



LifeWatch Italy data management strategy

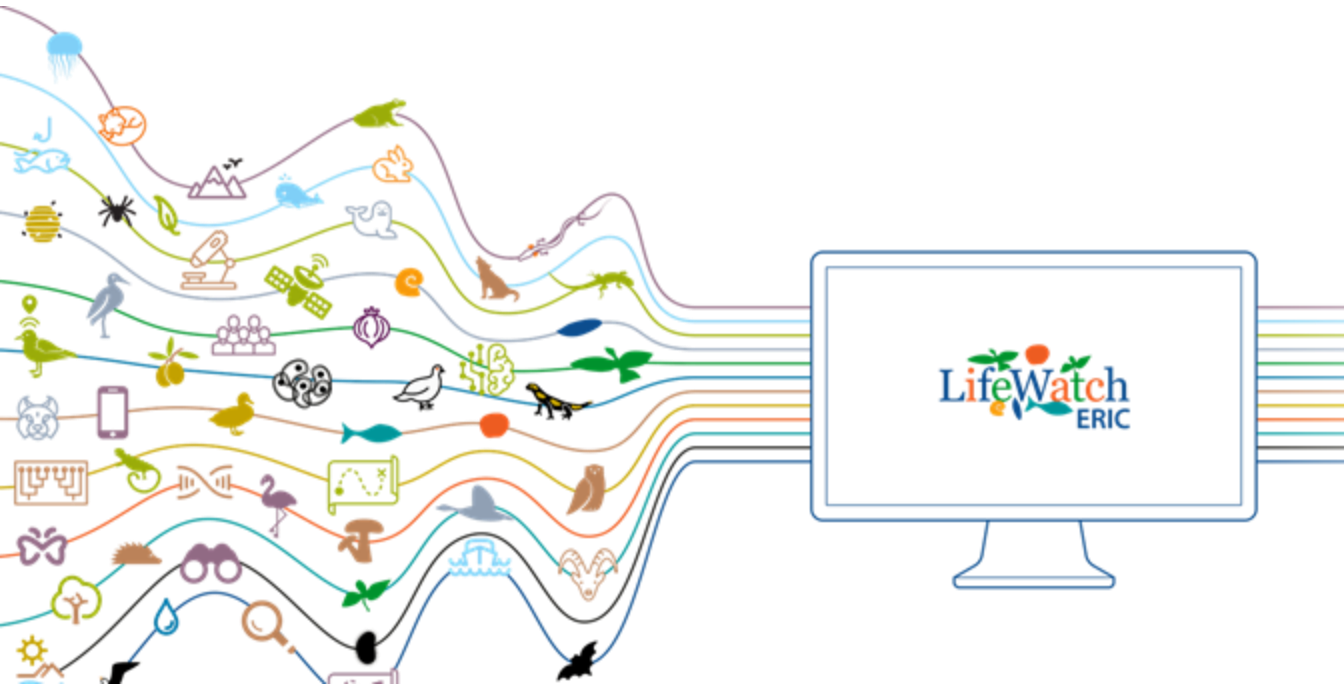
Martina Pulieri



UNIVERSITÀ
DEL SALENTO

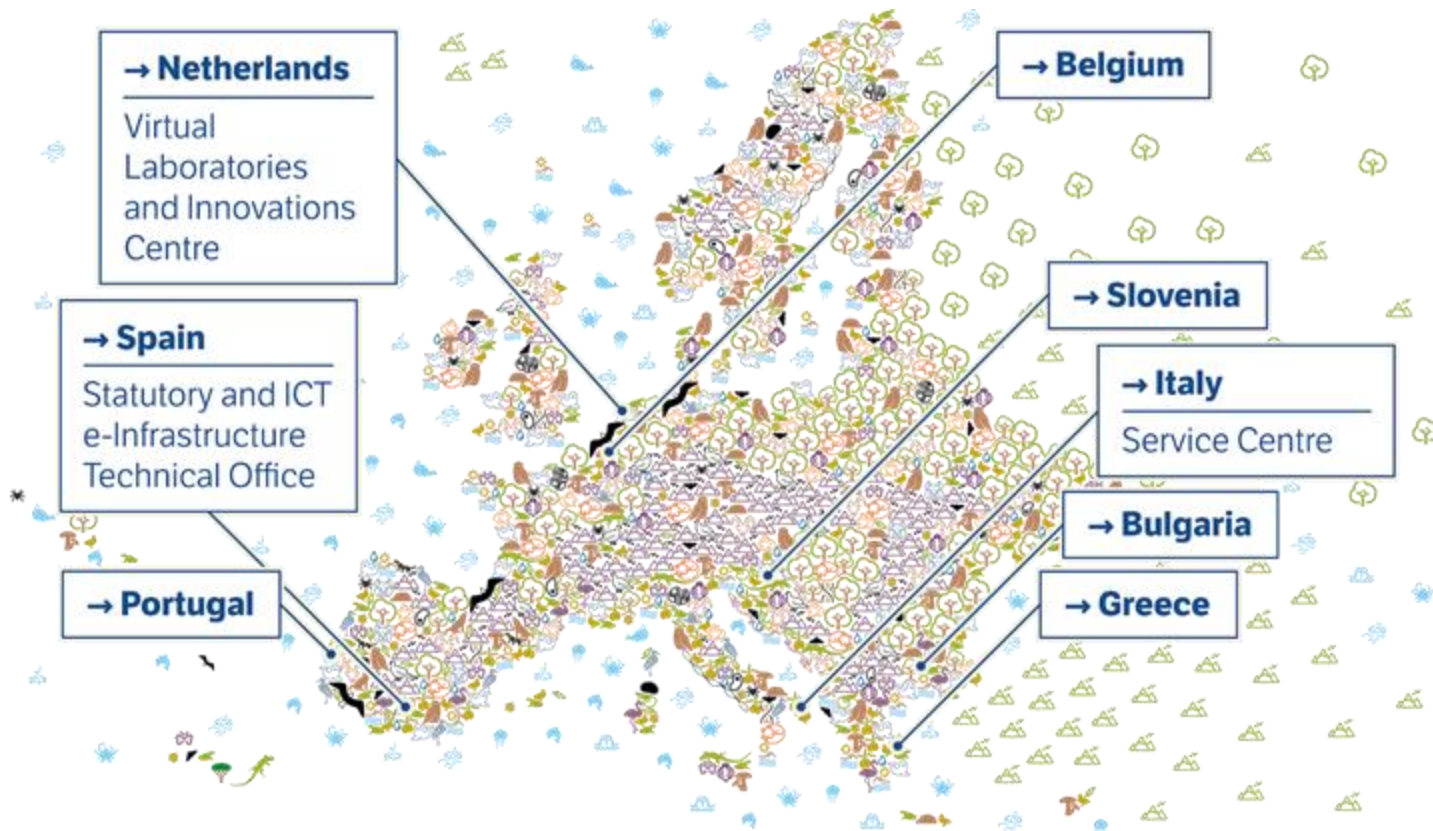


LifeWatch ERIC



LifeWatch ERIC is a European Research Infrastructure Consortium providing **e-Science research facilities** to scientists investigating biodiversity and ecosystem functions and **services** in order to support society in addressing key planetary challenges.

