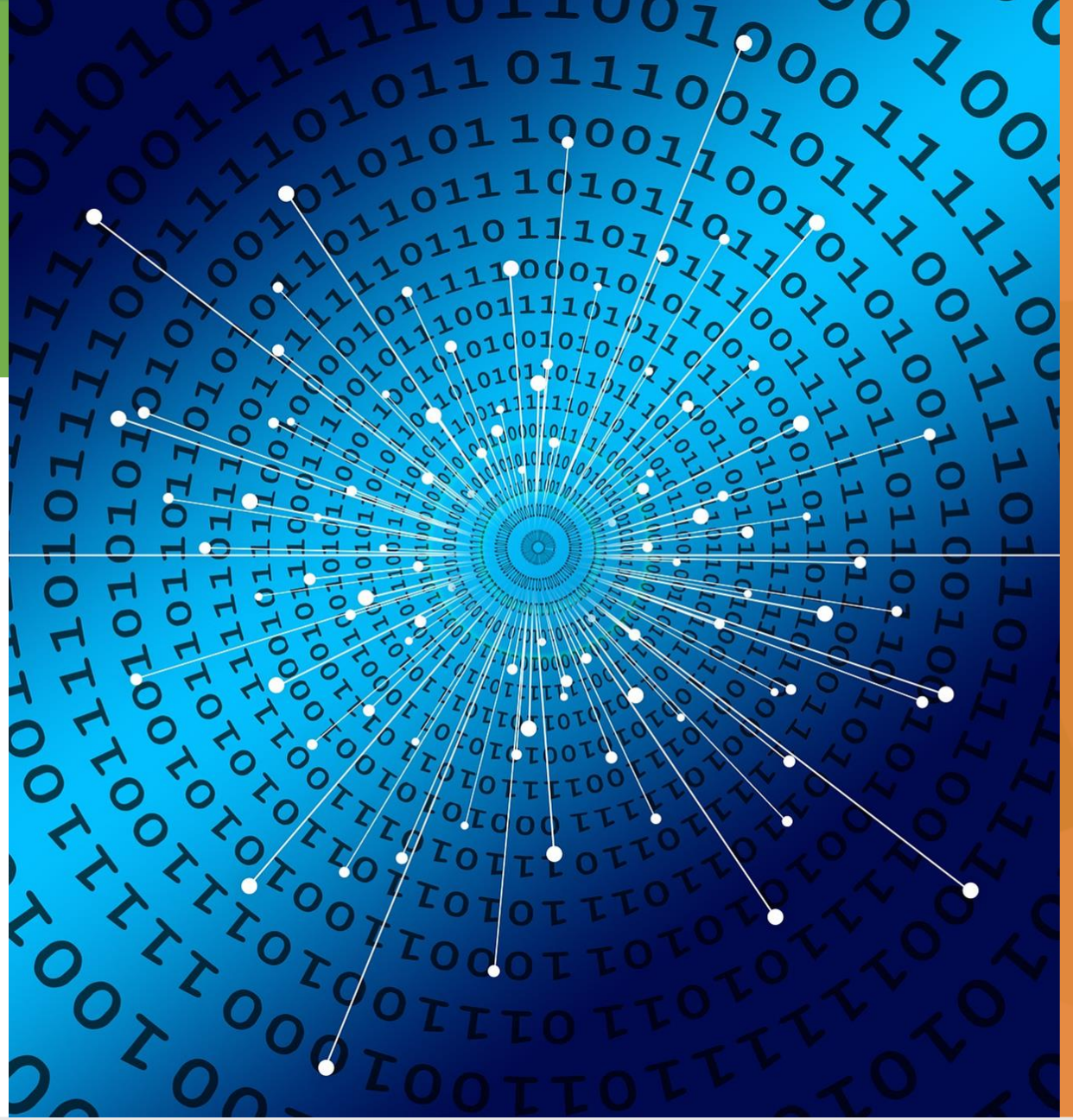


Image by Gerd Altmann

BIODIVERSITY INFORMATION STANDARDS

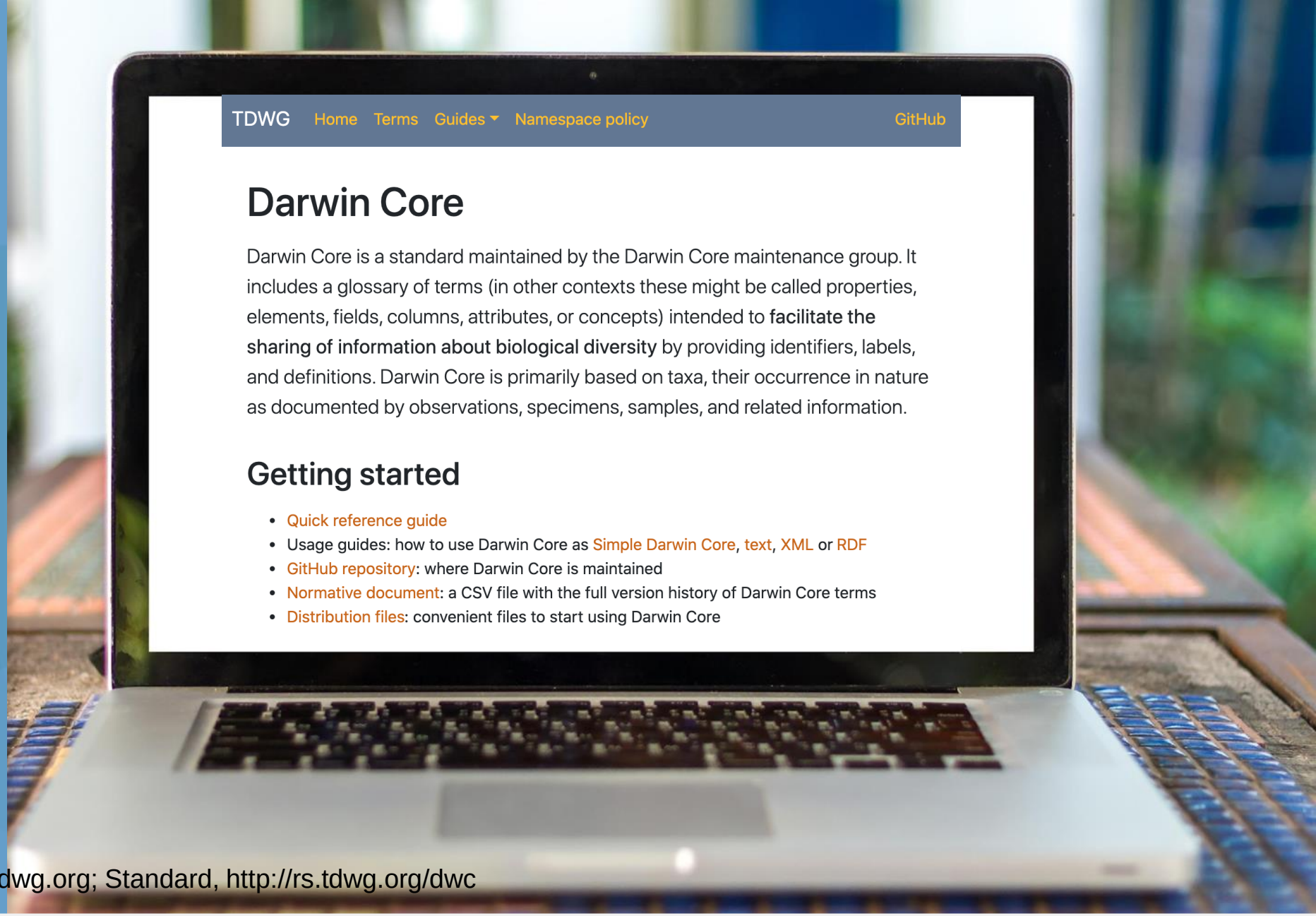
"Data standards are the rules by which data are described and recorded. In order to share, exchange, and understand data, we must standardize the format as well as the meaning." (USGS)

Ecological Metadata Language Standard (EML)
Audubon Media Description (aka Audubon Core)
Global Genome Biodiversity Network(GGBN)
Ocean Data Standards and Best Practices Project (ODSBP)
Darwin Core

GBIF Secretariat (2021) GBIF Biodiversity Data Mobilization Course. 12th edition. GBIF Secretariat: Copenhagen. https://doi.org/10.35035/ce-c6cr-6w42.

WHAT IS DARWIN CORE?

"List of fields and their definitions, as they relate to biodiversity data."



Reference(s): Governance, http://www.tdwg.org; Standard, http://rs.tdwg.org/dwc

SIMPLE DARWIN CORE

Field classes

- Record & Dataset
- Occurrence
- Organism
- Material Sample
- Event
- Location
- Geological Context
- Identification
- Taxon

Auxiliary classes:

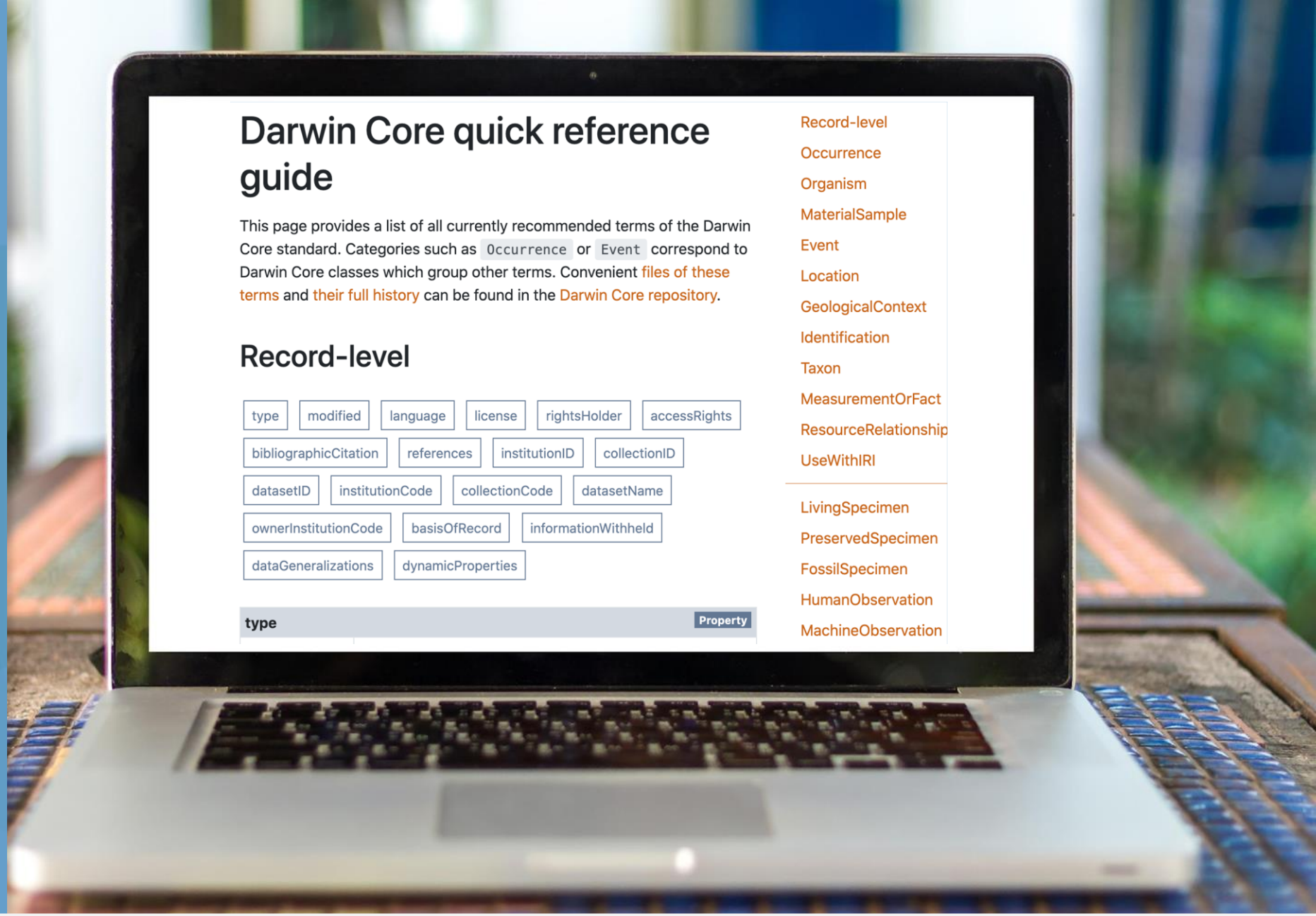
- ResourceRelationship
- MeasurementOrFact

https://dwc.tdwg.org/simple/
Image by Vladimir Tkalčić

GBIF Secretariat (2021) GBIF Biodiversity Data Mobilization Course. 12th edition. GBIF Secretariat: Copenhagen. https://doi.org/10.35035/ce-c6cr-6w42.

DWC QUICK REFERENCE GUIDE

https://dwc.tdwg.org/terms/



GBIF Secretariat (2021) GBIF Biodiversity Data Mobilization Course. 12th edition. GBIF Secretariat: Copenhagen. https://doi.org/10.35035/ce-c6cr-6w42.

DWC TERMS: BASISOFRECORD

basisOfRecord	Property
Identifier	http://rs.tdwg.org/dwc/terms/basisOfRecord
Definition	The specific nature of the data record.
Comments	Recommended best practice is to use the standard label of one of the Darwin Core classes.
Examples	`PreservedSpecimen`, `FossilSpecimen`, `LivingSpecimen`, `MaterialSample`, `Event`, `HumanObservation`, `MachineObservation`, `Taxon`, `Occurrence`

- **Record-level**
- Occurrence
- Organism
- MaterialSample
- Event
- Location

https://dwc.tdwg.org/terms/#dwc:basisOfRecord

GBIF Secretariat (2021) GBIF Biodiversity Data Mobilization Course. 12th edition. GBIF Secretariat: Copenhagen. https://doi.org/10.35035/ce-c6cr-6w42.

DWC TERMS: OCCURRENCEID

occurrenceID	Property
Identifier	http://rs.tdwg.org/dwc/terms/occurrenceID
Definition	An identifier for the Occurrence (as opposed to a particular digital record of the occurrence). In the absence of a persistent global unique identifier, construct one from a combination of identifiers in the record that will most closely make the occurrenceID globally unique.
Comments	Recommended best practice is to use a persistent, globally unique identifier.
Examples	`http://arctos.database.museum/guid/MSB:Mamm:233627`, `000866d2-c177-4648-a200-ead4007051b9`, `urn:catalog:UWBM:Bird:89776`

Record-level

- **Occurrence**
- Organism
- MaterialSample
- Event
- Location
- GeologicalContext
- Identification

https://dwc.tdwg.org/terms/#dwc:occurrenceID

GBIF Secretariat (2021) GBIF Biodiversity Data Mobilization Course. 12th edition. GBIF Secretariat: Copenhagen. https://doi.org/10.35035/ce-c6cr-6w42.

DWC TERMS: COUNTRY AND COUNTRYCODE

country	Property
Identifier	http://rs.tdwg.org/dwc/terms/country
Definition	The name of the country or major administrative unit in which the Location occurs.
Comments	Recommended best practice is to use a controlled vocabulary such as the Getty Thesaurus of Geographic Names.
Examples	`Denmark`, `Colombia`, `España`

countryCode	Property
Identifier	http://rs.tdwg.org/dwc/terms/countryCode
Definition	The standard code for the country in which the Location occurs.
Comments	Recommended best practice is to use an ISO 3166-1-alpha-2 country code.
Examples	`AR`, `SV`

- MaterialSample
- Event
- **Location**
- GeologicalContext
- Identification
- Taxon
- MeasurementOrFact
- ResourceRelationship
- UseWithIRI

- LivingSpecimen
- PreservedSpecimen

https://dwc.tdwg.org/terms/#dwc:country; https://dwc.tdwg.org/terms/#dwc:countryCode

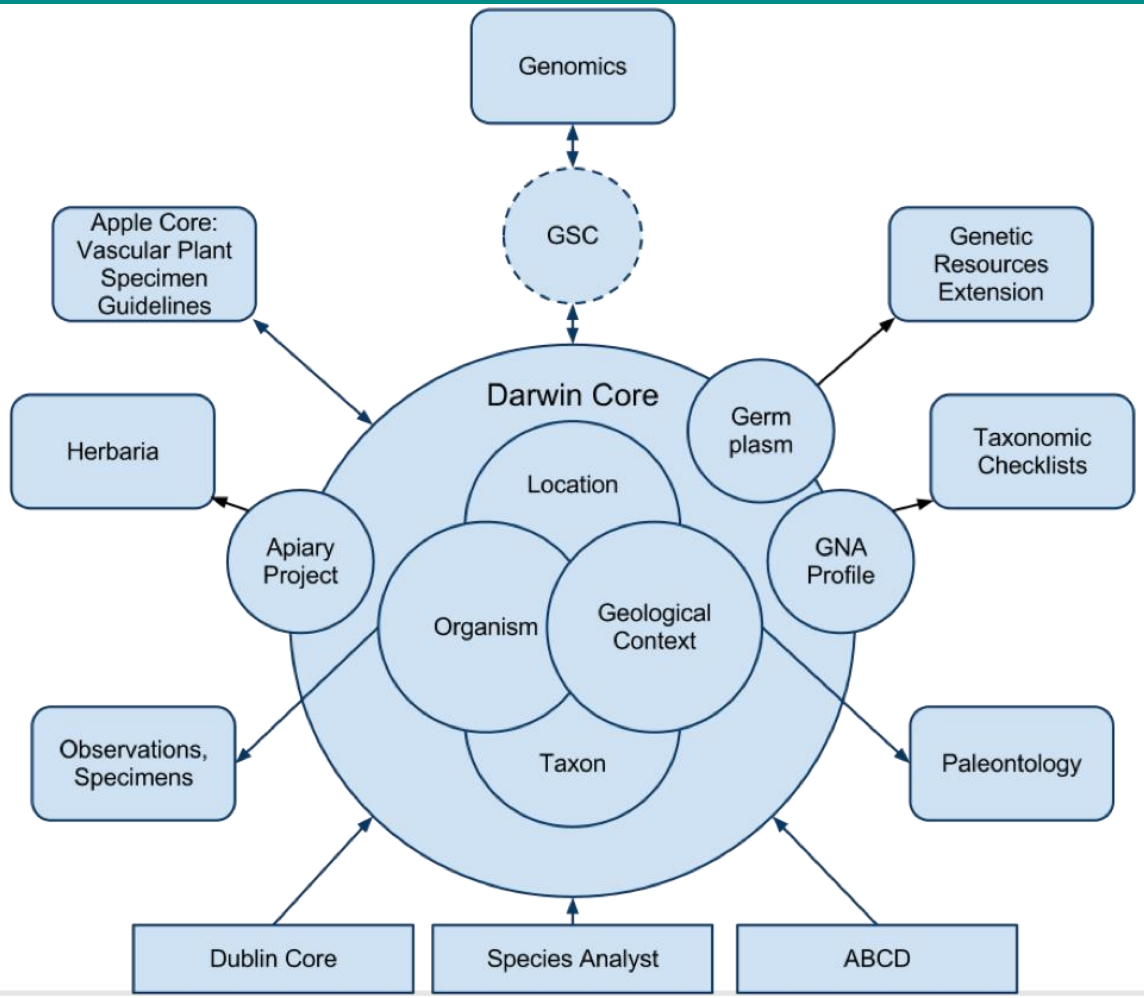
GBIF Secretariat (2021) GBIF Biodiversity Data Mobilization Course. 12th edition. GBIF Secretariat: Copenhagen. https://doi.org/10.35035/ce-c6cr-6w42.

DARWIN CORE EXTENSIONS

- Audubon Media Description (aka Audubon Core)
- Measurements or Facts
- Identification History
- And many more!

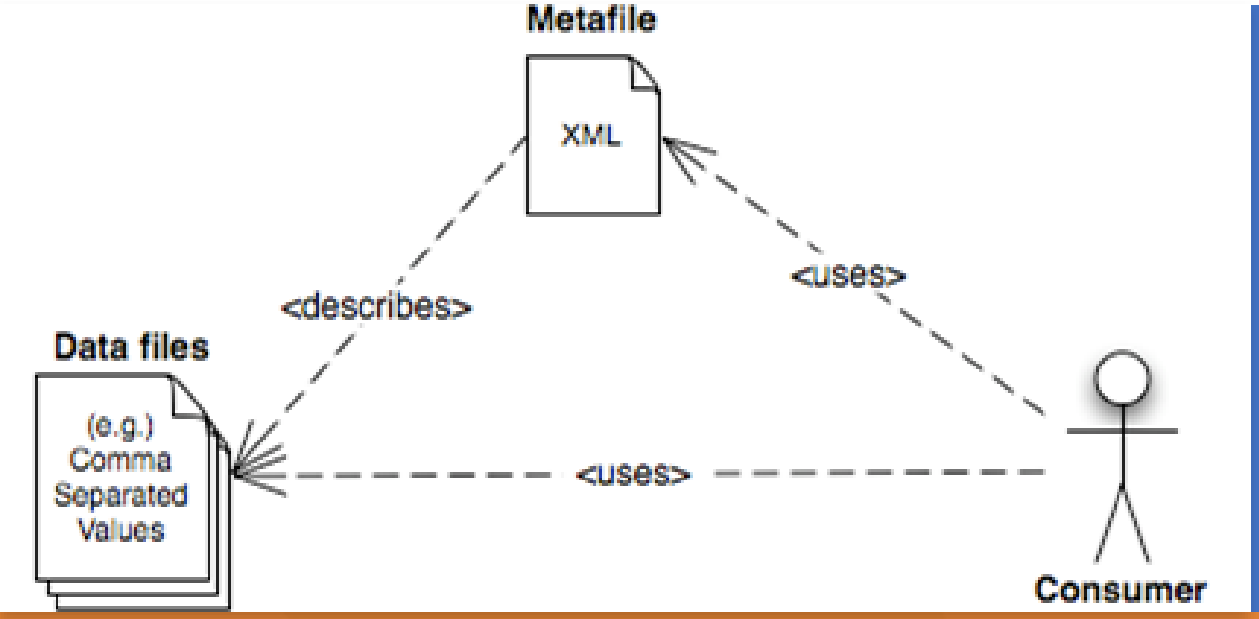
https://tools.gbif.org/dwca-validator/extensions.do
Image by Elizabeth Byers

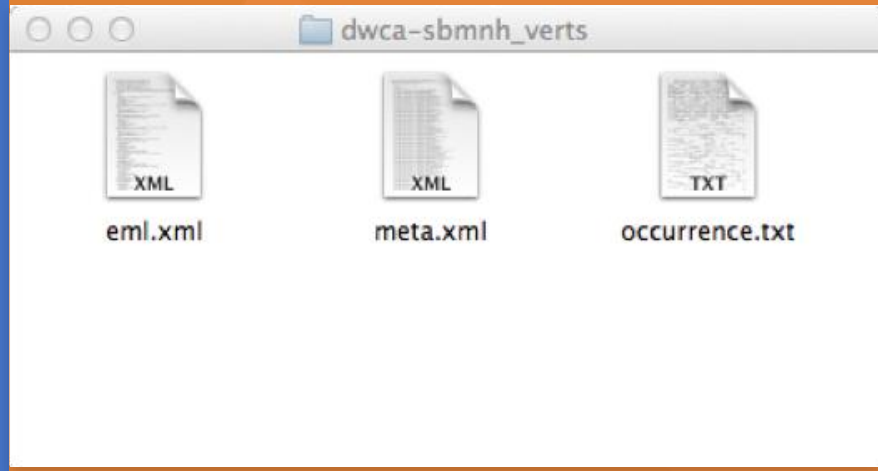
COMMUNITY AND STANDARDS RELATIONSHIPS



GBIF Secretariat (2021) GBIF Biodiversity Data Mobilization Course. 12th edition. GBIF Secretariat: Copenhagen. https://doi.org/10.35035/ce-c6cr-6w42.