LifeWatch ERIC common facilities and members

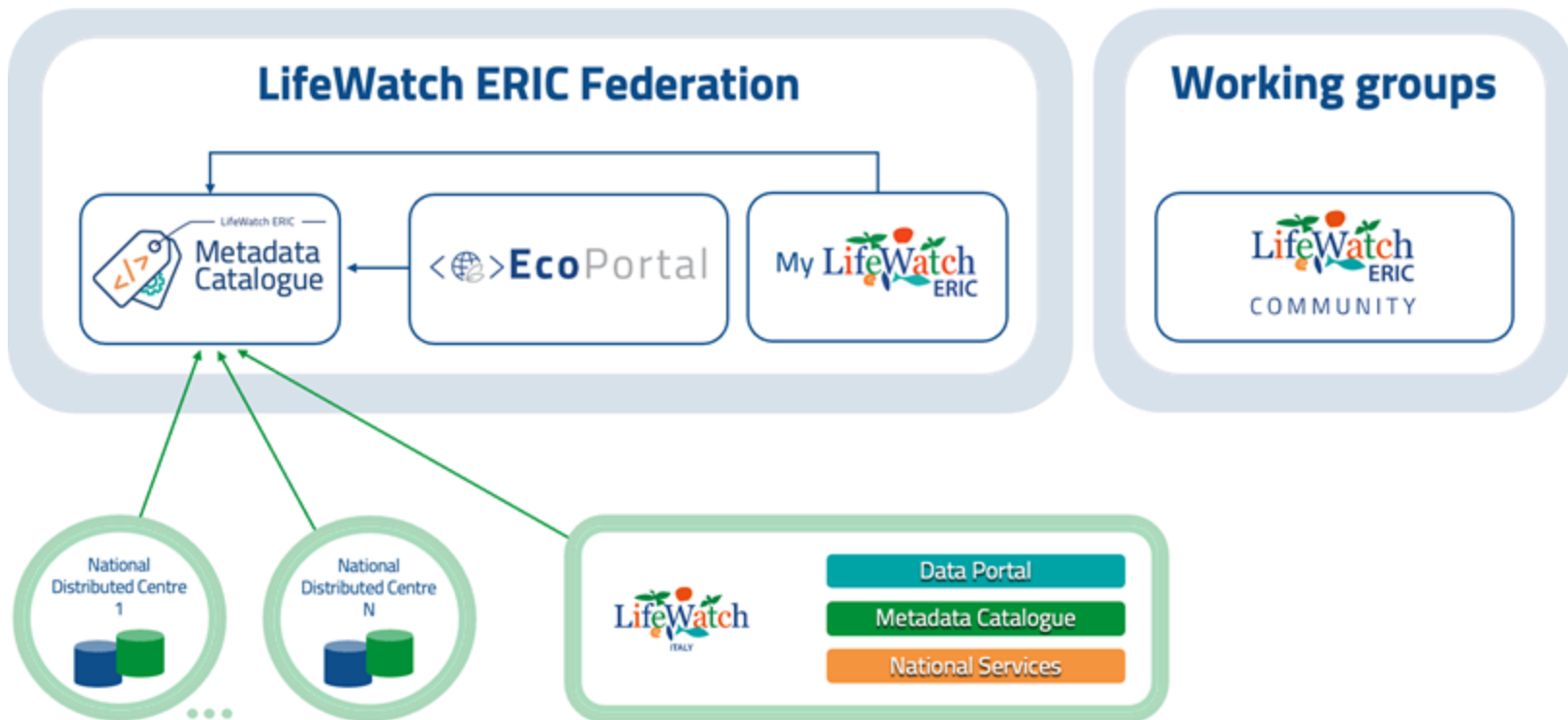


LifeWatch Italy JRU

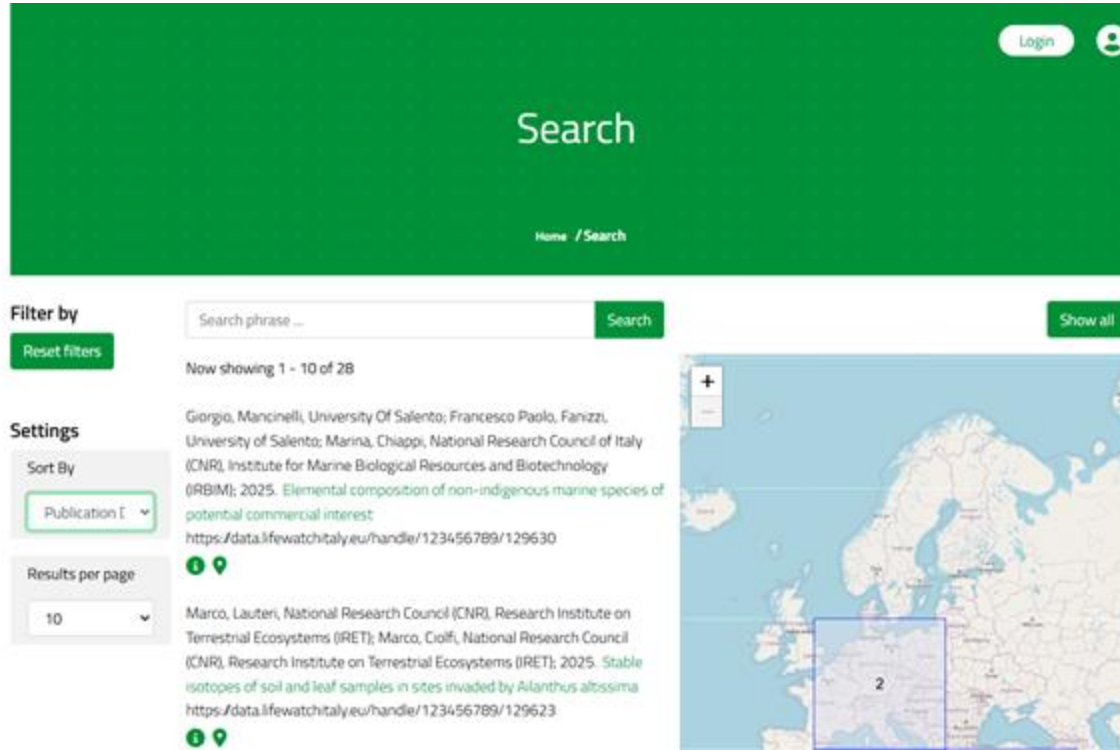
LifeWatch Italy has been established as a **Joint Research Unit (JRU)** in 2010 with 9 founding members led by **CNR**. It has now 35 members including research institutes, universities, associations and private companies distributed, seeking to reinforce integrated scientific research on biodiversity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