DARWIN CORE ARCHIVES

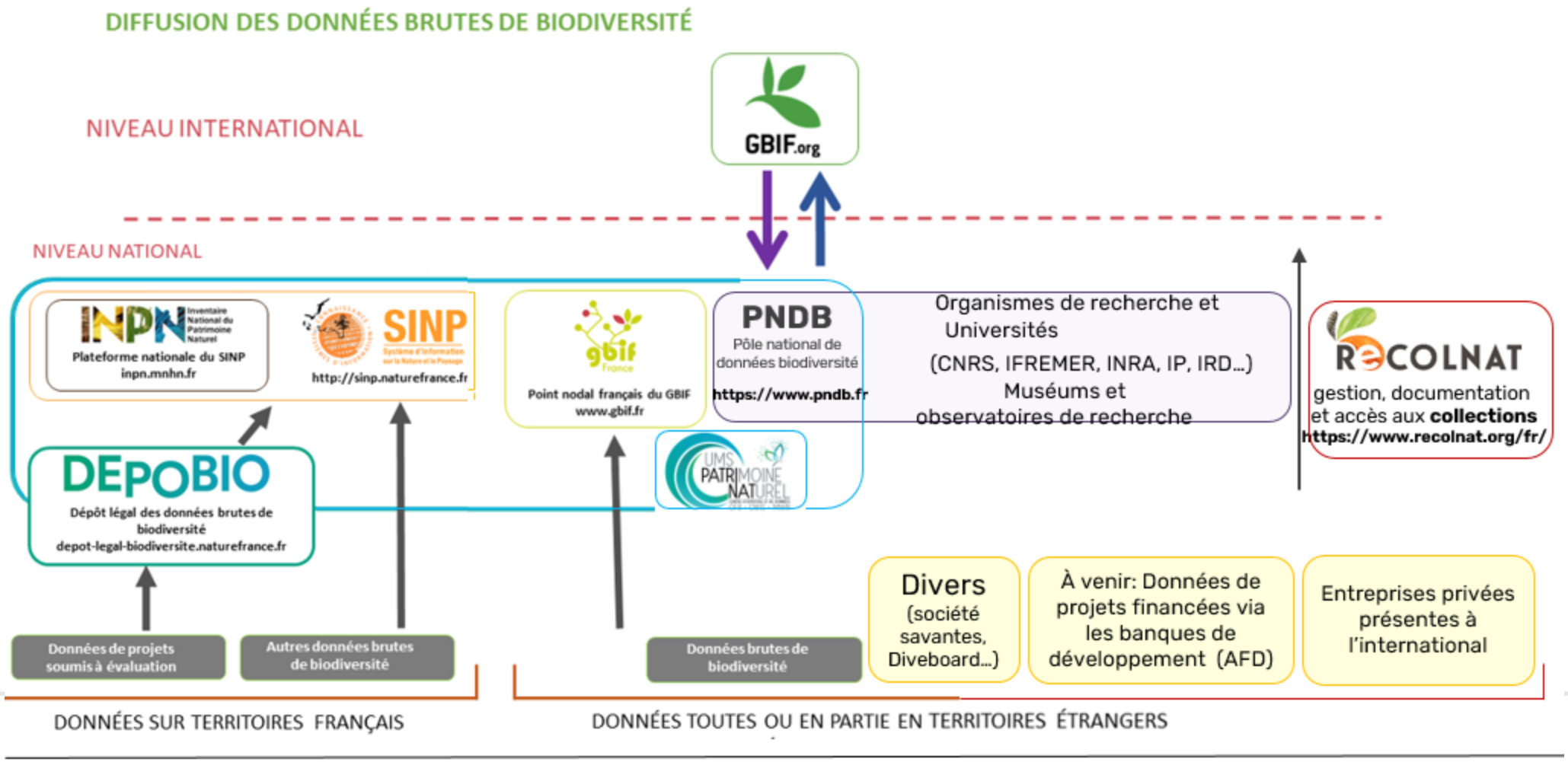




A DwC Archive is an expression of the Darwin Core text guide. It is a compressed file containing a minimum of three files. It is encoded as UTF-8.

https://dwc.tdwg.org/text/

FRENCH NATIONAL DATA WORKLOWS



FRENCH NATIONAL DATA WORKLOWS

PNDB portal (research)

GBIF-compatible data

GBIF.org

metadata resources by project

metadata resources

SINP standard occurrence data

DwC occurrence data

SIB catalog (public policies)

SINP metada resources

SINP / INPN (national biodiversity data agregator)

functional workflows
WIP workflows
TBI workflows

Darwin core and monitoring data

Electrofishing protocol

Darwin core and monitoring data

Electrofishing monitoring data

Code Sandre de la station	Nom de la station	Code ASPE du point de prélèvement	Code Sandre du point de prélèvement	Numéro d'opération	Date de l'opération	Niveau de validation	Qualification de la donnée	Bassin hydrographique	Cours d'eau
4613007	R DEME A LOUESTAULT		6	90170	12/10/23	Validé niveau 1	Correcte	LOIRE-BRETAGNE	la Dême
4613007	R DEME A LOUESTAULT		6	90170	12/10/23	Validé niveau 1	Correcte	LOIRE-BRETAGNE	la Dême
4613007	R DEME A LOUESTAULT		6	90170	12/10/23	Validé niveau 1	Correcte	LOIRE-BRETAGNE	la Dême

Code de l'entité hydrographique	Région	Département	Commune	Lieu-dit	Coordonnées X du point de prélèvement	Coordonnées Y du point de prélèvement	Système de projection des coordonnées du point	Objectif de pêche	Espèce ciblée
M1345800	CENTRE-VAL-DE-LOIRE	Indre-et-Loire	BEAUMONT-LA-RONCE	PONT DE LA D29 - LIEU-DIT MOQUE-SOURIS	524028.00000000	6728114.00000000	RGF93 / Lambert 93	RRP – Réseau de Référence Pérenne	
M1345800	CENTRE-VAL-DE-LOIRE	Indre-et-Loire	BEAUMONT-LA-RONCE	PONT DE LA D29 - LIEU-DIT MOQUE-SOURIS	524028.00000000	6728114.00000000	RGF93 / Lambert 93	RRP – Réseau de Référence Pérenne	
M1345800	CENTRE-VAL-DE-LOIRE	Indre-et-Loire	BEAUMONT-LA-RONCE	PONT DE LA D29 - LIEU-DIT MOQUE-SOURIS	524028.00000000	6728114.00000000	RGF93 / Lambert 93	RRP – Réseau de Référence Pérenne	

Protocole de pêche	Numéro du Passage	Type de point	Nom de l'ambiance	Surface échantillonnée (m²)	Identifiant du lot	Numéro du lot	Type du lot	Effectif de la ligne	Effectif du lot
Pêche complète à un ou plusieurs passages	1			351.00	5648513	1	N	1	1
Pêche complète à un ou plusieurs passages	1			351.00	5648575	63	I	1	3
Pêche complète à un ou plusieurs passages	1			351.00	5648569	57	S/L	1	158

Poids du lot (g)	Taille maximale des individus du lot (mm)	Taille minimale des individus du lot (mm)	Type de longueur du lot	Identifiant du poisson	Taxon - Code alternatif	Taxon - Code taxref	Taxon - Nom commun	Taille de l'individu (mm)	Mesure réelle de la taille (non dégroupée)	Poids du poisson (g)	Présence de pathologie	Sexe de l'individu	Opération confidentielle
36			Totale	22483116	TRF	67778	Truite de riviere	162	Oui		Non		Non
34			Totale	22483352	LOF	67550	Loche franche	56	Oui		Non		Non
864			Totale	22483172	VAI	67404	Vairon	79	Oui		Non		Non

40 columns

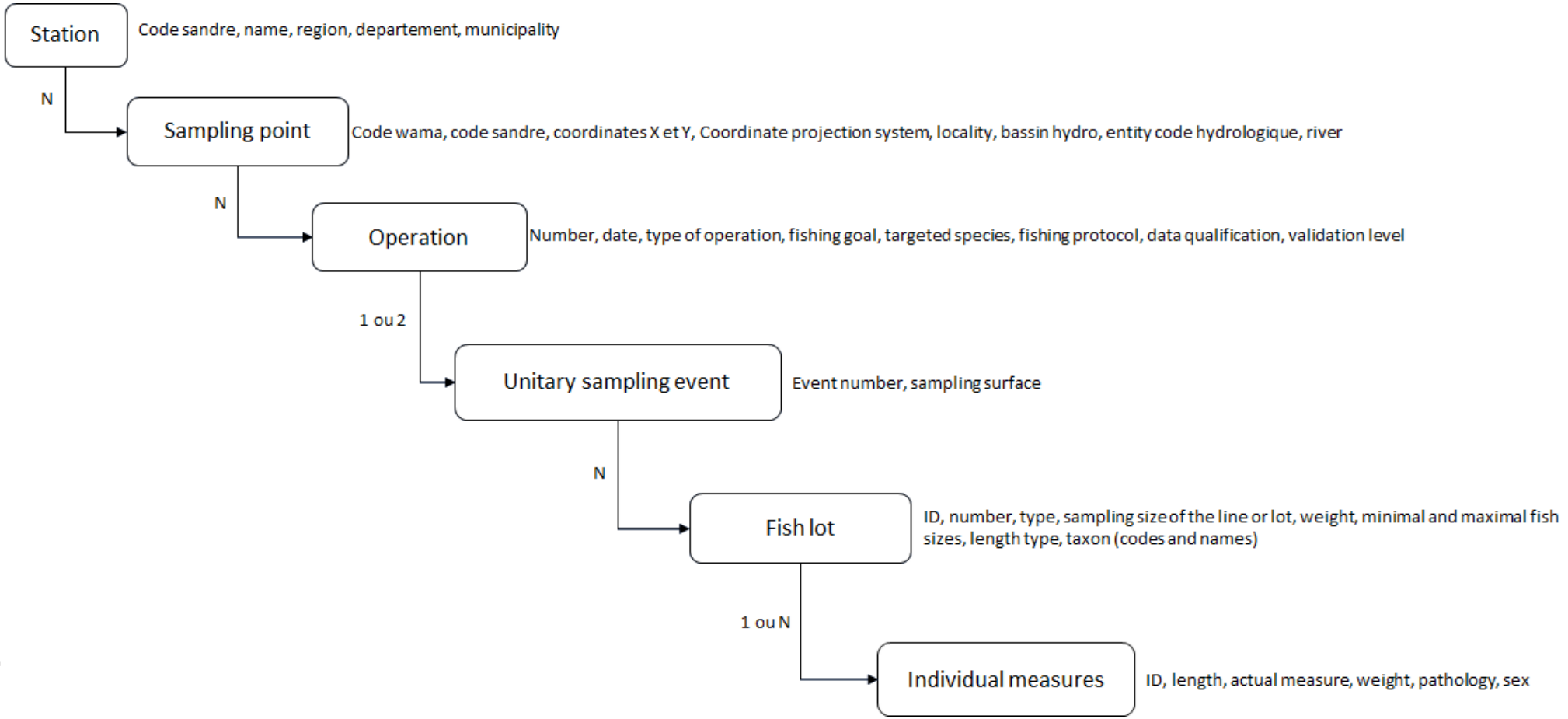
Figure 5: Sample of electrofishing raw data extracted from ASPE database (data has been subdivided for easier reading)

Darwin core and monitoring data

Electrofishing protocol

Darwin core and monitoring data

Electrofishing protocol



Event core:
for sampling event description

eventID	term2	term3
value	value	value

Occurrence extension:
for observations

occurenceID	term2	term3
value	value	value

Humbold extension:
for protocol details

eventID	term2	term3
value	value	value

Measurement of fact extension:
for measures or other facts linked to an event

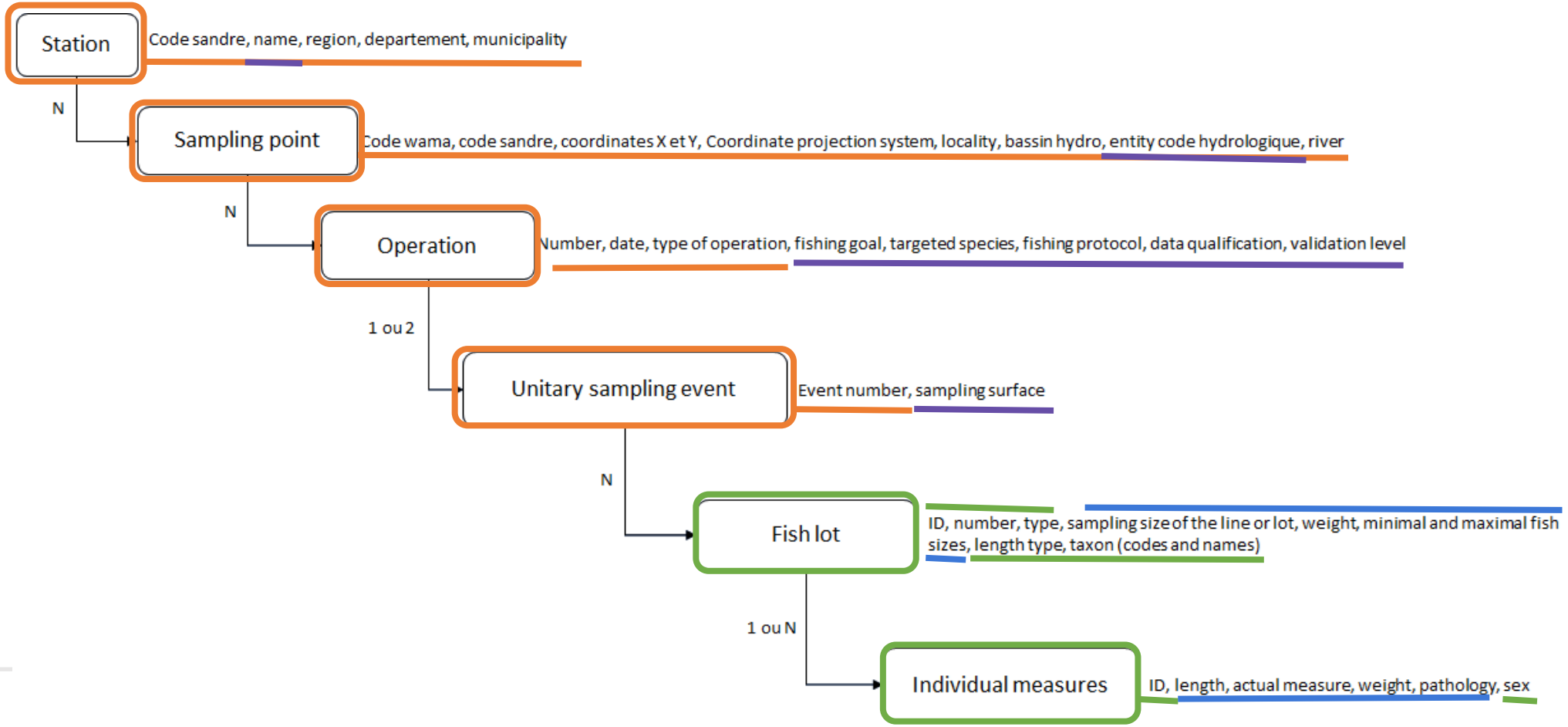
measurementID	measurementType	measurementValue	measurementUnit
value	value	value	value

Measurement of fact extension:
for measures or other facts linked to an observation

measurementID	measurementType	measurementValue	measurementUnit
value	value	value	value

Darwin core and monitoring data

Electrofishing protocol



Electrofishing protocol

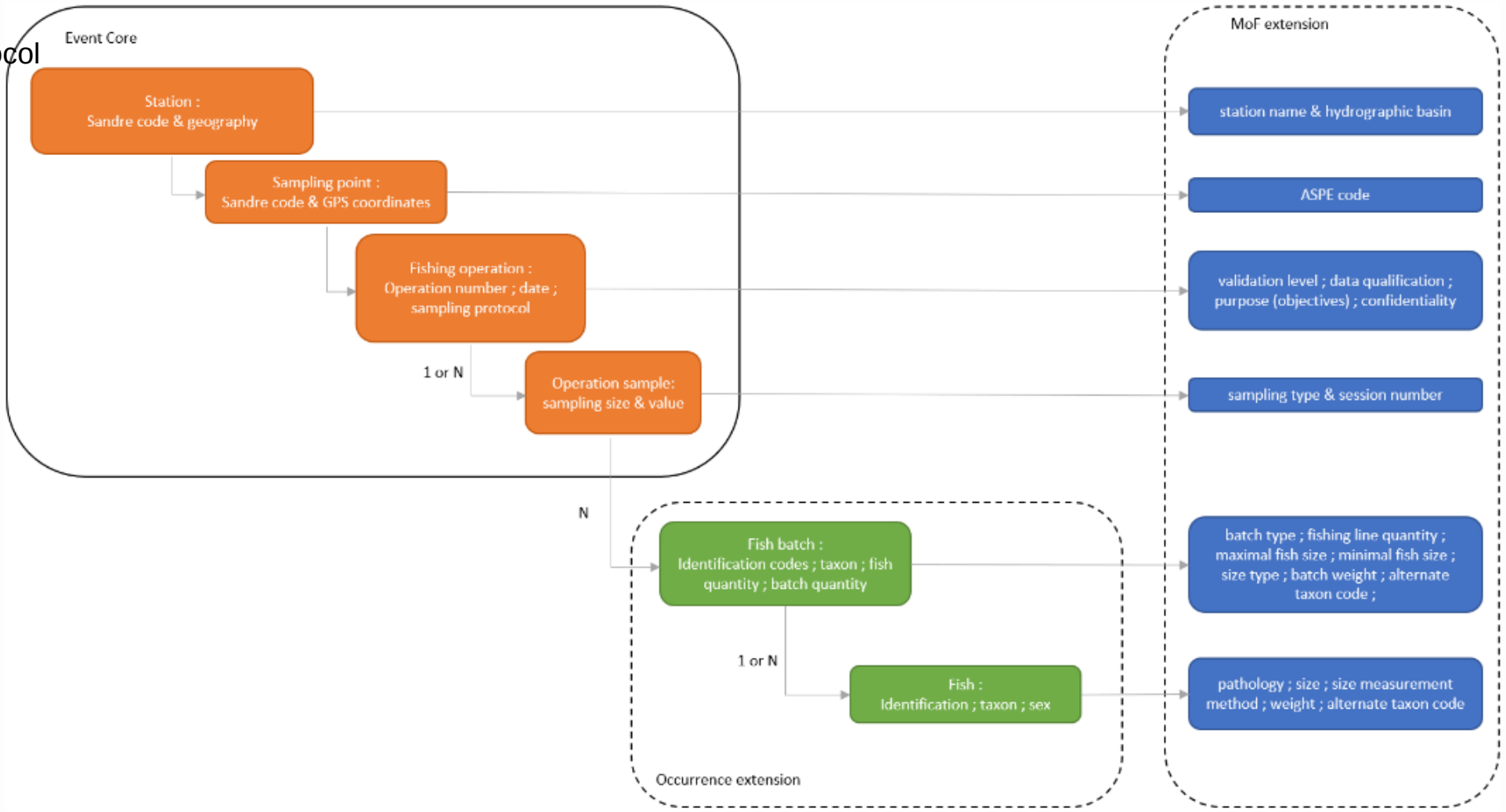


Figure 6: Scheme of variables in electrofishing data as DwC terms

eventID	locationID	waterBody	stateProvince	county	municipality	verbatimLocality
event1	4613007	La Dême	Centre-Val-de-Loire	Indre-et-Loire	Beaumont-La-Ronce	Pont de la D29 – Lieu-dit Moque-Souris

eventID	parentEventID	locationID	decimalLatitude	decimalLongitude	geodeticDatum
event1:1	event1	6	524028.00000000	6728114.00000000	EPSG:2154

eventID	parentEventID	fieldNumber	eventDate	samplingProtocol	sampleSizeValue	sampleSizeUnit
event1:1:1	event1:1	90170	23-10-12	Complete electrofishing with one or more sessions	351.00	square meter

eventID	parentEventID
event1:1:1:1	event1:1:1

measurementID	eventID	measurementType	measurementValue
event1/mea1	event1	Station name	R DEME A LOUESTAULT
event1/mea2	event1	Hydrographic basin name	Loire-Bretagne
event1/mea3	event1	Hydrographic entity code	M1345800
event1:1/mea1	event1:1	ASPE (Wama) code	
event1:1:1/mea1	event1:1:1	operation is confidential	No
event1:1:1/mea2	event1:1:1	operation validity level	Validity level 1
event1:1:1/mea3	event1:1:1	data qualification	correct
event1:1:1/mea4	event1:1:1	fishery purpose	RRP – Réseau de Référence Pérenne
event1:1:1:1/mea1	event1:1:1:1	operation sampling type	session
event1:1:1:1/mea2	event1:1:1:1	session number	1

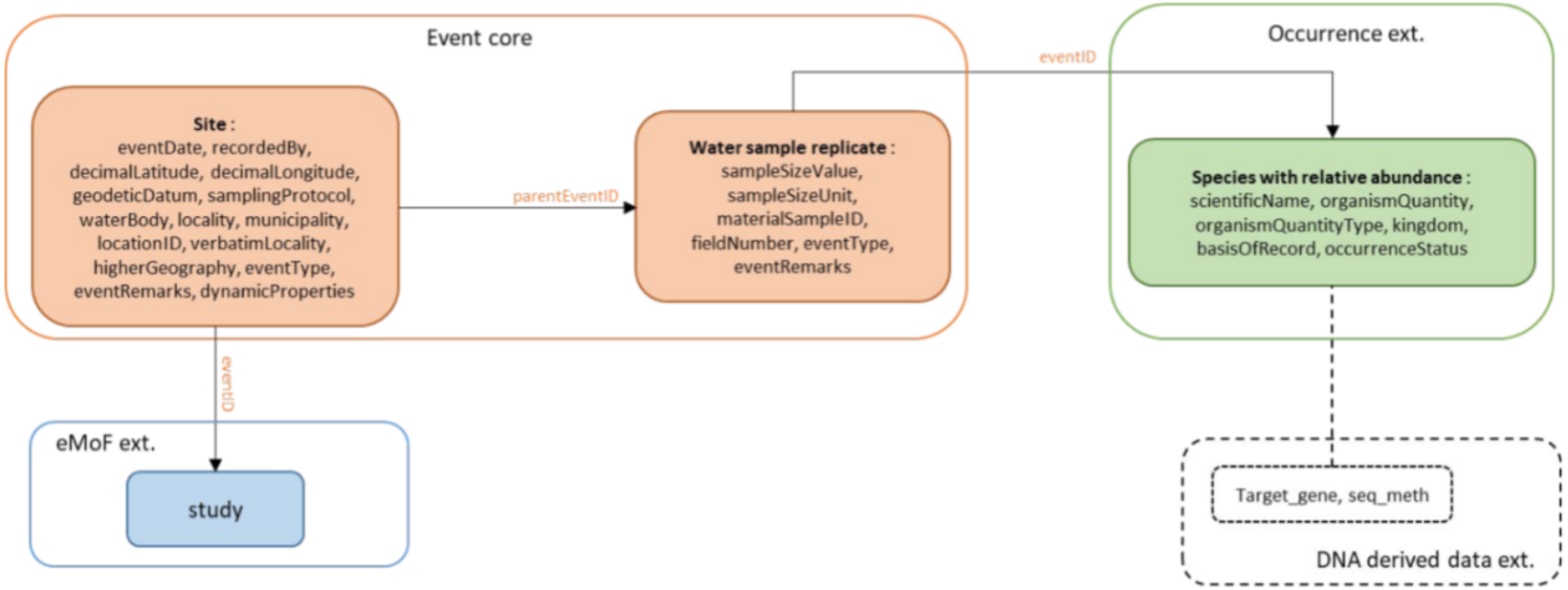
Figure 7: Transcription of the event level as seen in Figure 6

occurrenceID	eventID	catalogNumber	recordNumber	basisOfRecord	occurrenceStatus	scientificName	scientificNameAutorship	taxonID	nameAccordingTo	vernacularName	individualCount
event1:1:1:1/occ1	event1:1:1:1	5648513	1	HumanObservation	present	Salmo trutta fario	Linnaeus, 1758	67778	TaxRef	River trout (riverine ecotype)	1
event1:1:1:1/occ2	event1:1:1:1	5648575	63	HumanObservation	present	Barbatula barbatula	Linnaeus, 1758	67550	TaxRef	Stone loach	3
event1:1:1:1/occ3	event1:1:1:1	5648569	57	HumanObservation	present	Phoxinus phoxinus	Linnaeus, 1758	67404	TaxRef	Common minnow	158

measurementID	occurrenceID	measurementType	measurementValue	measurementUnit
event1:1:1:1/occ1/mea1	event1:1:1:1/occ1	Batch type	N	
event1:1:1:1/occ1/mea2	event1:1:1:1/occ1	Line quantity	1	
event1:1:1:1/occ1/mea3	event1:1:1:1/occ1	Total batch weight	36	g gram
event1:1:1:1/occ1/mea4	event1:1:1:1/occ1	Batch individuals maximal size		mm millimeter
event1:1:1:1/occ1/mea5	event1:1:1:1/occ1	Batch individuals minimal size		mm millimeter
event1:1:1:1/occ1/mea6	event1:1:1:1/occ1	Batch size type	total	
event1:1:1:1/occ1/mea7	event1:1:1:1/occ1	Taxa alternate code	TRF	

occurrenceID	eventID	parentOccurrenceID	individualCount	catalogNumber	basisOfRecord	occurrenceStatus	scientificName	scientificNameAutorship	taxonID	nameAccordingTo	vernacularName	sex
event1:1:1:1/occ1:1	event1:1:1:1	event1:1:1:1/occ1	1	22483116	HumanObservation	present	Salmo trutta fario	Linnaeus, 1758	67778	TaxRef	River trout (riverine ecotype)	
event1:1:1:1/occ2:1	event1:1:1:1	event1:1:1:1/occ2	1	22483352	HumanObservation	present	Barbatula barbatula	Linnaeus, 1758	67550	TaxRef	Stone loach	
event1:1:1:1/occ2:2	event1:1:1:1	event1:1:1:1/occ2	1	22483353	HumanObservation	present	Barbatula	Linnaeus, 1758	67550	TaxRef	Stone loach	
event1:1:1:1/occ2:3	event1:1:1:1	event1:1:1:1/occ2	1	22483354	HumanObservation	present	Barbatula	Linnaeus, 1758	67550	TaxRef	Stone loach	
event1:1:1:1/occ3:1	event1:1:1:1	event1:1:1:1/occ3	1	22483355	HumanObservation	present	Phoxinus	Linnaeus, 1758	67404	TaxRef	Common minnow	
event1:1:1:1/occ3:2	event1:1:1:1	event1:1:1:1/occ3	1	22483356	HumanObservation	present	Phoxinus	Linnaeus, 1758	67404	TaxRef	Common minnow	

measurementID	occurrenceID	measurementType	measurementValue	measurementUnit
event1:1:1:1/occ1:1/mea1	event1:1:1:1/occ1:1	Individual size	162	mm
event1:1:1:1/occ1:1/mea2	event1:1:1:1/occ1:1	Is real size	yes	
event1:1:1:1/occ1:1/mea3	event1:1:1:1/occ1:1	Individual weight		g
event1:1:1:1/occ1:1/mea4	event1:1:1:1/occ1:1	Pathology	no	



eDNA protocole

Citizen science bird count

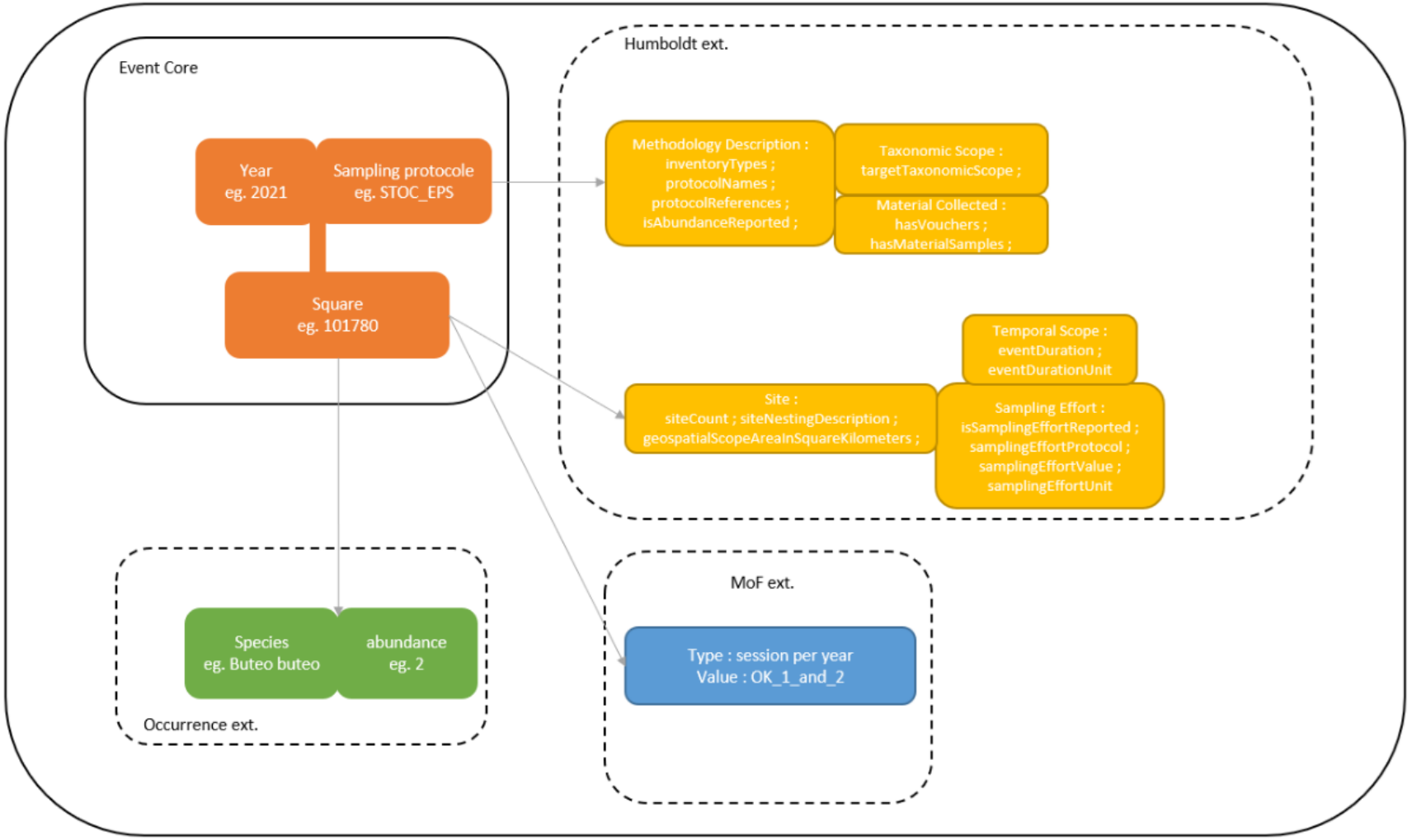


Figure 3 : Scheme of variables in STOC-EPS data as DwC terms

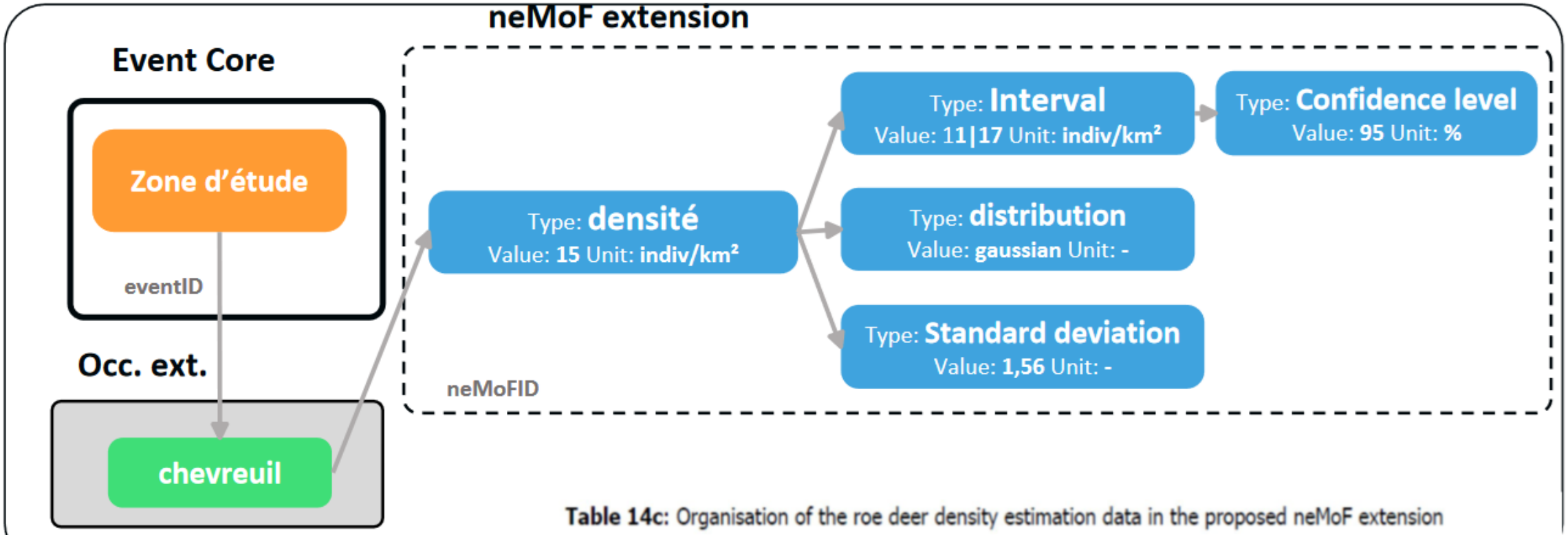
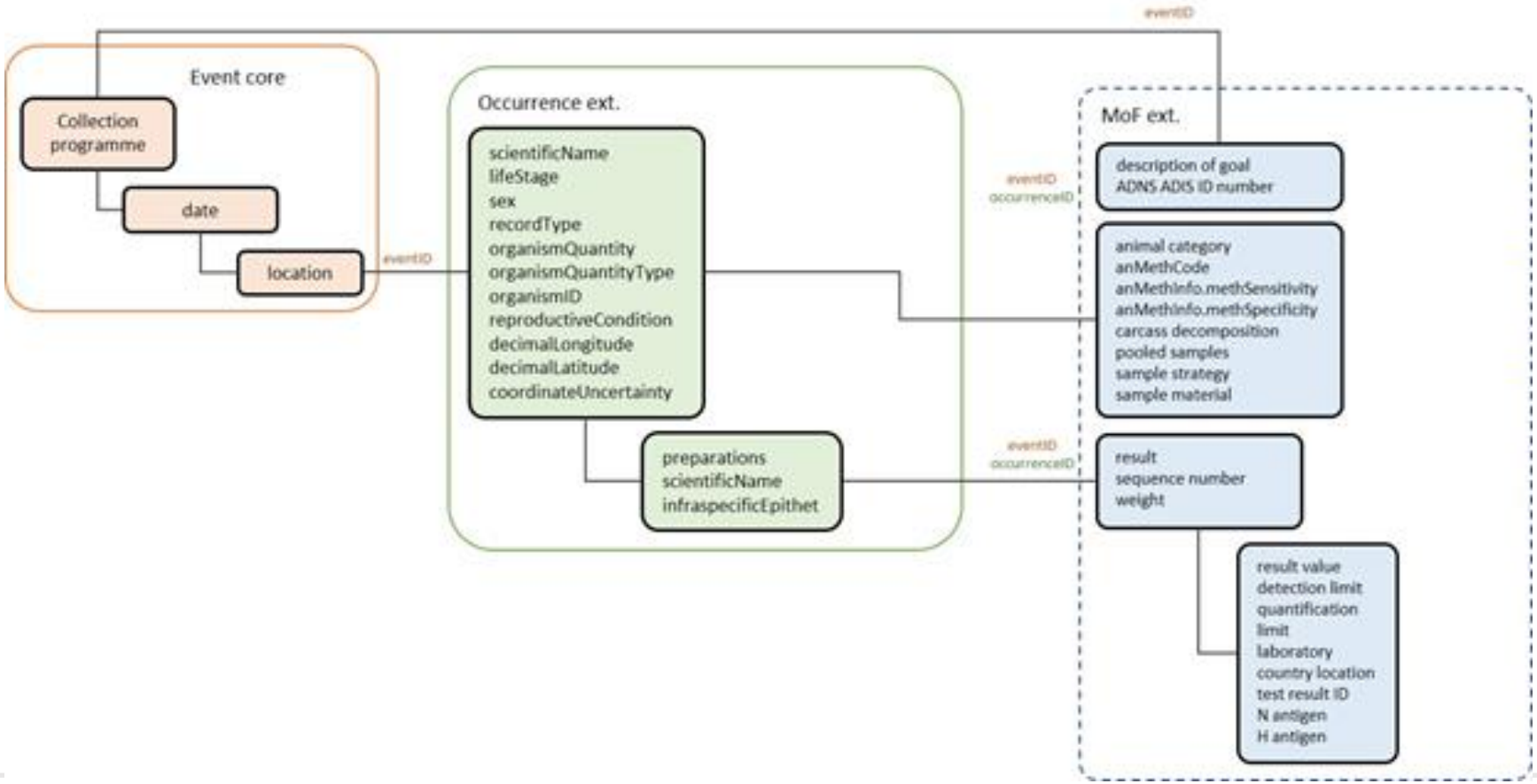


Table 14c: Organisation of the roe deer density estimation data in the proposed neMoF extension

Roe deer density estimate

Epidemiological data: the SIGMA database



Difficulties and advices

It works !! most of the time.

Request recursivity in occurrences and in measurements,

(not yet included in Darwin Core official, nor in GBIF/OBIS, but new GBIF model should allows it)

“EventType” would be useful to clearly state the kind of event we are talking about

Humbolt core to precise/ajust while trying.

Still some issues:

Habitat/Ecosystem approach not so well developed. Darwin core remains taxon oriented

Non biological observation (boats, pollution) not included, but should be easy to develop

Which event first : location or time or sampling => no rules, but easier while following the protocol logic

Graphical metadata useful, and color code useful as well to ease the understanding

More documentation necessary, and guideline for monitoring

Reference:

- De Pooler D. et al. (2017) Toward a new data standard for combined marine biological and environmental datasets - expanding OBIS beyond species occurrences. Biodiversity Data Journal 5: e10989 https://bdj.pensoft.net/articles.php?id=10989

- Enetwild conApplying the Darwin core standard to the monitoring of wildlife species, their management and estimated records: https://doi.org/10.2903/sp.efsa.2020.EN-1841

- Applying the Darwin Core data standard to wildlife disease – advancements toward a new data model https://doi.org/10.2903/sp.efsa.2022.EN-7667

- Darwin core terms: https://dwc.tdwg.org/terms/

Tools to develop: try this Shiny app (and improve it): https://github.com/fja062/WLDM.standardisation

Thank you

sophie.pamerlon@mnhn.fr
guillaume.body@ofb.gouv.fr

Co-funded by the European Union

biodiversa+
European Biodiversity Partnership

Biodiversa+ Capacity building workshop on Darwin Core standard

7th of June from 9:00 to 12:40 CEST

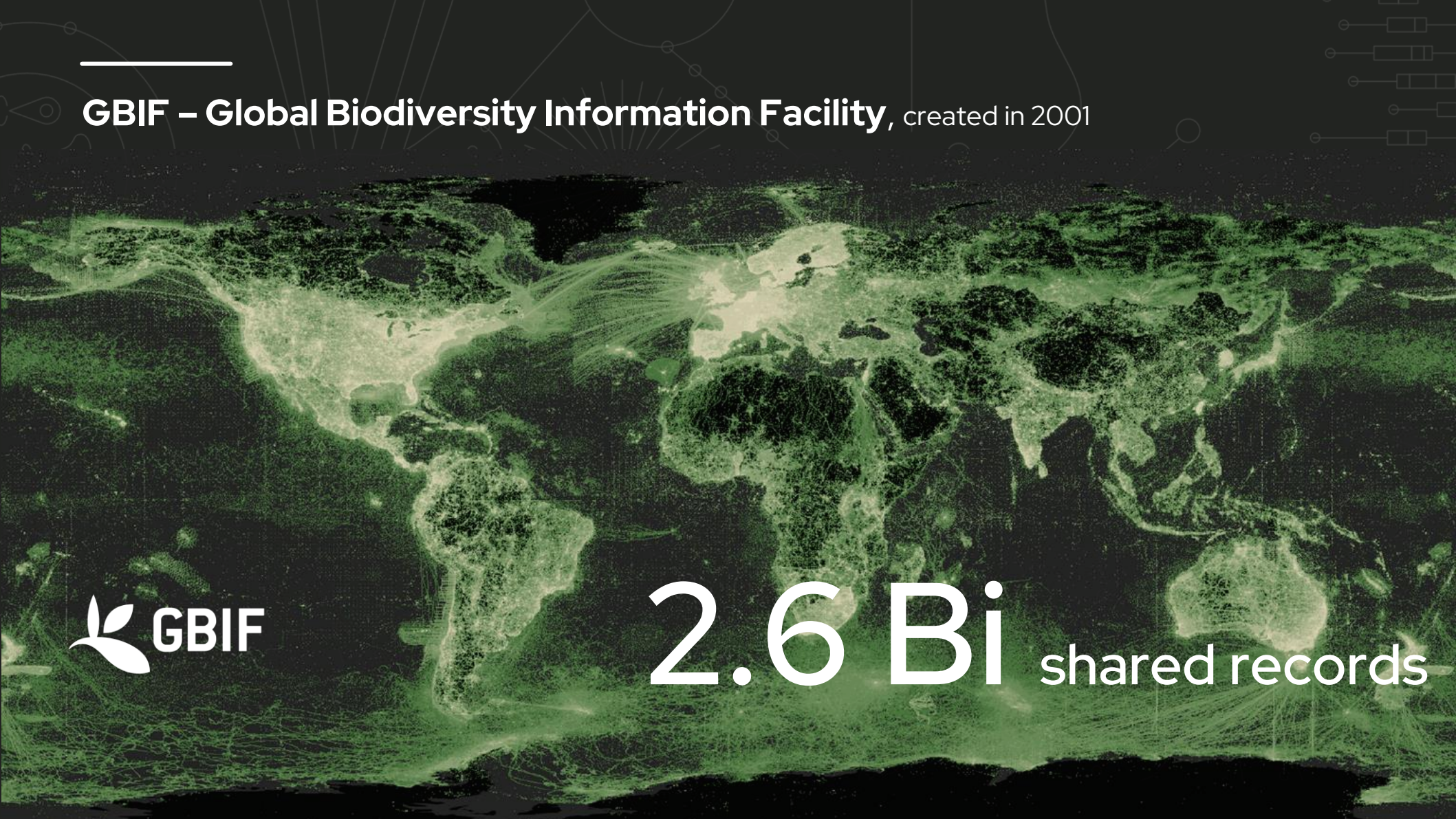
Use of DwC standard in the context of a national participation in GBIF