LifeWatch Italy contribution



LifeWatch Italy Data Portal



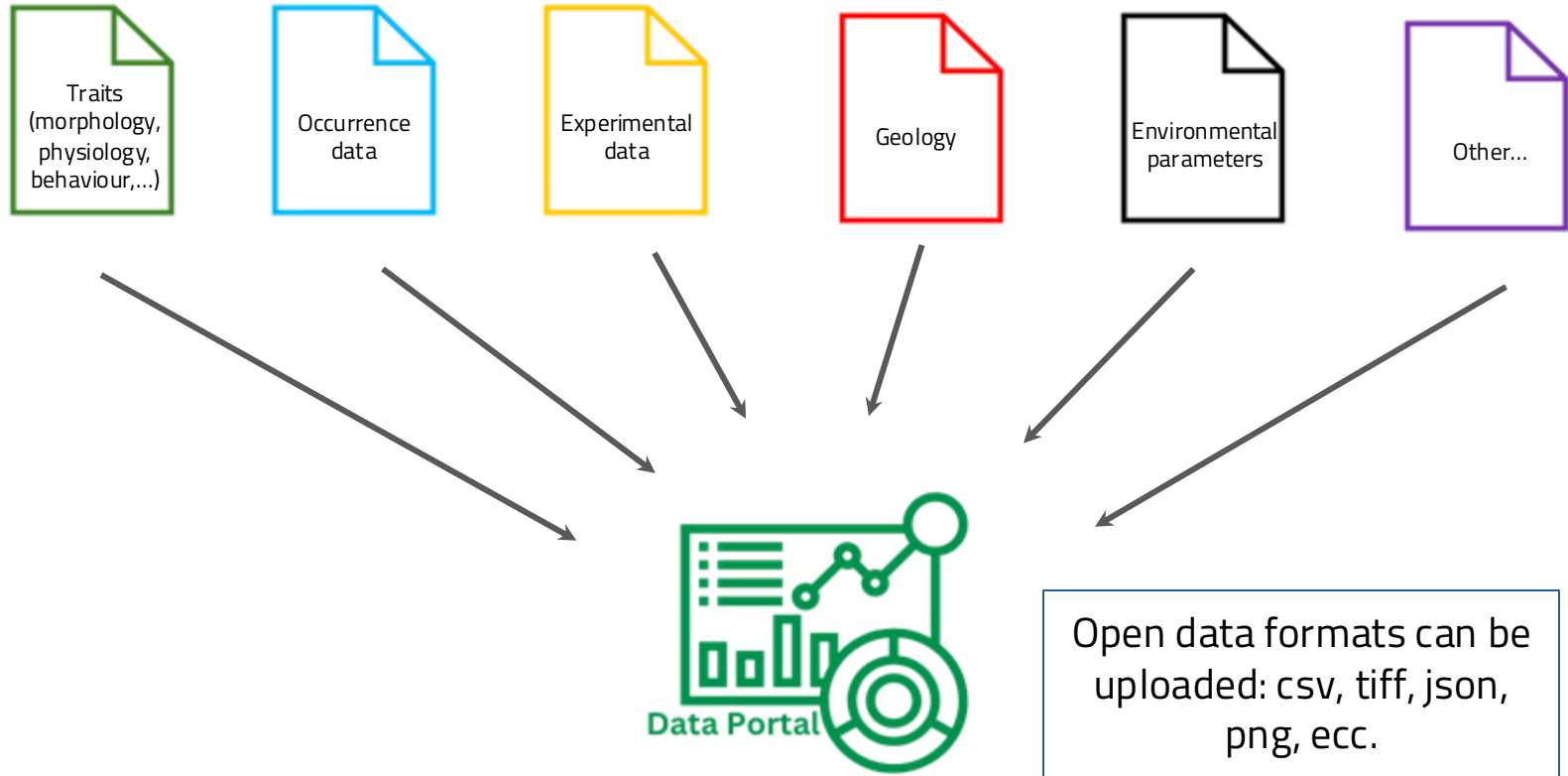
The screenshot displays the LifeWatch Italy Data Portal interface. At the top, a green header bar contains the 'Search' text and a 'Login' button with a user icon. Below the header, a sidebar on the left lists navigation options: DATA, SUBMIT DATA, SUPPORT, and ABOUT. The main content area features a search bar with the placeholder 'Search phrase ...' and a 'Search' button. Below the search bar, it indicates 'Now showing 1 - 10 of 28'. A 'Filter by' section includes a 'Reset filters' button. A 'Settings' section on the left allows users to 'Sort By' (currently set to 'Publication') and 'Results per page' (currently set to '10'). The main content area displays two search results. The first result is by Giorgio Mancinelli, Francesco Paolo Fanizzi, and Marina Chiappi, titled 'Elemental composition of non-indigenous marine species of potential commercial interest', with a URL starting with 'https://data.lifewatchitaly.eu/handle/123456789/129630'. The second result is by Marco Lauteri and Marco Colli, titled 'Stable isotopes of soil and leaf samples in sites invaded by *Alanthus altissima*', with a URL starting with 'https://data.lifewatchitaly.eu/handle/123456789/129623'. A map of Europe is shown on the right, with a blue box highlighting Italy and a '2' indicating the number of results from that region.