Rui Figueira
Nó Português do GBIF

INSTITUTO SUPERIOR DE AGRONOMIA
Universidade de Lisboa
Hinc Patriam Sustinet

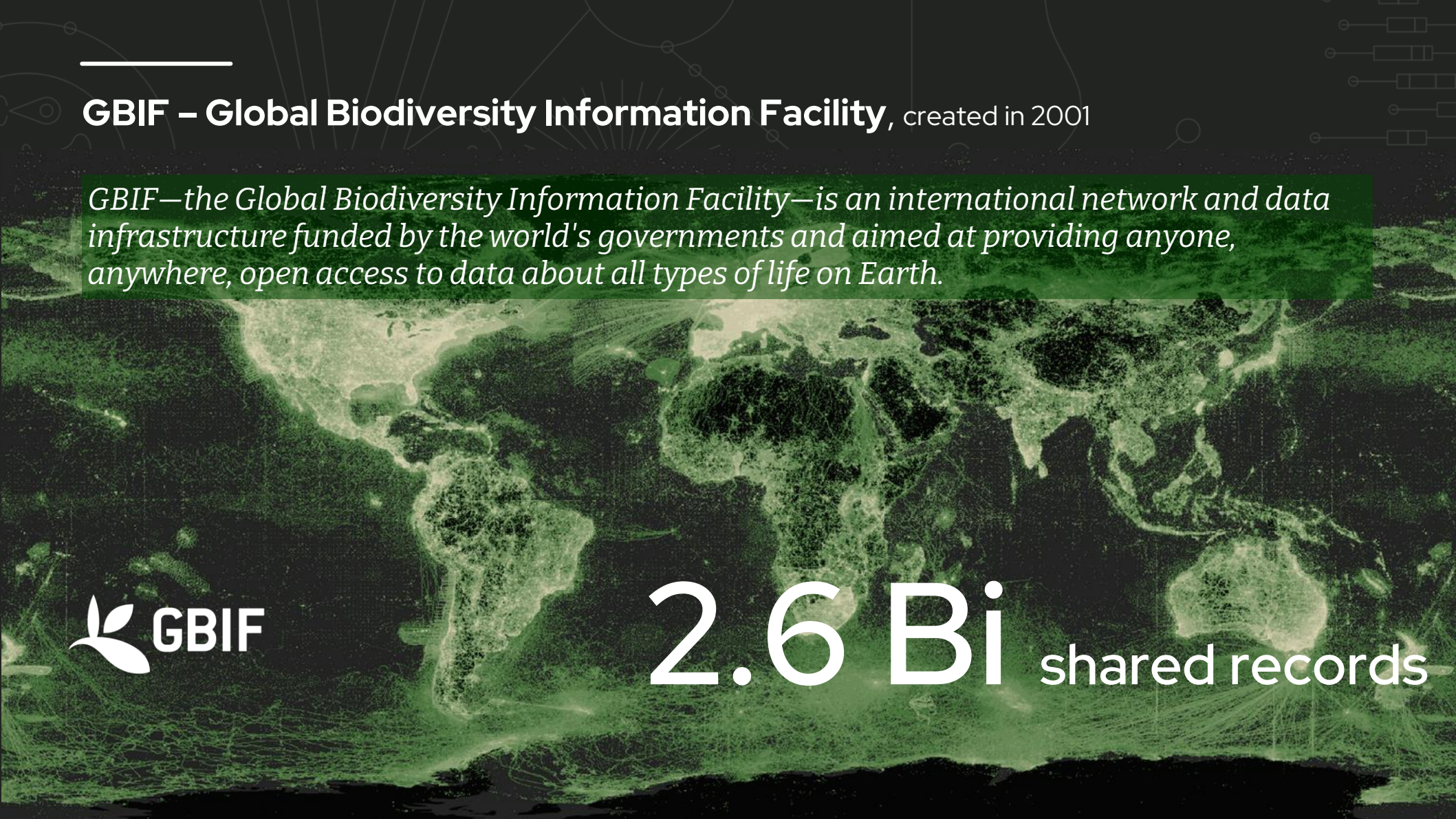
nó português do GBIF

GBIF – Global Biodiversity Information Facility, created in 2001



GBIF – Global Biodiversity Information Facility, created in 2001

GBIF—the Global Biodiversity Information Facility—is an international network and data infrastructure funded by the world's governments and aimed at providing anyone, anywhere, open access to data about all types of life on Earth.



What is GBIF? - example

INVASIVE SPECIES:

Yellow-legged Asian hornet (*Vespa velutina*) expansion in Europe

2005

What is GBIF? - example

INVASIVE SPECIES:

Yellow-legged Asian hornet (*Vespa velutina*) expansion in Europe

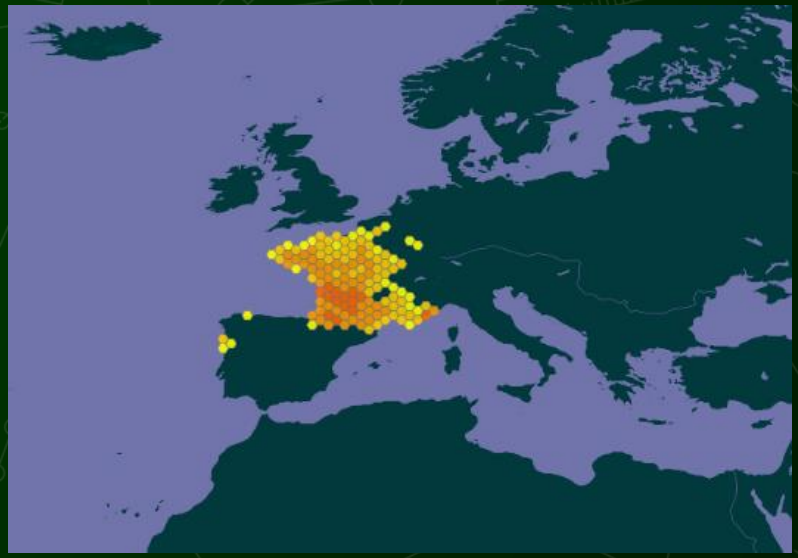
2010

What is GBIF? - example

INVASIVE SPECIES:

Yellow-legged Asian hornet (*Vespa velutina*) expansion in Europe

2015

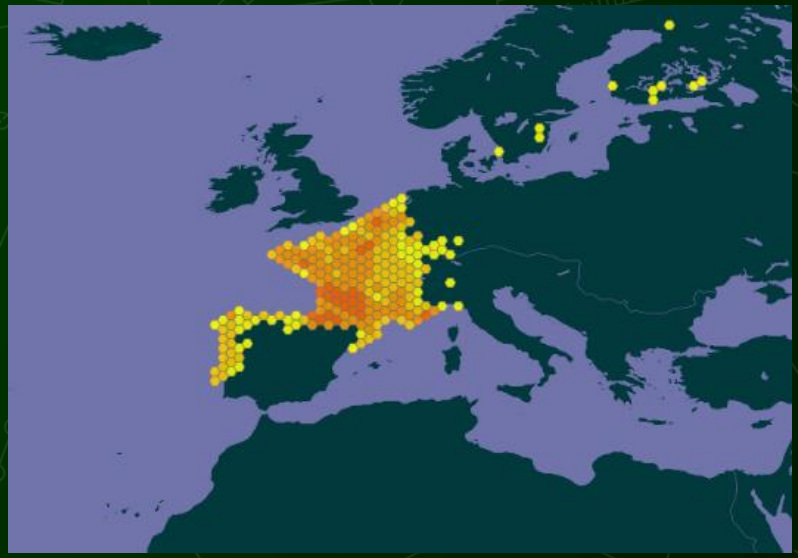


What is GBIF? - example

INVASIVE SPECIES:

Yellow-legged Asian hornet (*Vespa velutina*) expansion in Europe

2020

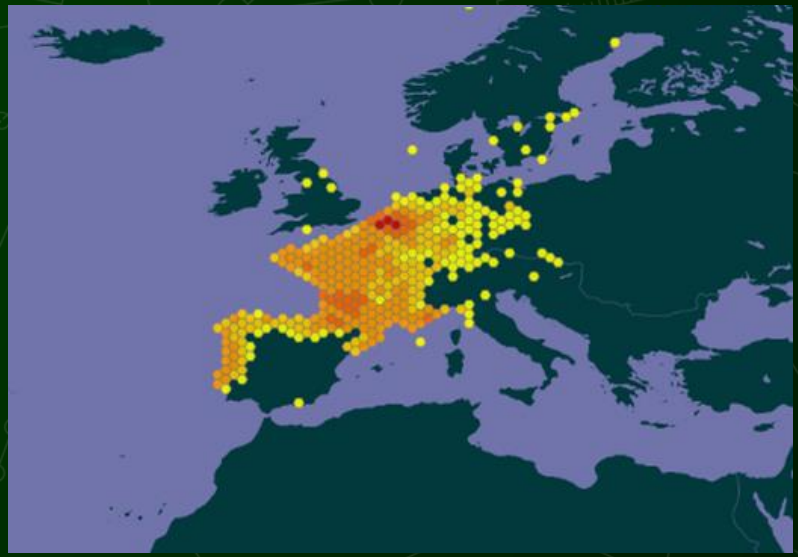


What is GBIF? - example

INVASIVE SPECIES:

Yellow-legged Asian hornet (*Vespa velutina*) expansion in Europe

2024



Datasets

91,826

Hosted portals

19

Country Participants

62

Peer-review papers using data

9,923

Organizational Participants

43

Average records downloaded per month (2023)

173.8 billion

Publishers

2,170

Species occurrence records

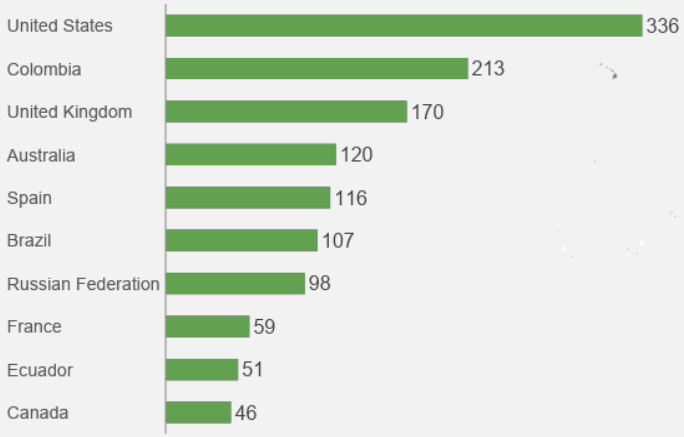
2,627,518,860

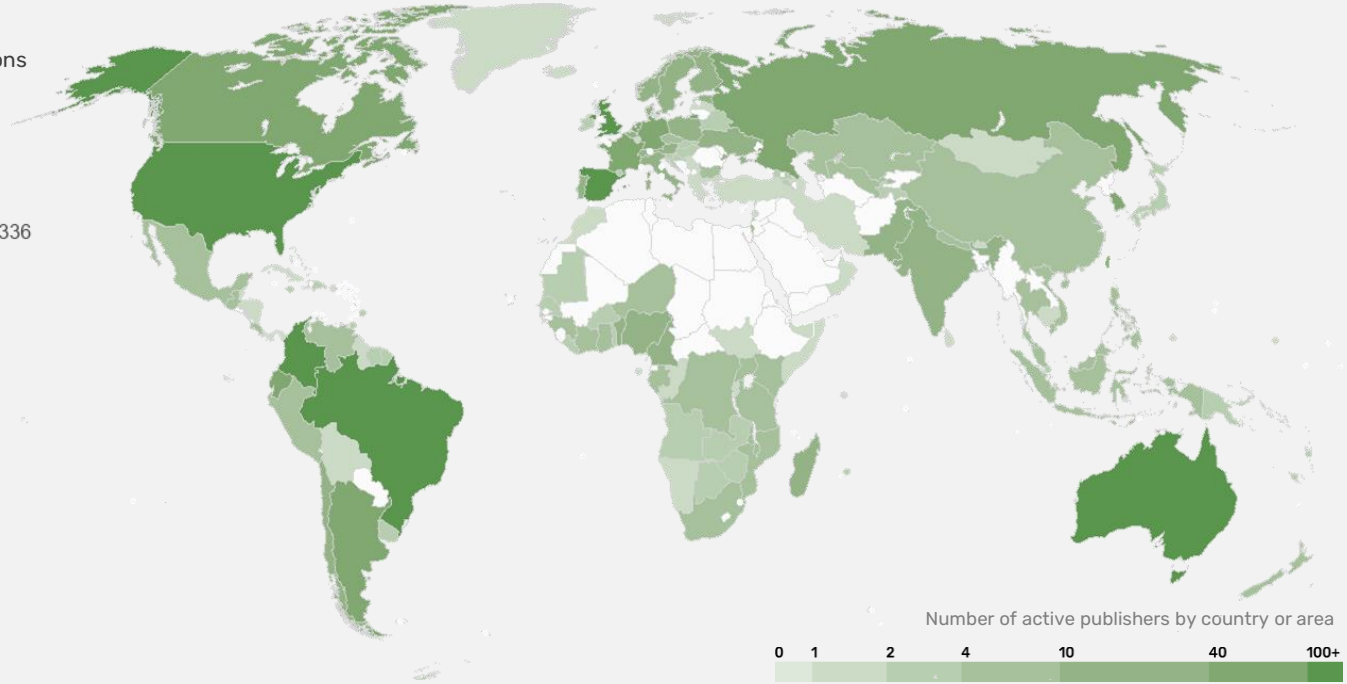
By the numbers | 1 January 2024

www.gbif.org

GBIF network of data publishing institutions

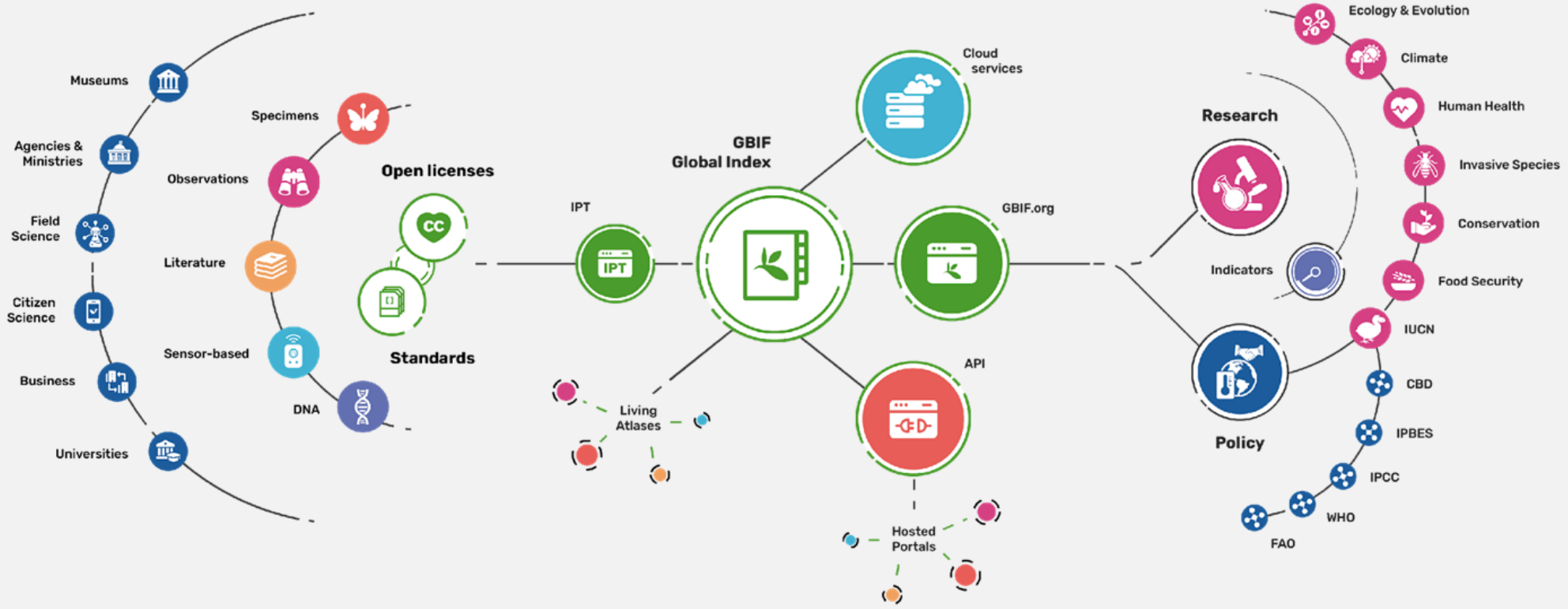
142 countries/areas with institutions sharing data through GBIF





https://www.gbif.org/publisher/search

Providing biodiversity evidence for research and policy



GBIF IN PORTUGAL

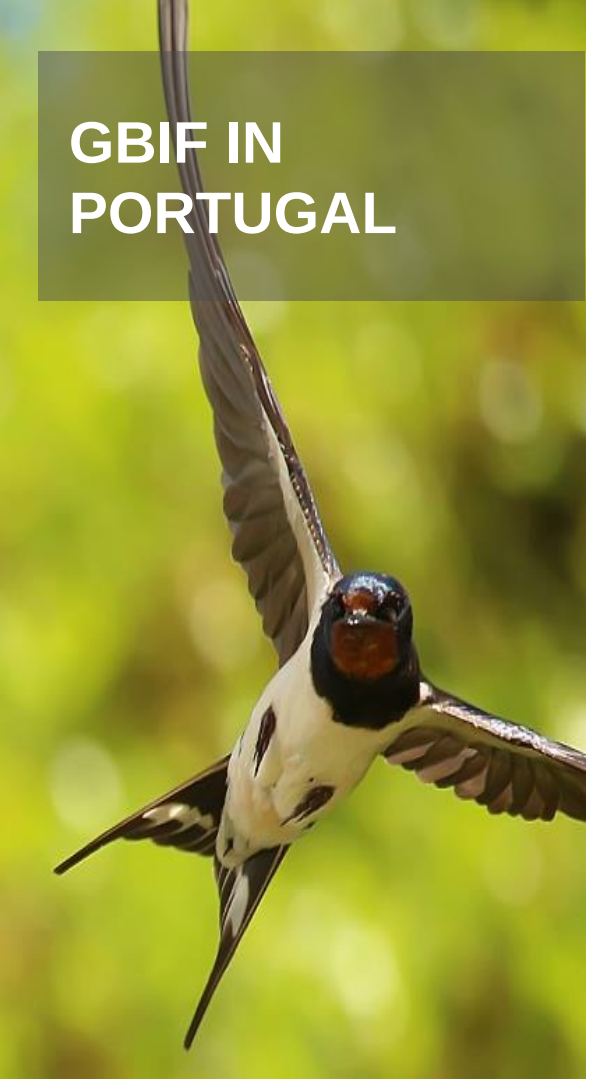
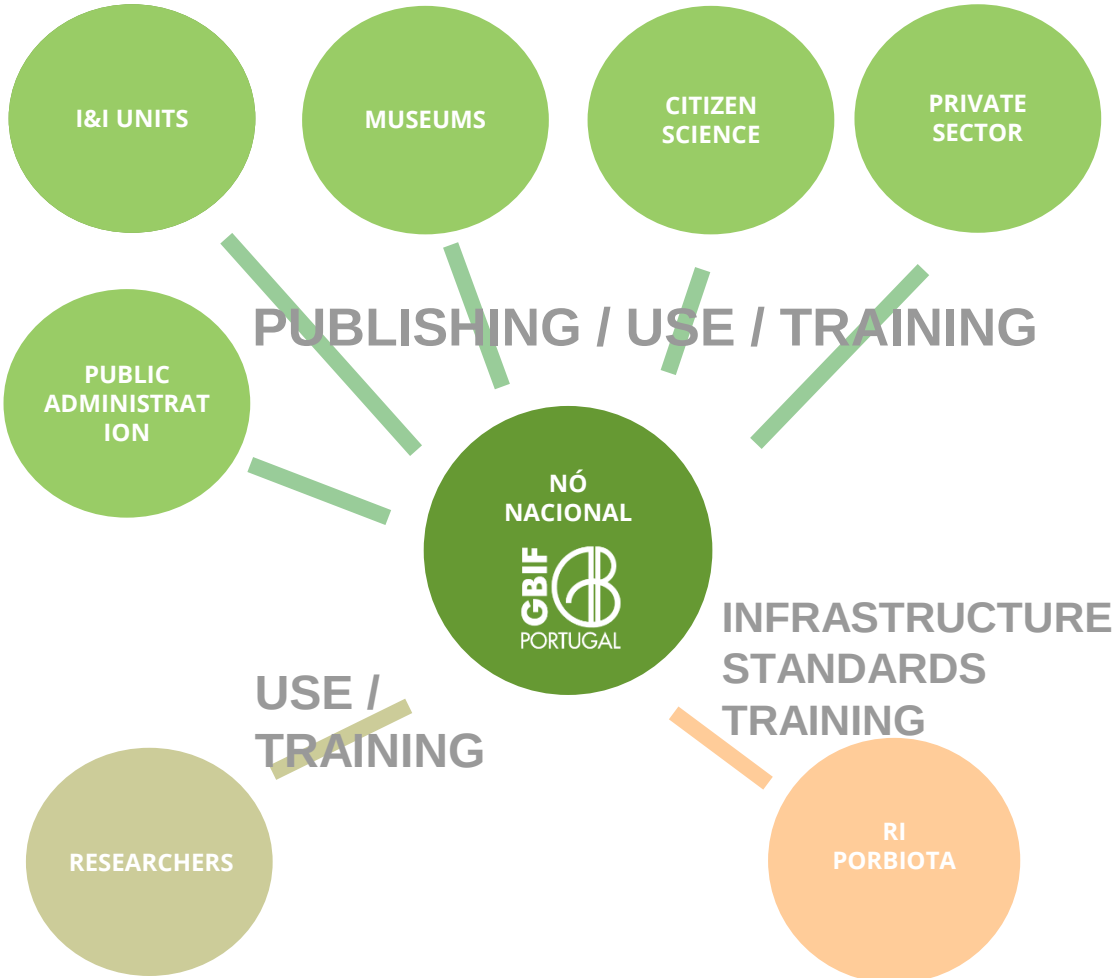


Imagem: César Garcia



Information dimensions

What
Taxonomic/nomenclatural data

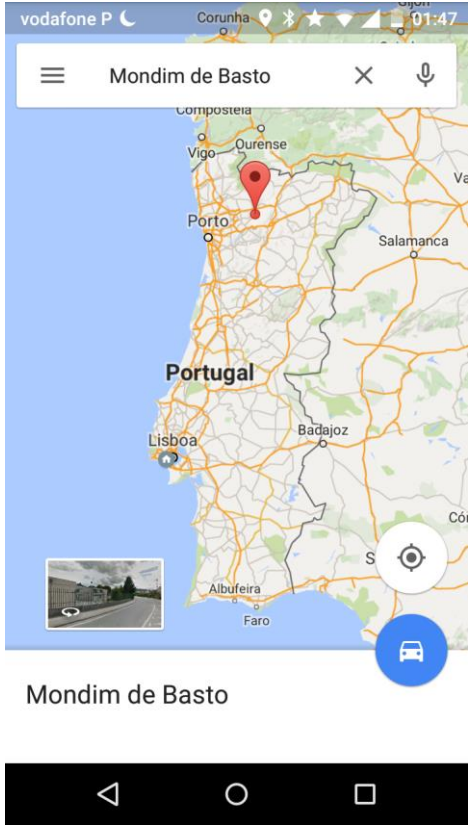
Where
Spatial data

Who
Occurrence data

When
Sampling data

What
Descriptive data

@CésarGarcia, 2010/11/23
Amanita muscaria



Mondim de Basto

Information dimensions

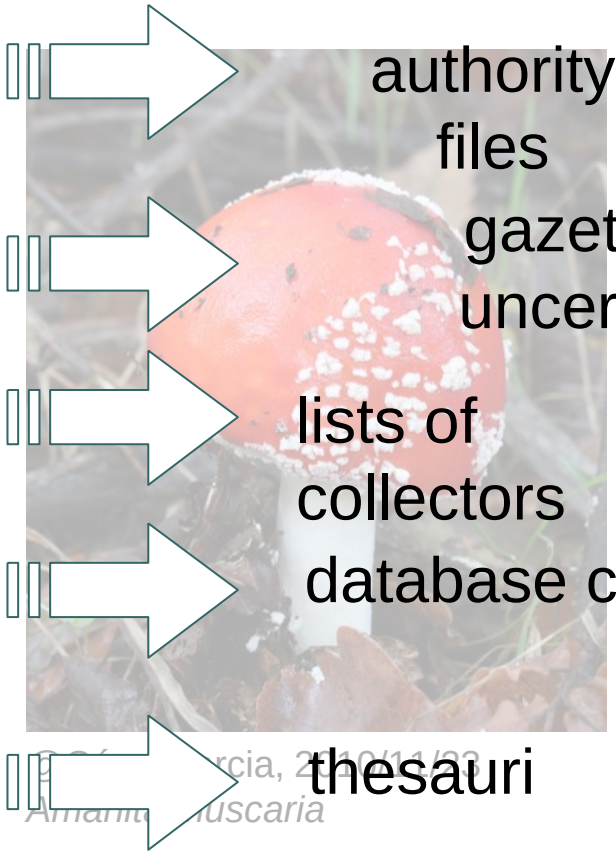
What
Taxonomic/nomenclatural data

Where
Spatial data

Who
Occurrence data

When
Sampling data

What
Descriptive data



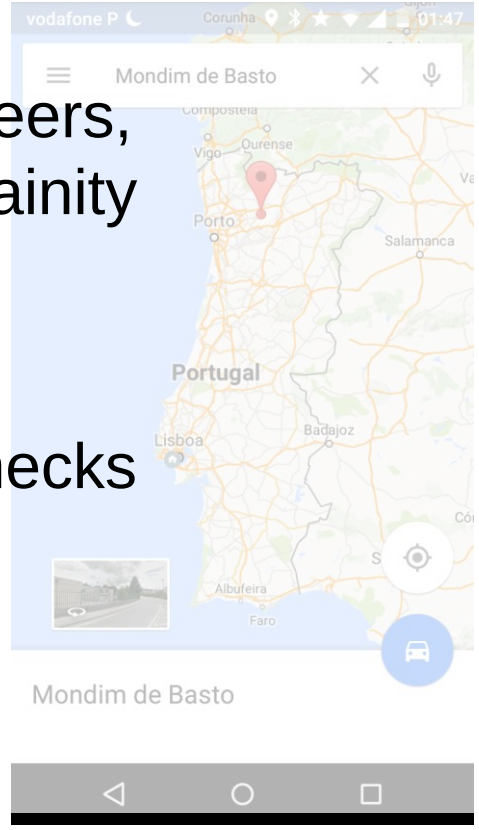
authority files

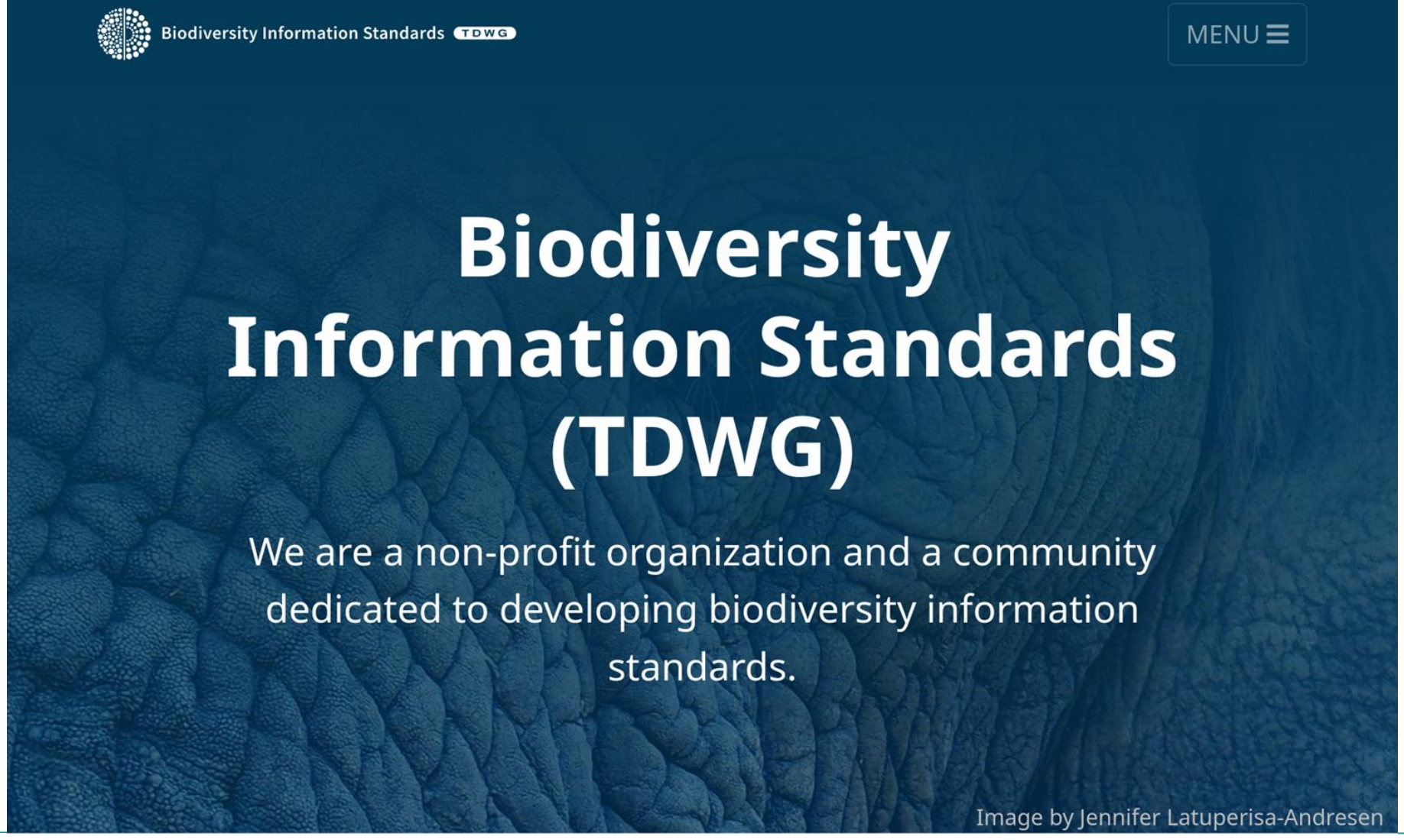
gazetteers, uncertainity

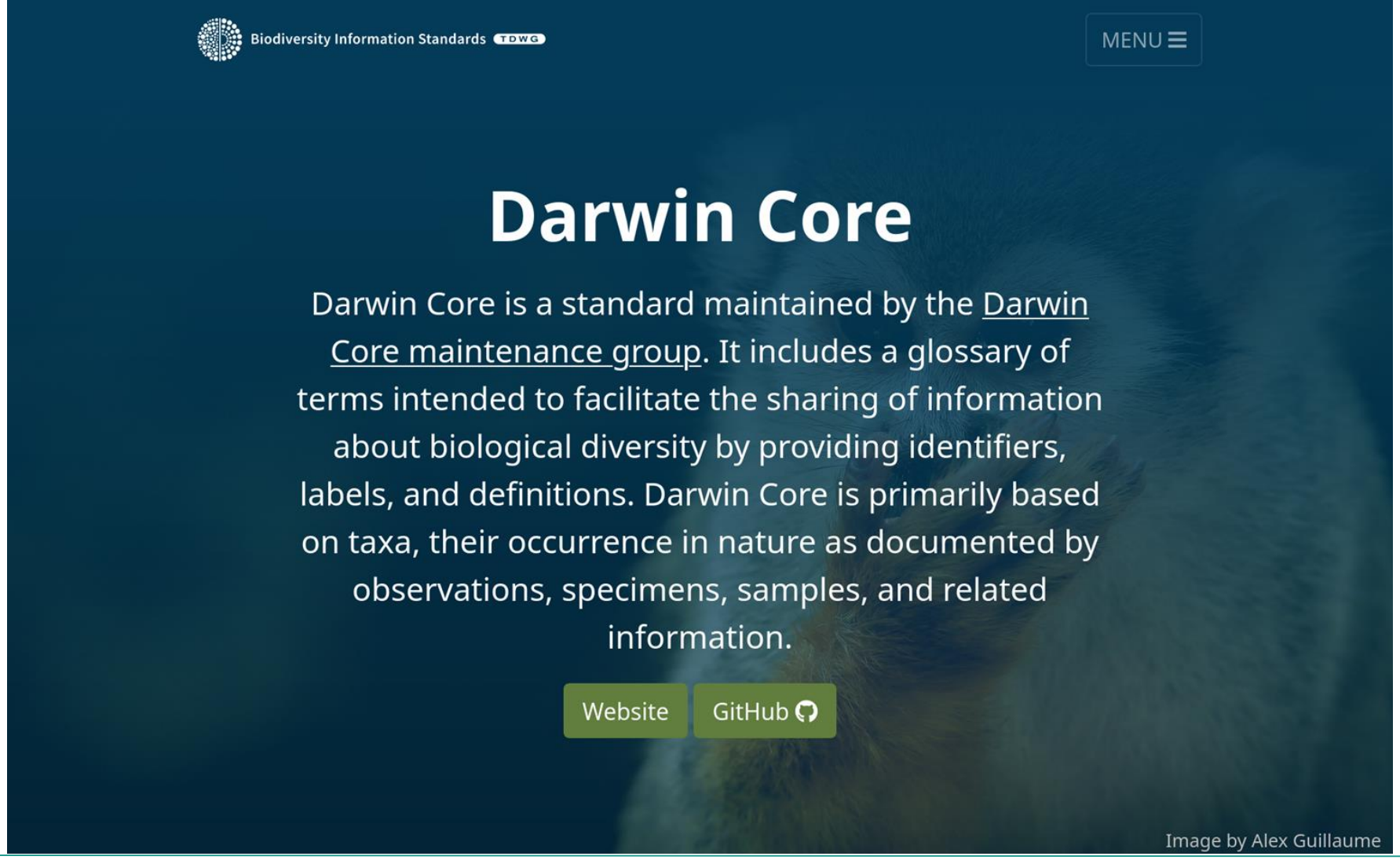
lists of collectors

database checks

thesauri




Biodiversity Information Standards TDWG
MENU
Biodiversity Information Standards (TDWG)
We are a non-profit organization and a community dedicated to developing biodiversity information standards.
Image by Jennifer Latuperisa-Andresen


Biodiversity Information Standards TDWG
MENU
Darwin Core
Darwin Core is a standard maintained by the Darwin Core maintenance group. It includes a glossary of terms intended to facilitate the sharing of information about biological diversity by providing identifiers, labels, and definitions. Darwin Core is primarily based on taxa, their occurrence in nature as documented by observations, specimens, samples, and related information.
Website
GitHub
Image by Alex Guillaume

TDWG Darwin Core

Record-level
Occurrence
Organism
MaterialEntity
MaterialSample
Event
Location
GeologicalContext
Identification
Taxon
MeasurementOrFact
ResourceRelationship
UseWithIRI
LivingSpecimen
PreservedSpecimen
FossilSpecimen
MaterialCitation
HumanObservation
MachineObservation
Cite Darwin Core

Record-level

This category contains terms that are generic in that they might apply to any type of record in a dataset.

type | modified | language | license | rightsHolder | accessRights | bibliographicCitation | references
institutionID | collectionID | datasetID | institutionCode | collectionCode | datasetName | ownerInstitutionCode
basisOfRecord | informationWithheld | dataGeneralizations | dynamicProperties

Occurrence

occurrenceID | catalogNumber | recordNumber | recordedBy | recordedByID | individualCount | organismQuantity
organismQuantityType | sex | lifeStage | reproductiveCondition | caste | behavior | vitality | establishmentMeans
degreeOfEstablishment | pathway | georeferenceVerificationStatus | occurrenceStatus | associatedMedia
associatedOccurrences | associatedReferences | associatedTaxa | otherCatalogNumbers | occurrenceRemarks

Event

eventID | parentEventID | eventType | fieldNumber | eventDate | eventTime | startDayOfYear | endDayOfYear
year | month | day | verbatimEventDate | habitat | samplingProtocol | sampleSizeValue | sampleSizeUnit
samplingEffort | fieldNotes | eventRemarks

TDWG Darwin Core

https://dwc.tdwg.org/terms/

Record-level

type | modified | language | license | rightsHolder | accessRights | bibliographicCitation | references | institutionID
collectionID | datasetID | institutionCode | collectionCode | datasetName | ownerInstitutionCode | basisOfRecord
informationWithheld | dataGeneralizations | dynamicProperties

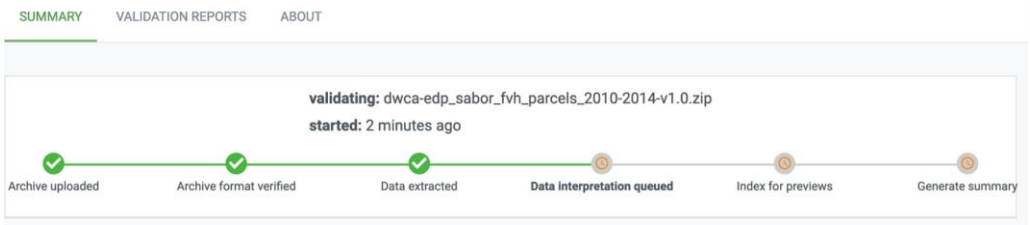
basisOfRecord	Property
Identifier	http://rs.tdwg.org/dwc/terms/basisOfRecord
Definition	The specific nature of the data record.
Comments	Recommended best practice is to use the standard label of one of the Darwin Core classes.
Examples	`PreservedSpecimen`, `FossilSpecimen`, `LivingSpecimen`, `MaterialSample`, `Event`, `HumanObservation`, `MachineObservation`, `Taxon`, `Occurrence`

Darwin Core - training

Yearly training workshops in one of the topics:

- data publication through GBIF
- data quality / data cleaning
- georeferencing
- Specify - management of biological collections

Extensive use of GBIF's and other resources, e.g., GBIF Data Validator



https://www.gbif.org/tools/data-validator/89a13f02-e55e-4632-9b35-7b82f1b3f08e

Darwin Core – core

Occurrences

http://tools.gbif.org/dwca-validator/extension.do?id=dwc:Occurrence

Events

http://tools.gbif.org/dwca-validator/extension.do?id=dwc:Event

Taxa

http://tools.gbif.org/dwca-validator/extension.do?id=dwc:Taxon

Darwin Core – core

Occurrences

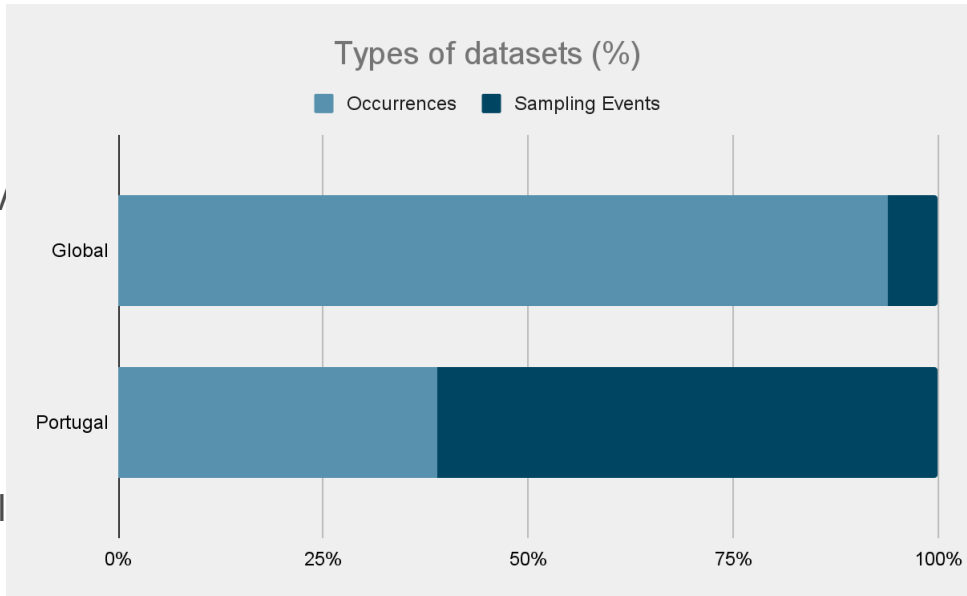
http://tools.gbif.org/dwca-validator/extension.do?id=dwc:Occurrence

Events

http://tools.gbif.org/dwca-v

Taxa

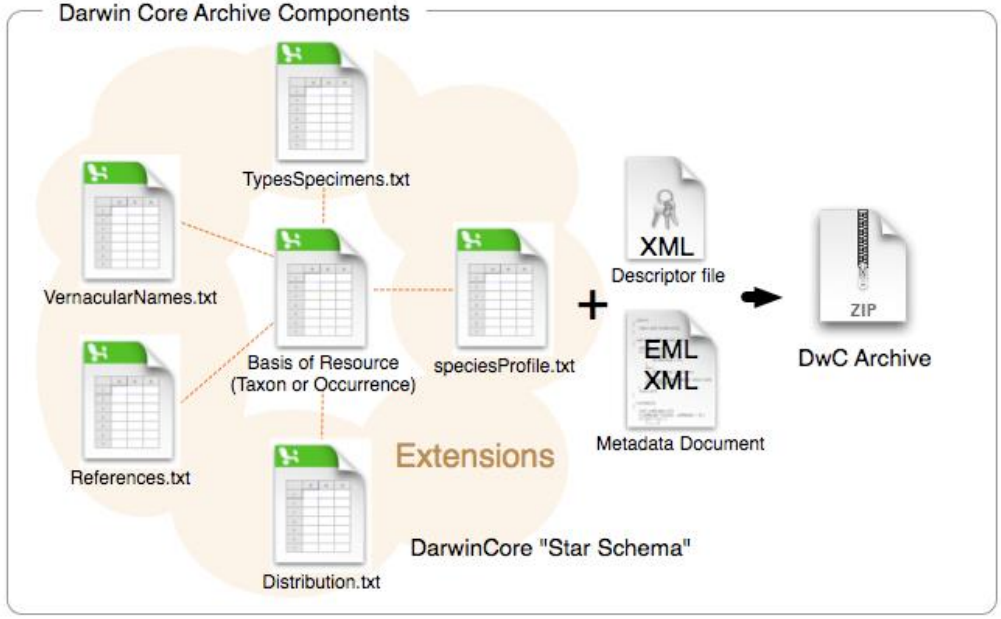
http://tools.gbif.org/dwca-val



https://rs.gbif.org/extensions.html

Darwin Core – Star schema, supporting extensions for several types of data

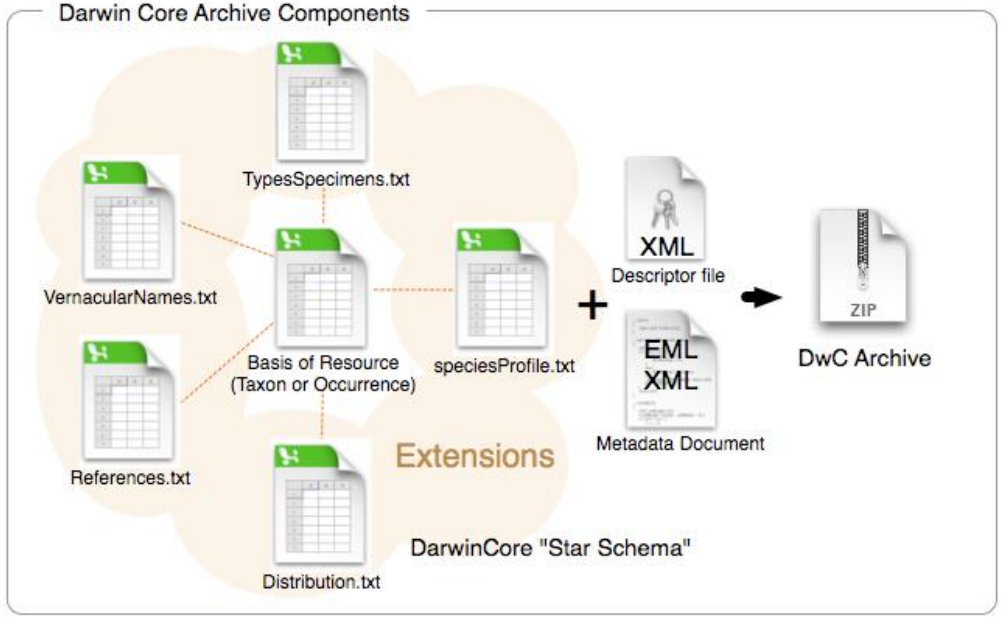
- germplasm
- multiple identifications
- types and specimens
- common names
- alternative identifiers
- species profile
- references in literature
- taxon description
- traits
- images, audio, videos
- others...



https://rs.gbif.org/extensions.html

Darwin Core – Star schema, supporting extensions for several types of data, but a constrain to more richer datasets...

- germplasm
- multiple identifications
- types and specimens
- common names
- alternative identifiers
- species profile
- references in literature
- taxon description
- traits
- images, audio, videos
- others...



TDWG Darwin Core

Darwin Core - MoF

MeasurementOrFact

measurementID | parentMeasurementID | measurementType | measurementValue | measurementAccuracy | measurementUnit | measurementDeterminedBy | measurementDeterminedDate | measurementMethod | measurementRemarks

MeasurementOrFact	Class
Identifier	http://rs.tdwg.org/dwc/terms/MeasurementOrFact
Definition	A measurement of or fact about an rdfs:Resource (http://www.w3.org/2000/01/rdf-schema#Resource).
Comments	Resources can be thought of as identifiable records or instances of classes and may include, but need not be limited to instances of dwc:Occurrence, dwc:Organism, dwc:MaterialEntity, dwc:Event, dcterms:Location, dwc:GeologicalContext, dwc:Identification, or dwc:Taxon.

Examples	`the weight of a dwc:Organism in grams`
	`the number of placental scars`
	`surface water temperature in Celsius`

Darwin Core – dynamicProperties

Record

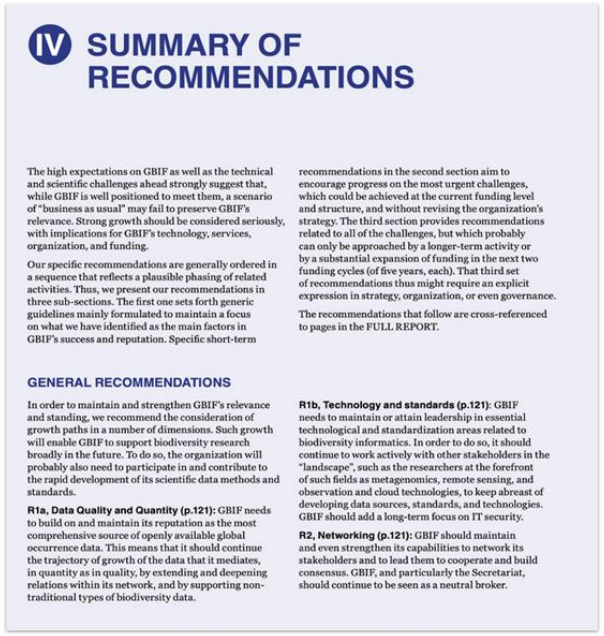
Basis of record	Human observation
Collection code	edp_sabor_fvh_parcels
Dataset name	EDP Baixo Sabor: Construction and Reservoir Filling Phases - Vascular Flora and Habitats: species from pri-oritary habitats/communities - [2010-2014]
Dynamic properties	{“lithology”:"schists", “% rock”:35 , “exposureInCardinalDirections”:"W" , “slopeInDegrees”:30}
Institution code	EDP
Institution ID	e5150835-f502-424c-b470-24dd496b1b18
Owner institution code	EDP

https://globalnodes.gbif.org/en/gnm/data-model#meeting

GBIF 20-year review

"...biodiversity data is **more complicated than 'just' the occurrence of species** in time and space; organisms interact, co-occur, move and evolve.

This implies a need for **richer and more varied types of data** than GBIF has thus far supported"



IV SUMMARY OF RECOMMENDATIONS

The high expectations on GBIF as well as the technical and scientific challenges ahead strongly suggest that, while GBIF is well positioned to meet them, a scenario of "business as usual" may fail to preserve GBIF's relevance. Strong growth should be considered seriously, with implications for GBIF's technology, services, organization, and funding.

Our specific recommendations are generally ordered in a sequence that reflects a plausible phasing of related activities. Thus, we present our recommendations in three sub-sections. The first one sets forth generic guidelines mainly formulated to maintain a focus on what we have identified as the main factors in GBIF's success and reputation. Specific short-term recommendations in the second section aim to encourage progress on the most urgent challenges, which could be achieved at the current funding level and structure, and without revising the organization's strategy. The third section provides recommendations related to all of the challenges, but which probably can only be approached by a longer-term activity or by a substantial expansion of funding in the next two funding cycles (of five years, each). That third set of recommendations thus might require an explicit expression in strategy, organization, or even governance.

The recommendations that follow are cross-referenced to pages in the FULL REPORT.

GENERAL RECOMMENDATIONS

In order to maintain and strengthen GBIF's relevance and standing, we recommend the consideration of growth paths in a number of dimensions. Such growth will enable GBIF to support biodiversity research broadly in the future. To do so, the organization will probably also need to participate in and contribute to the rapid development of its scientific data methods and standards.

R1a, Data Quality and Quantity (p.121): GBIF needs to build on and maintain its reputation as the most comprehensive source of openly available global occurrence data. This means that it should continue the trajectory of growth of the data that it mediates, in quantity as in quality, by extending and deepening relations within its network, and by supporting non-traditional types of biodiversity data.

R1b, Technology and standards (p.121): GBIF needs to maintain or attain leadership in essential technological and standardization areas related to biodiversity informatics. In order to do so, it should continue to work actively with other stakeholders in the "landscape", such as the researchers at the forefront of such fields as metagenomics, remote sensing, and observation and cloud technologies, to keep abreast of developing data sources, standards, and technologies. GBIF should add a long-term focus on IT security.

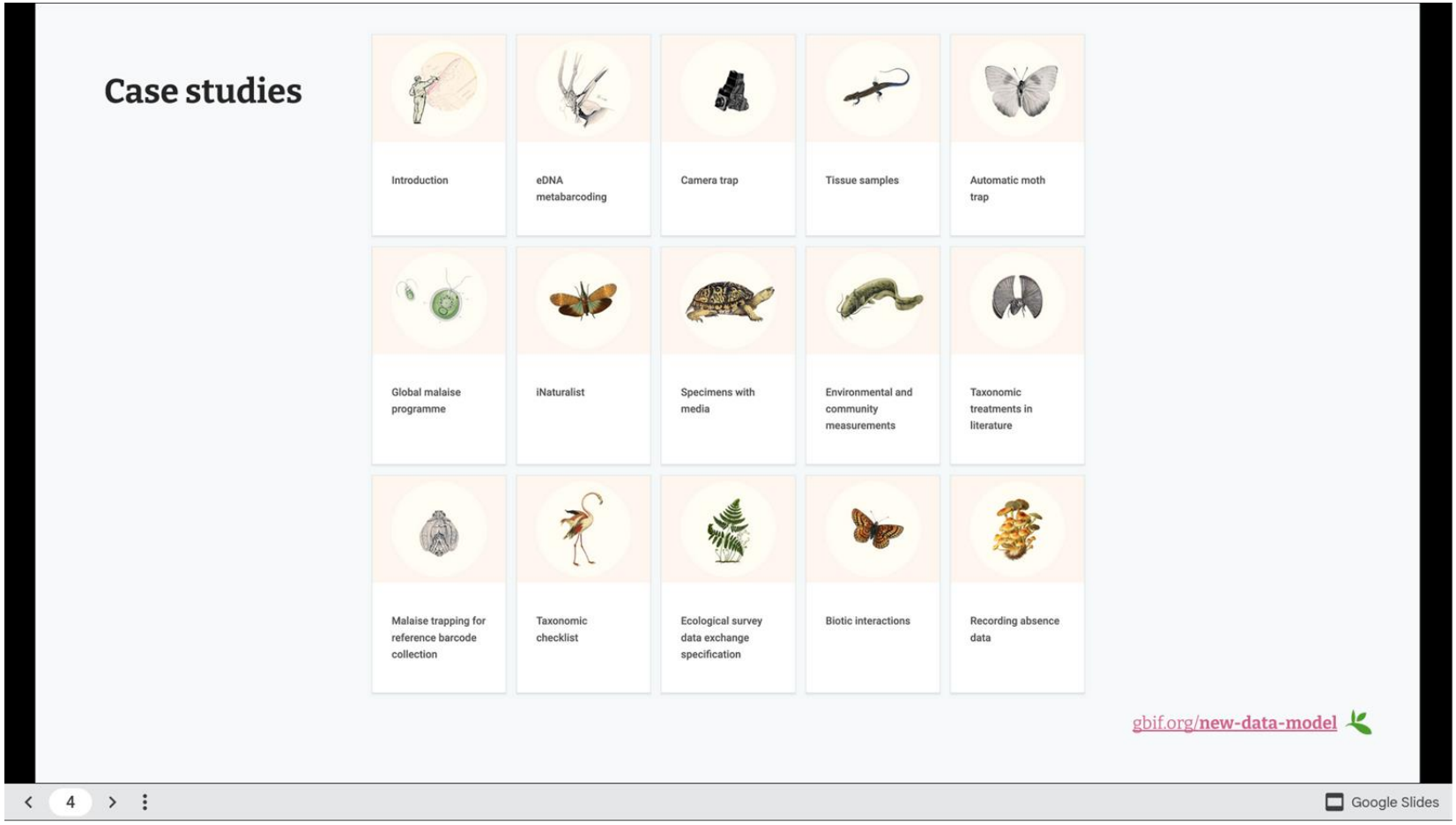
R2, Networking (p.121): GBIF should maintain and even strengthen its capabilities to network its stakeholders and to lead them to cooperate and build consensus. GBIF, and particularly the Secretariat, should continue to be seen as a neutral broker.

CODATA, the Committee on Data of the International Science Council, Pfeiffenberger, Hans, Uhlir, Paul, & Hodson, Simon. (2020). Twenty-Year Review of GBIF. https://doi.org/10.35035/ctzm-hz97

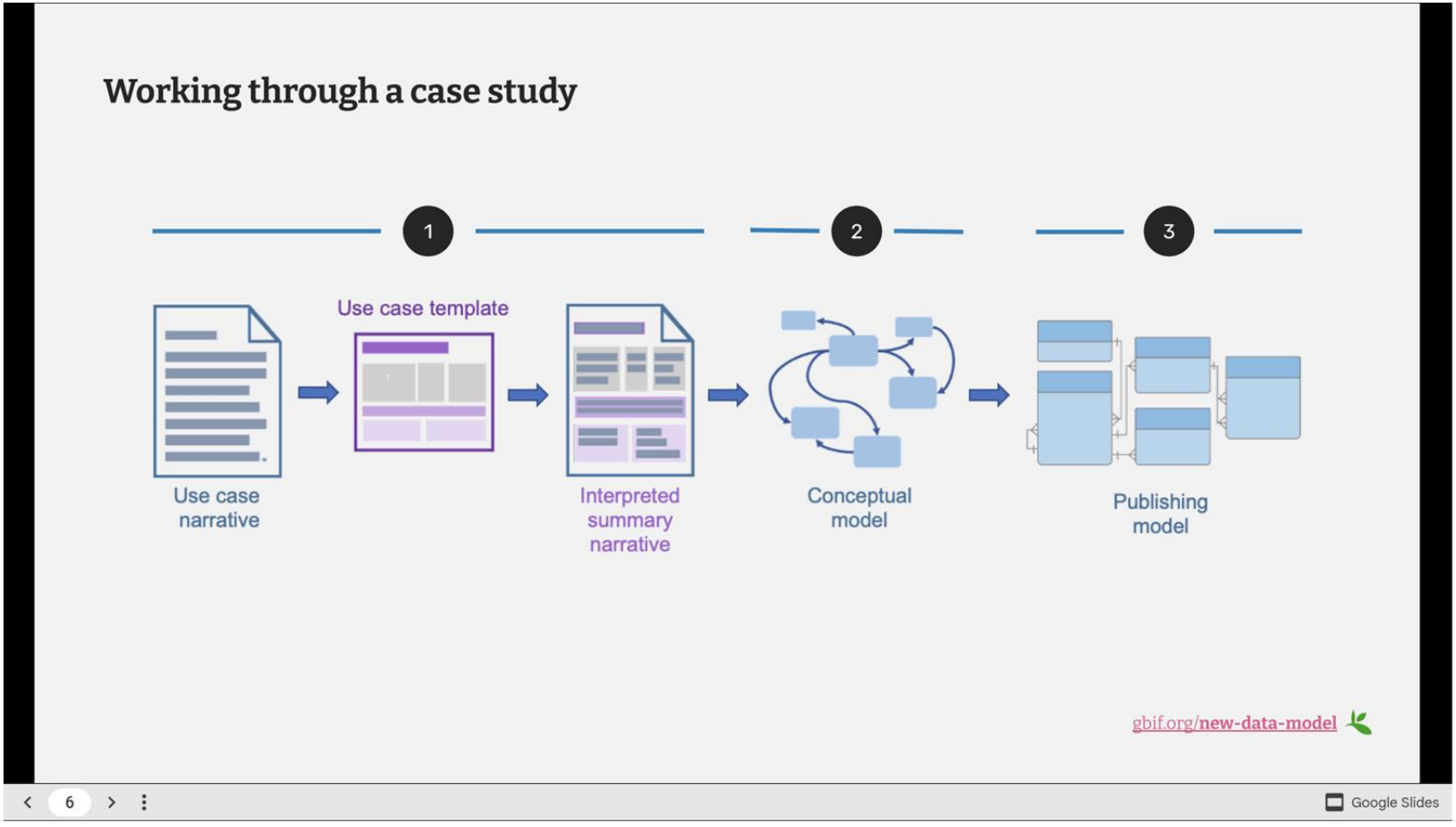
3

Google Slides

https://globalnodes.gbif.org/en/gnm/data-model#meeting

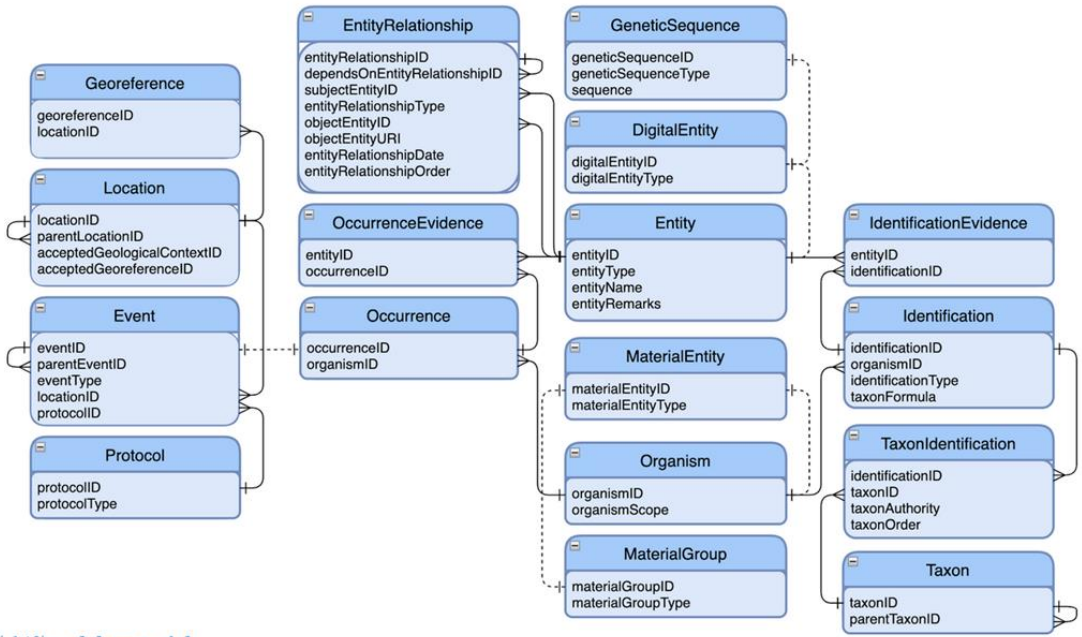


https://globalnodes.gbif.org/en/gnm/data-model#meeting



https://globalnodes.gbif.org/en/gnm/data-model#meeting

Instruction: “Map your data to this, please”



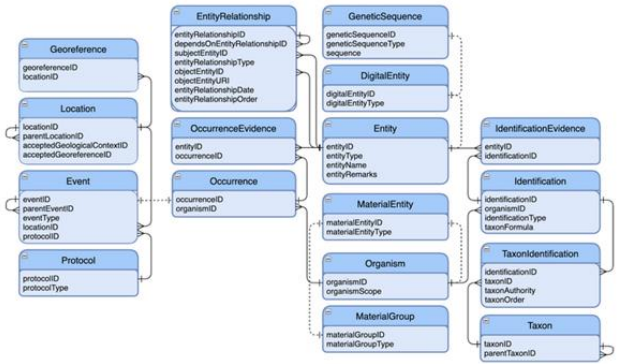
http://github.com/gbif/**model-material**

16

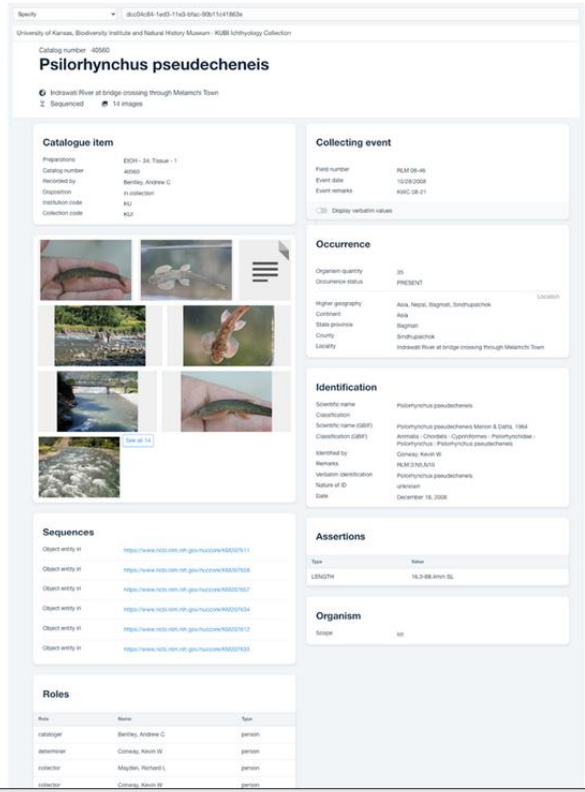
Google Slides

https://globalnodes.gbif.org/en/gnm/data-model#meeting

Instruction: “Map your data to this, please”



http://github.com/gbif/**model-material**

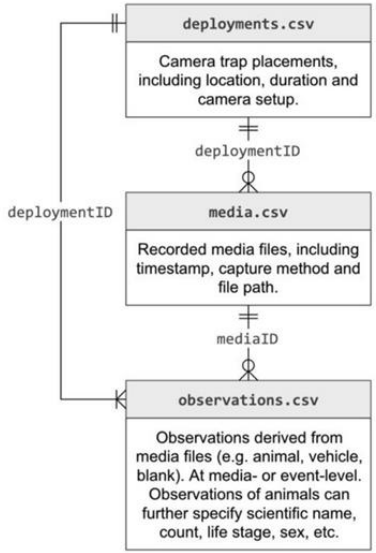


17

Google Slides

https://globalnodes.gbif.org/en/gnm/data-model#meeting

Publishing model: Camtrap DP



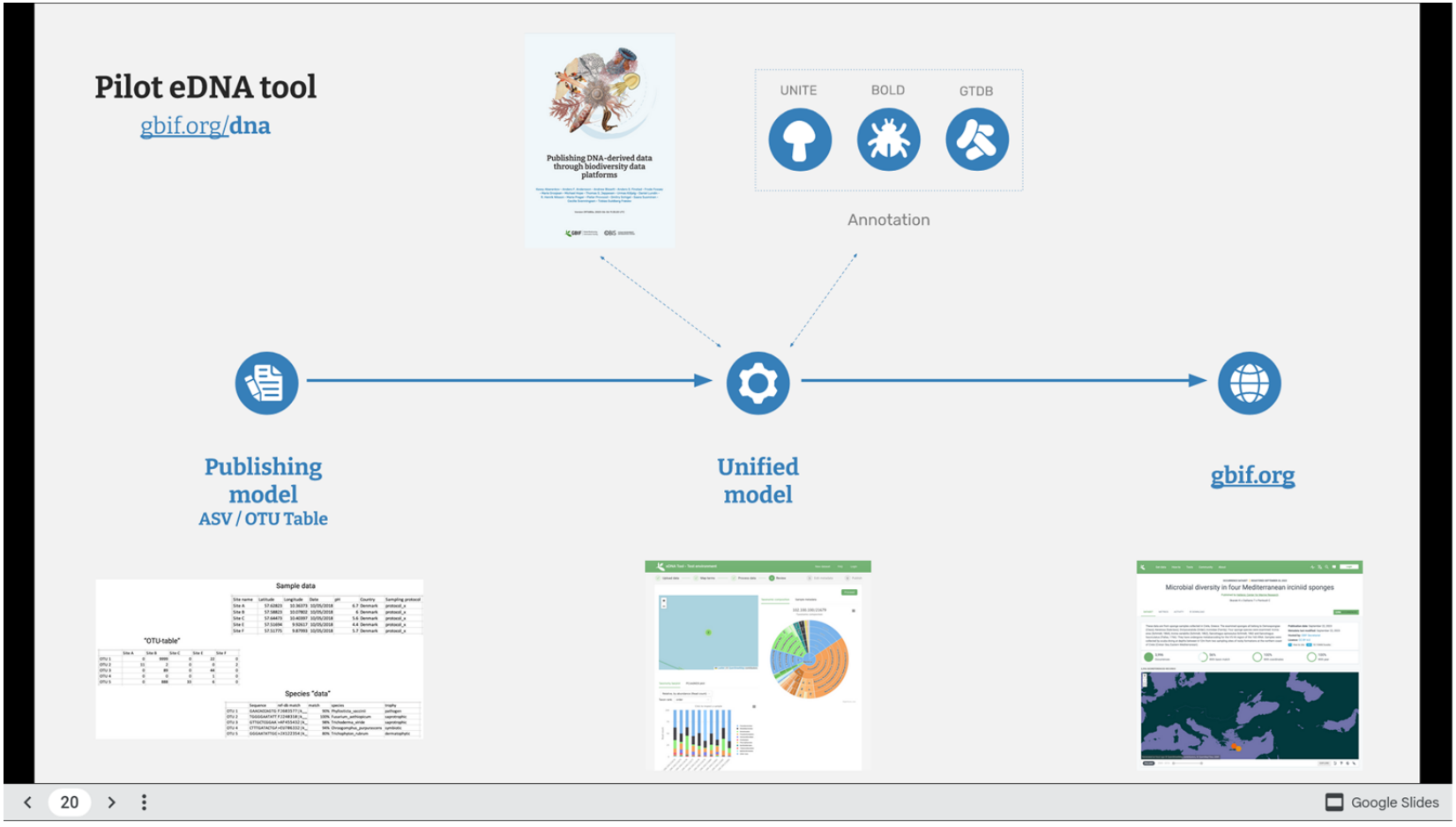
Credit: Tom Rickman, CC BY-NC
https://www.inaturalist.org/observations/148735438

https://bit.ly/camtrap-dp-tdwg-2023

19

Google Slides

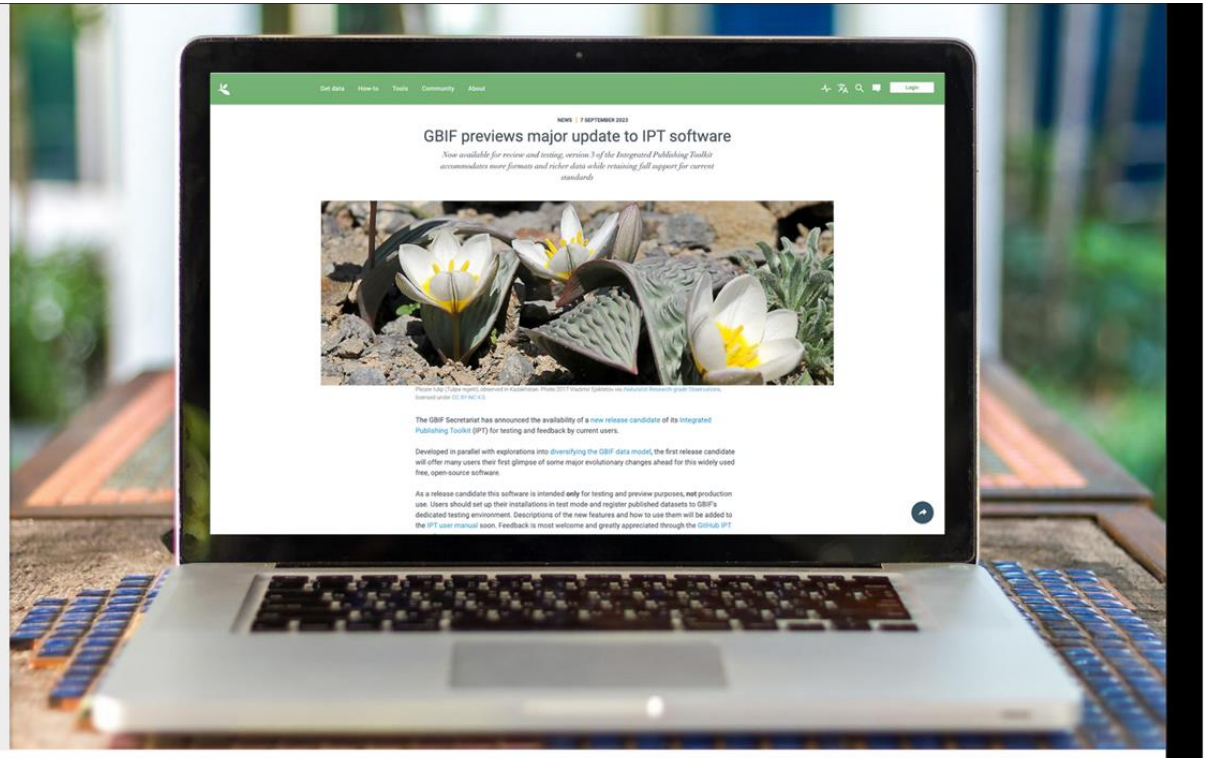
https://globalnodes.gbif.org/en/gnm/data-model#meeting



https://globalnodes.gbif.org/en/gnm/data-model#meeting

Integrated Publishing Toolkit

- Version 3.0 nearing release
- Supports mapping to Frictionless Data standards
- Critical to the support of new publishing models

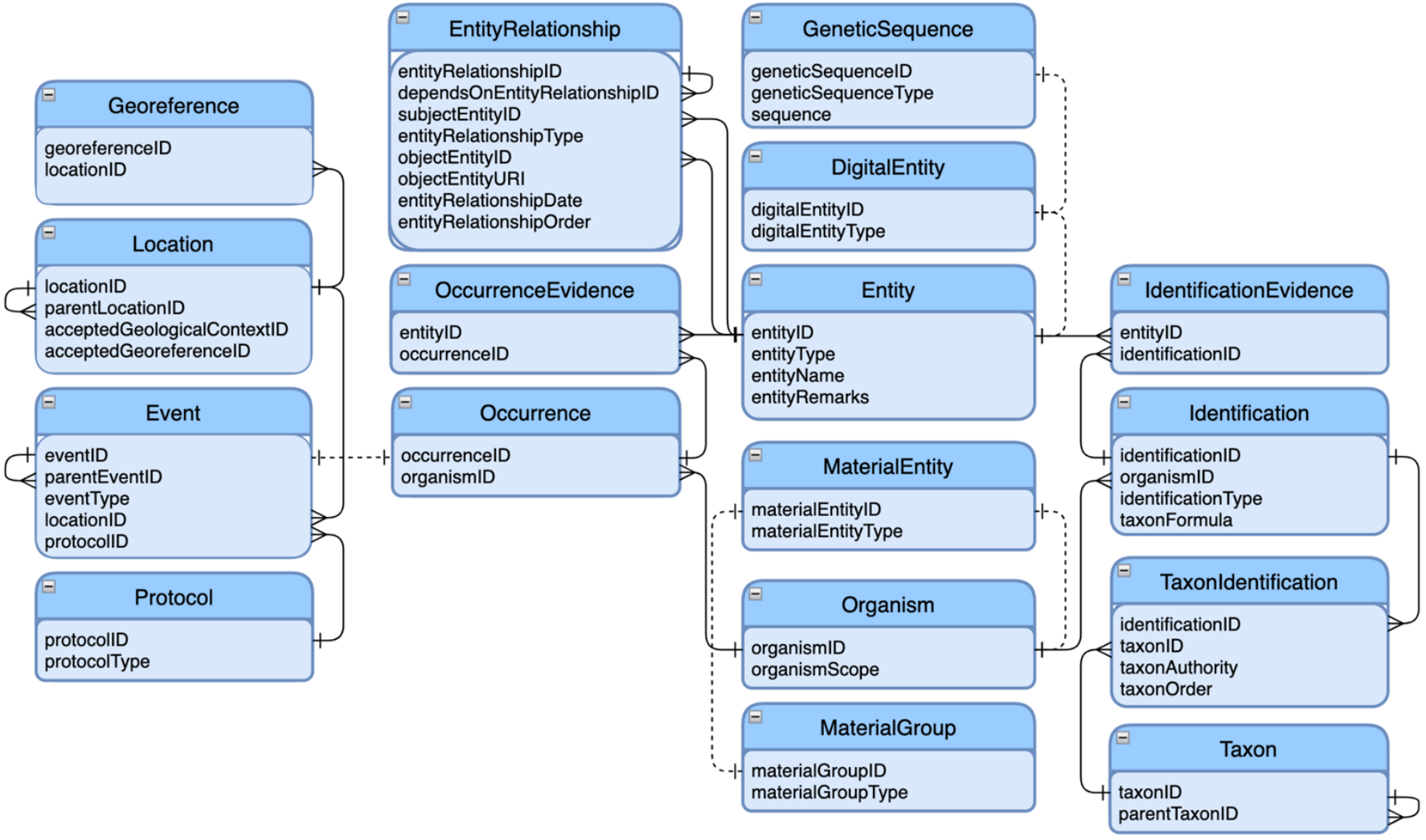


https://gbif.org/ipt

22

Google Slides

https://globalnodes.gbif.org/en/gnm/data-model#meeting

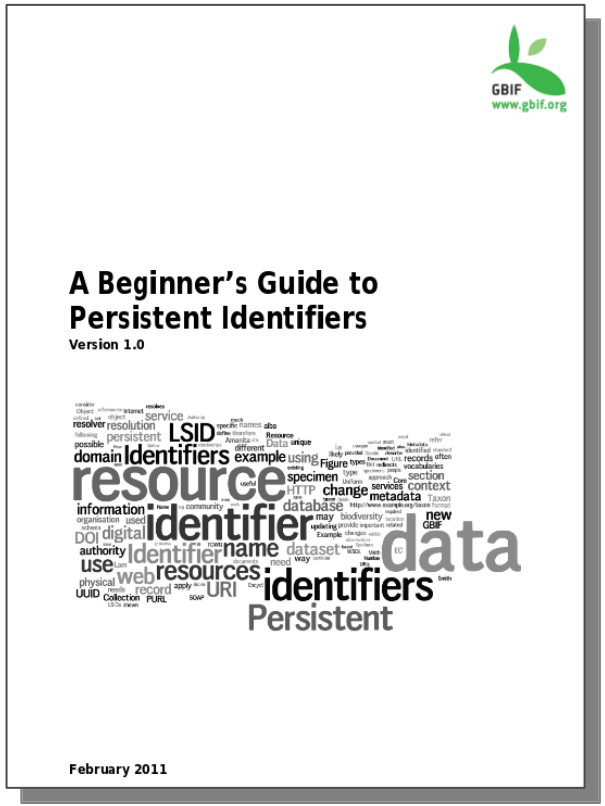

Georeference
georeferenceID
locationID
Location
locationID
parentLocationID
acceptedGeologicalContextID
acceptedGeoreferenceID
Event
eventID
parentEventID
eventType
locationID
protocolID
Protocol
protocolID
protocolType
EntityRelationship
entityRelationshipID
dependsOnEntityRelationshipID
subjectEntityID
entityRelationshipType
objectEntityID
objectEntityURI
entityRelationshipDate
entityRelationshipOrder
OccurrenceEvidence
entityID
occurrenceID
Occurrence
occurrenceID
organismID
GeneticSequence
geneticSequenceID
geneticSequenceType
sequence
DigitalEntity
digitalEntityID
digitalEntityType
Entity
entityID
entityType
entityName
entityRemarks
MaterialEntity
materialEntityID
materialEntityType
Organism
organismID
organismScope
MaterialGroup
materialGroupID
materialGroupType
IdentificationEvidence
entityID
identificationID
Identification
identificationID
organismID
identificationType
taxonFormula
TaxonIdentification
identificationID
taxonID
taxonAuthority
taxonOrder
Taxon
taxonID
parentTaxonID

Persistent identifiers

Why do we find so many xxxxIDs under the DwC standard?
What do they mean and what are they for?

A Beginner's Guide to Persistent Identifiers

https://www.gbif.org/document/80575



Persistent identifiers - relation to FAIR

Why do we find so many xxxxIDs under the DwC standard?
What do they mean and what are they for?

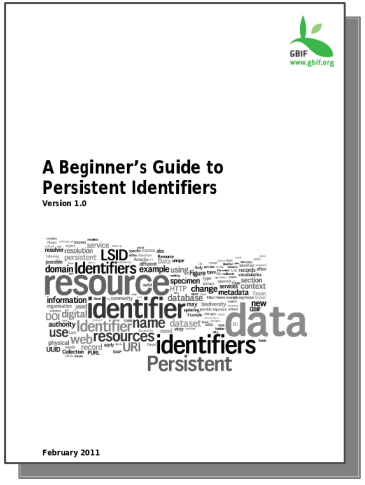


Table 9: Overall summary of the FAIRness assessment of the EGI-ACE Data Spaces.

Data Space	Findability (F)	Accessible (A)	Interoperable (I)	Reusability (R)
GBIF	**100%** Compliant	**100%** Complaint	**86%** Compliant	**100%** Compliant

https://documents.egi.eu/document/3815

Thank you!

RUI FIGUEIRA
NÓ PORTUGUÊS DO GBIF
INSTITUTO SUPERIOR DE AGRONOMIA
ruifigueira@isa.ulisboa.pt

The Portuguese Node is hosted by ISA-ULisboa with the support of

Fundação para a Ciência e a Tecnologia

biodiversa+
European Biodiversity Partnership

Biodiversa+ Capacity building workshop on Darwin Core standard

7th of June from 9:00 to 12:40 CEST

INSTITUTO SUPERIOR DE AGRONOMIA
Universidade de Lisboa

GBIF PORTUGAL

PORBIOTA
E-INFRAESTRUTURA PORTUGUESA DE INFORMAÇÃO E INVESTIGAÇÃO EM BIODIVERSIDADE

This content is available under an open access license
https://creativecommons.org/licenses/by-sa/4.0/

The use of Darwin Core as standard terminology to harmonise biodiversity data in LifeWatch Italy

Andrea Tarallo

Institute of Research on Terrestrial Ecosystems (IRET)
National Research Council (CNR), Lecce, Italy

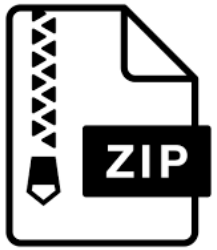
Biodiversa+ Data management and capacity building workshop on Darwin Core standard – 7 June 2024

LifeWatch Italy Joint Research Unit

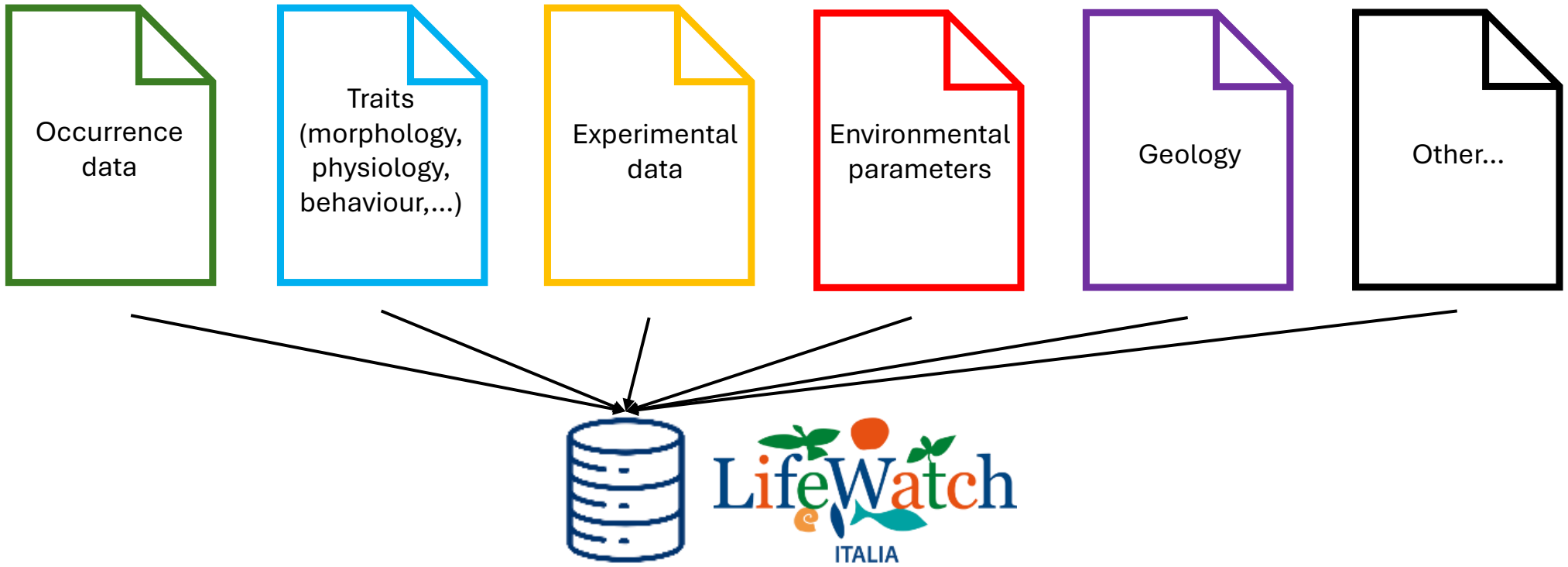
LifeWatch ITA has been established as a **Joint Research Unit (JRU)** in 2010 with 9 founding members **leaded by CNR**. It has now **35 members** including research institutes, universities, associations and private companies distributed, seeking to reinforce integrated scientific research on biodiversity and ecosystems.

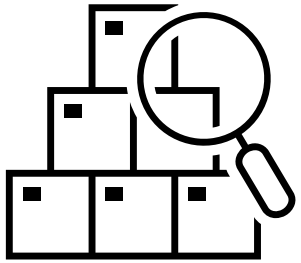
A required core data file:
Occurrence, checklist or sampling-event
+
Optional Extensions
+
Metafile (annotation)
+
Resource Metadata

1. Data file in tabular format
2. Annotation
3. Metadata



The Darwin Core Archive is an effective way to mobilise and exchange specimen occurrence and observational records



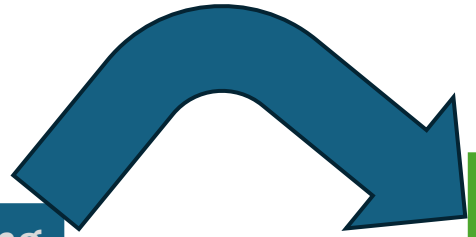


Flexible solution but that allow complex queries

Map variables of CSV files with standard vocabularies and thesauri

XLS files

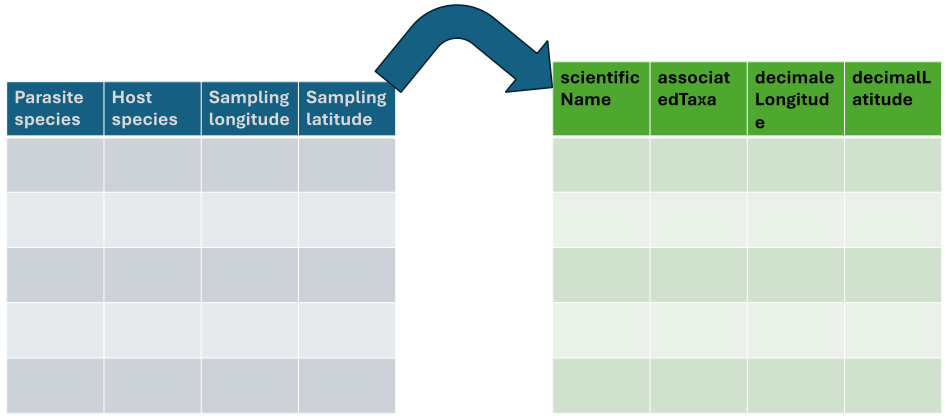
Parasite species	Host species	Sampling longitude	Sampling latitude



CSV UTF-8 files

scientificName	associatedTaxa	decimaleLongitude	decimalLatitude

Map variables of CSV files with standard vocabularies and thesauri



- Darwin Core
- EcoPortal (LifeWatch Italy thesauri and others)
- NERC Vocabulary Server
- BioPortal and other OntoPortal Alliance repositories
- Others

NERC Environmental Data Service

Biodiversity Information Standards TDWG

LifeWatch ITALIA

eLTER

EcoPortal

DATA PROVIDER FILE

Organism				
Id	**Species**	**Taxonomy**	**Life.stage**	**SEX**
1021315821	Carabus auratus	Coleoptera	adult	female
1134943631	Carabus auratus	Coleoptera	adult	male
1591440928	Oniscus asellus	Isopoda; Oniscidae; Oniscus	NA	female
633288236	Oniscus asellus	Isopoda; Oniscidae; Oniscus	NA	male
409193800	Oniscus asellus	Isopoda; Oniscidae; Oniscus	NA	unsexed
1050793085	Oreochromis mossambicus	fish	Fry	NA
1050793084	Oreochromis mossambicus	fish	Fry	NA
1050793083	Oniscus asellus	Oniscus asellus	Juvenile	unsexed

Metabolism						
Harmonized_MR_watts	**Originalg_unit_MR**	**Harmonized_Body_mass_grams**	**Dry.weight**	**Measurment_Temperature_C**	**Coordinates**	**Ecosystem**
3,57E+09	Watts	11.9	4.95	5	NA	Terrestrial
3,25E+09	ml/h	13.4	0.00191	10	NA	Terrestrial
2,19E+09	ml/h	0.512	0.1617	15	67 30' N, 26 40' E	Terrestrial
2,03E-02	µLO2 h-1	12.9	0.00253	20	67 30' N, 26 40' E	Terrestrial
1,05E+09	µLO2 h-1	12.9	0.00672	25	NA	Terrestrial
0.000288972	µLO2 h-1	43.7	0.0078	30	41°30N; 145°50E-44°00N; 155°00E	Aquatic
0.009355331	µLO2 h-1	21.3	0.0309	20	41 46 N 65 28' W	Aquatic
0.009585727	J h-1	4.65	0.0138	22	NA	Terrestrial

REF			EXP		
Openess.of.data..reusable.policy	**Review.Doi**	**Original.Doi**	**Thermoregulatory**	**Salinity**	**Laboratory.Condition**
https://www.journals.uchicago.edu/t-and-c paragraph 3.1.7	https://doi.org/10.1086/505997•	10.1016/0198-0149(90)90104-4	ectotherm	NA	NA
CC-BY	https://doi.org/10.1086/505997•	10.1016/0198-0149(90)90104-4	ectotherm	NA	No
https://onlinelibrary.wiley.com/library-info/resources/text-and-datamining	https://doi.org/10.1086/505997•	10.1016/0300-9629(73)90241-7	ectotherm	NA	NA
https://onlinelibrary.wiley.com/library-info/resources/text-and-datamining	https://doi.org/10.1086/505997•	10.1016/0300-9629(73)90241-7	ectotherm	NA	NA
https://onlinelibrary.wiley.com/library-info/resources/text-and-datamining	https://doi.org/10.21203/rs.3.rs-1092818/v1	10.1016/0198-0149(90)90104-4	ectotherm	10	none specified
Creative Commons Attribution 4	https://doi.org/10.21203/rs.3.rs-1092818/v1	https://doi.org/10.1016/0300-9629(75)90146-2	ectotherm	30	Yes
Creative Commons Attribution 4	https://doi.org/10.21203/rs.3.rs-1092818/v1	https://doi.org/10.1016/0300-9629(75)90146-2	ectotherm	20	Yes
Creative Commons Attribution 4.0	https://doi.org/10.21203/rs.3.rs-1092818/v1	https://doi.org/10.1016/0300-9629(75)90146-2	ectotherm	30	Yes

MAPPING OF COLUMNS HEADER

Original label	Label	Terminology used	ID
ID	catalogNumber	DwC	http://rs.tdwg.org/dwc/terms/catalogNumber
Kingdom	kingdom	DwC	http://rs.tdwg.org/dwc/terms/kingdom
Phylum	phylum	DwC	http://rs.tdwg.org/dwc/terms/phylum
Class	class	DwC	http://rs.tdwg.org/dwc/terms/class
Order	order	DwC	http://rs.tdwg.org/dwc/terms/order
Family	family	DwC	http://rs.tdwg.org/dwc/terms/family
Genus	genus	DwC	http://rs.tdwg.org/dwc/terms/genus
Species	scientificName	DwC	http://rs.tdwg.org/dwc/terms/scientificName
Life.stage	lifeStage	DwC	http://rs.tdwg.org/dwc/terms/lifeStage
SEX	sex	DwC	http://rs.tdwg.org/dwc/terms/sex
Harmonized_MR_watts	metabolic rate	Environmental Thesaurus	http://vocabs.lter-europe.net/EnvThes/21316
Originalg_unit_MR	Original Result Unit	NCI Thesaurus OBO Edition	http://purl.obolibrary.org/obo/NCIT_C82586
Harmonized_Body_mass_grams	body mass	Environmental Thesaurus	http://vocabs.lter-europe.net/EnvThes/21364
Dry.weight	Dry Mass	zooplanktontraits	https://kos.lifewatch.eu/thesauri/zooplanktontraits/c_19
Measurment_Temperature_C	temperatureValue	ABCD	http://rs.tdwg.org/abcd/terms/temperatureValue
Coordinates	verbatimCoordinates	DwC	http://rs.tdwg.org/dwc/terms/verbatimCoordinates
N/A	decimalLatitude	DwC	http://rs.tdwg.org/dwc/terms/decimalLatitude
N/A	decimalLongitude	DwC	http://rs.tdwg.org/dwc/terms/decimalLongitude
Ecosystem	habitat	Environmental Thesaurus	http://vocabs.lter-europe.net/EnvThes/21894
Openess.of.data..reusable.policy	license	DwC	http://purl.org/dc/terms/license
Review.Doi	Is Referenced By	DC	http://purl.org/dc/terms/isReferencedBy
Original.Doi	references	DC	http://purl.org/dc/terms/references
Thermoregulatory	body temperature regulation trait	Vertebrate trait ontology	http://purl.obolibrary.org/obo/VT_0001777
Salinity	Water body salinity	EMODnet Chemistry aggregated parameter names	http://vocab.nerc.ac.uk/collection/P35/current/EPC00001/
Laboratory.Condition	laboratory study	Radiation Biology Ontology	http://purl.obolibrary.org/obo/RBO_00002024

Full taxonomy

Use of curated resources to describe variables (primary Darwin Core)

catalogNumber	kingdom	phylum	class	order	family	genus	scientificName
1	Animalia	Arthropoda	Insecta	Coleoptera	Carabidae	Carabus	Carabus auratus
2	Animalia	Arthropoda	Insecta	Coleoptera	Carabidae	Carabus	Carabus auratus
3	Animalia	Arthropoda	Malacostraca	Isopoda	Oniscidae	Oniscus	Oniscus asellus
4	Animalia	Arthropoda	Malacostraca	Isopoda	Oniscidae	Oniscus	
5	Animalia	Arthropoda	Malacostraca	Isopoda	Oniscidae	Oniscus	

lifeStage	sex	metabolic rate	Original Result Unit	body mass	Dry Mass	temperatureValue	country	verbatimlocality
adult	female	0,02617798	Watts	0,875	0,06365	5	NA	NA
adult	male	0,07338301	ml-hour	0,564	0,06365	10	NA	NA
NA	female	0,072122288	ml-hour	0,784	0,02038	15	Malaysia	Danum Valley Conservation Area, south-east Sabah
NA	male	0,178103254	µLO2-hour	0,684	0,03758	20	Malaysia	Danum Valley Conservation Area, south-east Sabah
NA	unsexed	0,016082891	µl O2-hour	0,967	0,06365	25	NA	NA
fry	NA	0,009523393						
fry	NA	0,012314669						
juvenile	unsexed	0,031881146						

verbatimCoordinates	decimalLatitude	decimalLongitude	habitat	license
NA	NA	NA	Terrestrial	https://www.journals.uchicago.edu/t-and-c paragraph 3.1.7
NA	NA	NA	Terrestrial	CC-BY 4.0
67 30' N, 26 40' E	67.53	26.66	Terrestrial	https://onlinelibrary.wiley.com/library-info/resources/text-and-datamining
67 30' N, 26 40' E	67.53	26.66	Terrestrial	https://onlinelibrary.wiley.com/library-info/resources/text-and-datamining
NA	NA	NA	Terrestrial	https://onlinelibrary.wiley.com/library-info/resources/text-and-datamining
41°30N; 145°50E-44°00N; 155°00E	41.76	-65.46	Aquatic	CC-BY 4.0
41 46 N 65 28' W	41.76	-65.46	Aquatic	CC-BY 4.0
NA				

Is Referenced By	references	body temperature regulation trait	Water body salinity	laboratory study
https://doi.org/10.1086/505997	10.1016/0198-0149(90)90104-4	ectotherm	NA	NA
https://doi.org/10.1086/505997	10.1016/0198-0149(90)90104-4	ectotherm	NA	No
https://doi.org/10.1086/505997	10.1016/0300-9629(73)90241-7	ectotherm	NA	NA
https://doi.org/10.1086/505997	10.1016/0300-9629(73)90241-7	ectotherm	NA	NA
	98-0149(90)90104-4	ectotherm	10	none specified
	org/10.1016/0300-9629(75)90146-2	ectotherm	30	Yes
https://doi.org/10.21203/rs.3.rs-1092818/v1	https://doi.org/10.1016/0300-9629(75)90146-2	ectotherm	20	Yes
https://doi.org/10.21203/rs.3.rs-1092818/v1	https://doi.org/10.1016/0300-9629(75)90146-2	ectotherm	30	Yes

1 Data file in tabular format

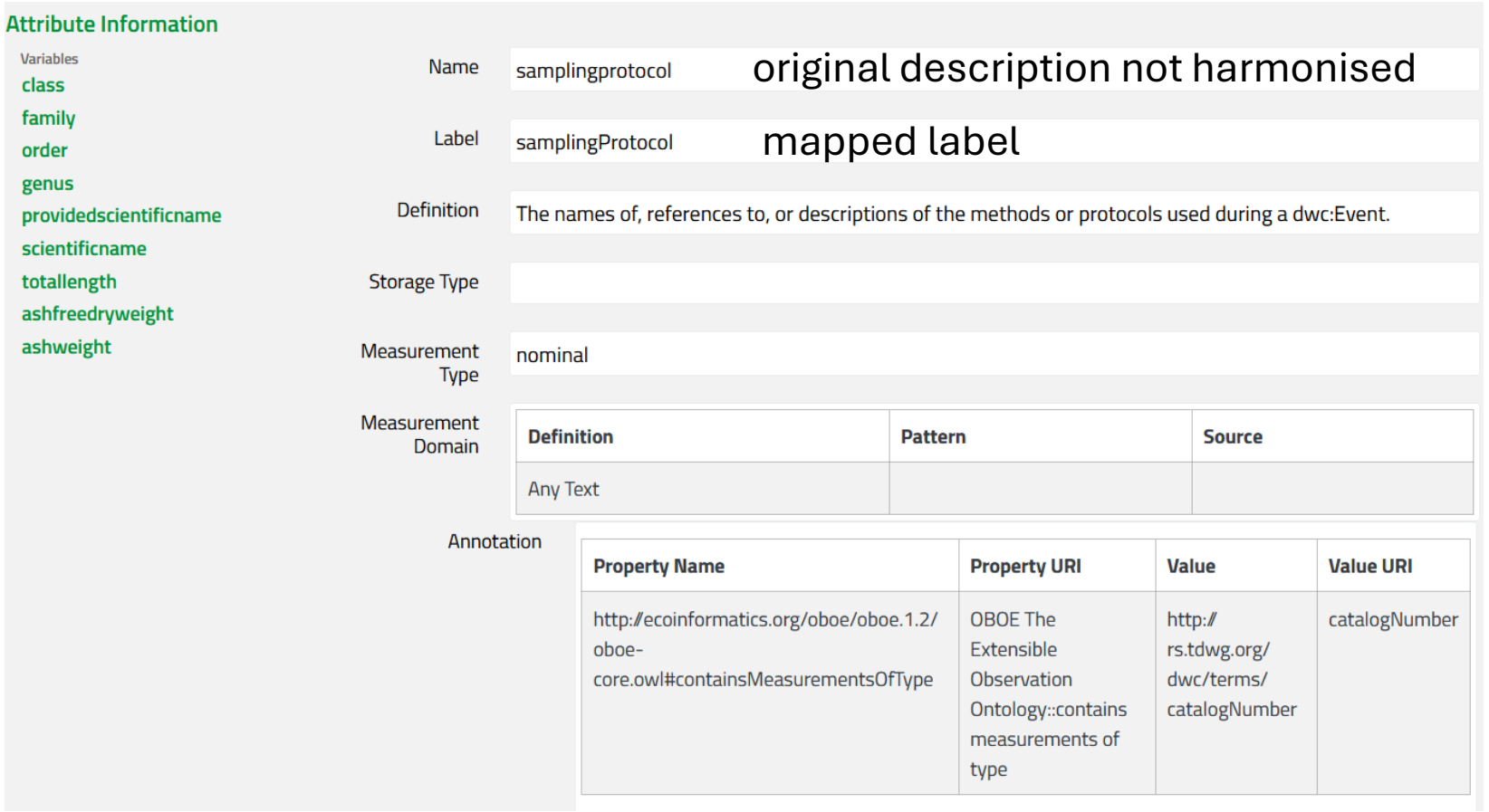
How we manage it now: semiautomatic annotation on data portal

UNIONE EUROPEA
Fondo Sociale Europeo
Fondo Europeo di Sviluppo Regionale

Ministero dell'Università e della Ricerca

PON RICERCA E INNOVAZIONE

LifeWatch PLUS ITALIA



URI

2 Annotation

Data harmonisation pipeline

Export

CSV
XML (EML 2.2.0)
JSON-LD

Data gathering

XLS
non standardised data

Structure and mapping

Structured and standardised CSV

Parasite species -> speciesName
Host species -> associatedTaxa
Sampling longitude-> decimaleLongitude
Sampling latitude -> decimalLatitude
Sampling coordinates-> verbatimCoordinates
...

Taxonomic check

LifeWatch ITALIA

Catalogue of Life

WoRMS World Register of Marine Species

Publication

Annotated dataset
+
Metadata in EML 2.2.0

LifeWatch ITALIA

LifeWatch ERIC

METADATA CATALOGUE

3 Metadata

Advantages of our approach

Integration with semantic platform and Data Labs

Extensible as user like

Interoperable with other systems

Export as Darwin Core Archive

Work in progress

Thank you ☺

Biodiversa+: Data management and capacity building on Darwin Core Standard workshop

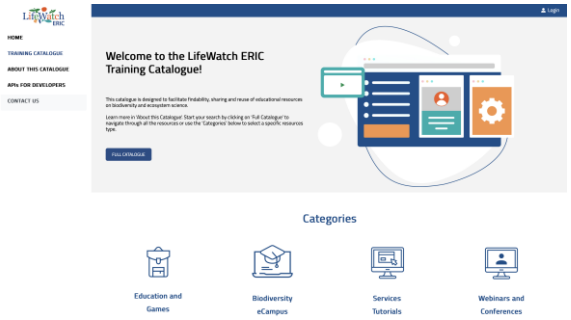
Building Capacity: developing skill and expertise & creating community of practice

Cosimo Vallo, LifeWatch ERIC Training Officer

LW ERIC Training Architecture

Training, Learning and Education

Training Catalogue

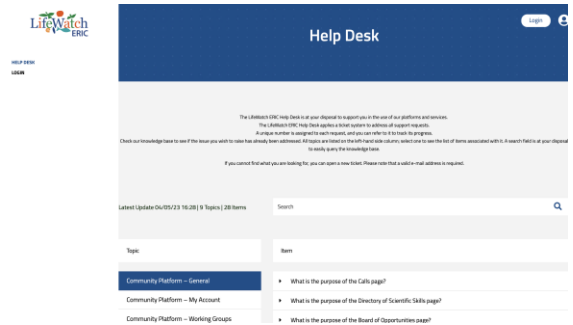


Training Platform

Users and Community Support

Community

Helpdesk



Training Catalogue

LW ERIC Training Catalogue hosts the metadata of learning resources

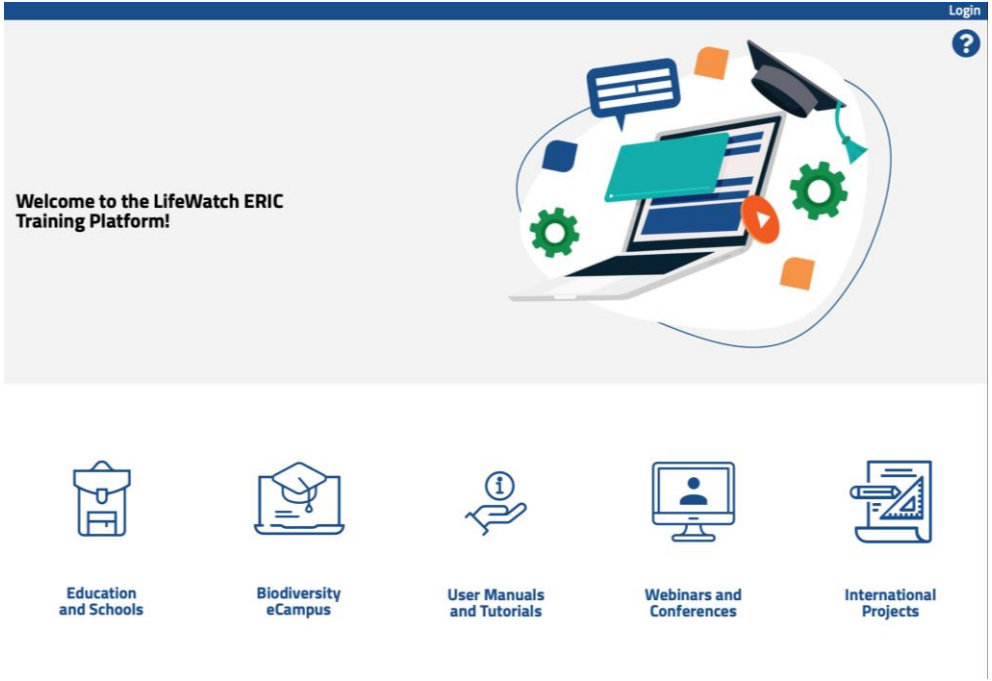
- **Enhances FAIRness**: ensures that resources can be shared, searched, discovered, accessed and reused
- **Increases visibility**: offers learning resources produced internally in LW ERIC and hosted on its Training Platform, as well as externally by partner institutions/communities
- **Allows contributors** to publish learning resources

https://trainingcatalogue.lifewatch.eu/home/

Training Platform

LW ERIC Training Platform addresses training, learning and education needs of the biodiversity research community

- **Training contents** constantly added and curated
- **User experience** is optimized with a user-friendly interface
- **Dedicated section(s)** can be developed to host project-specific contents (ex. EU projects)



https://training.lifewatch.eu/

Community

LW ERIC Community space dedicated to members of the wider scientific community

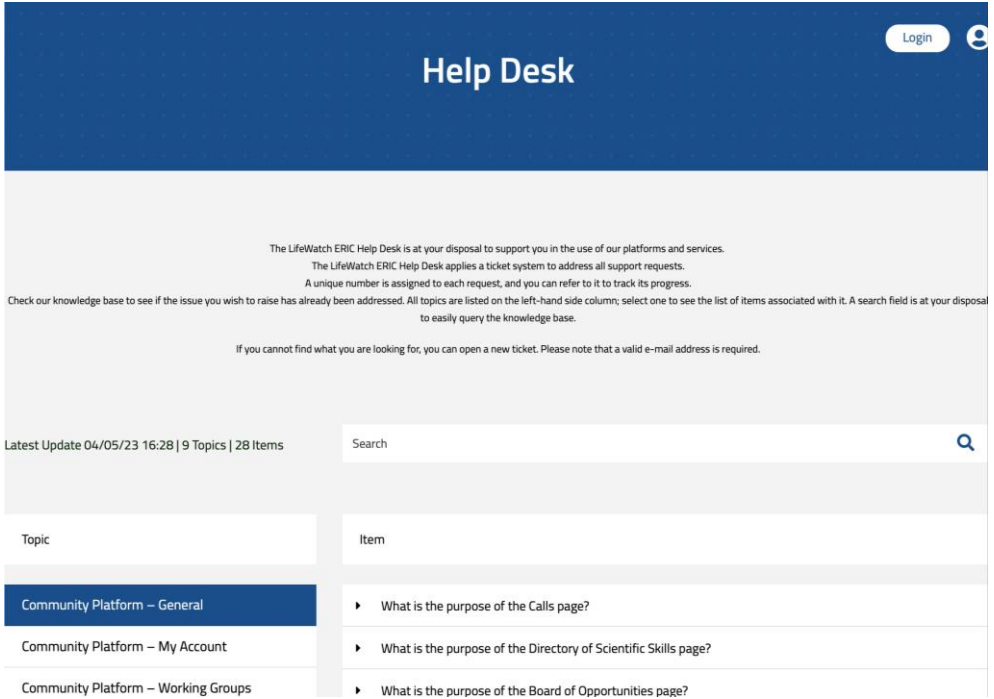
- **Meet, interact and collaborate**: interactive space for collaborative working and networking
- **Create Working Groups:** join researchers and experts working in the field of biodiversity and ecosystem research
- **Share, discover, find and promote** targeted opportunities and calls

https://community.lifewatch.eu

Helpdesk

LW ERIC Helpdesk supports the use of our platforms and services

- **Ticket system** to address all support requests. Unique number to refer and track progress of your request
- **Open new ticket** to get support and solve issues as well as to suggest improvements to our products
- **Knowledge base** of all the issues raised with a search field to easily query the knowledge base



https://helpdesk.lifewatch.eu

Meet the Team

Training Team at LifeWatch Service Center in Lecce

Cosimo Vallo

Training Officer

Eleonora Parisi

EU Projects Training Officer

Landing page: https://www.lifewatch.eu/training-and-education/

Training Action Coordination

LifeWatch ERIC created a **Training Working Group (TWG)** to collaborate among different offices and coordinate its training activities.

- **14 training focal points** from all its Common Facilities and National Distributed Centres that hold monthly coordination meetings
- **share, discuss, plan and design training** initiatives, activities and products
- members' engagement and information sharing contributes to **strengthen ERIC position in training within scientific community**
- Achievements: collaboratively developed Training Strategy, establishment of procedures and quality standards

Contact us

Get in touch with us at

training@lifewatch.eu

Biodiversa+: Data management and capacity building on Darwin Core Standard workshop

Building Capacity: developing skill and expertise & creating community of practice

Thank you for the attention!