A unique access point to **FAIR** and **Open** ecological and biodiversity data.

Three roles:

- Researcher
- Reviewer
- Administrator

What data we handle?



Our problem

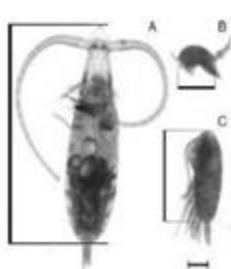
Ecological and biodiversity data are highly diverse in:

- **content** (traits, occurrence, behaviour, abiotic data);
- **structure** (long table, wide table);
- **semantics** (the same concept expressed with different name).

SiteCode	TaxonCode	species name	authorship	AphiaID	group	percent abundance
TY0166	AV	Acanthocyclops vernalis	(Fischer, 1853)	157729	Copepoda	60,00
TY0170	AV	Acanthocyclops vernalis	(Fischer, 1853)	157729	Copepoda	3,91
CA0005	AD	Acanthodiaptomus denticornis	(Wierzejski, 1887)	356841	Copepoda	33,95
CA0026	AD	Acanthodiaptomus denticornis	(Wierzejski, 1887)	356841	Copepoda	27,60
CA0042	AH	Acroperus harpae	(Baird, 1834)	148407	Cladocera	29,17
CA0047	AH	Acroperus harpae	(Baird, 1834)	148407	Cladocera	31,77
PT0026	AH	Acroperus harpae	(Baird, 1834)	148407	Cladocera	1,78
PT0027	AH	Acroperus harpae	(Baird, 1834)	148407	Cladocera	4,20
PT0030	AH	Acroperus harpae	(Baird, 1834)	148407	Cladocera	8,20
PT0058	AH	Acroperus harpae	(Baird, 1834)	148407	Cladocera	1,56

my_sp_name	body_length_cm	antenna_length_cm	...
Agonum_ericeti	0.587	0.42	
Agonum_gracilis	0.480	0.30	
...	...	...	

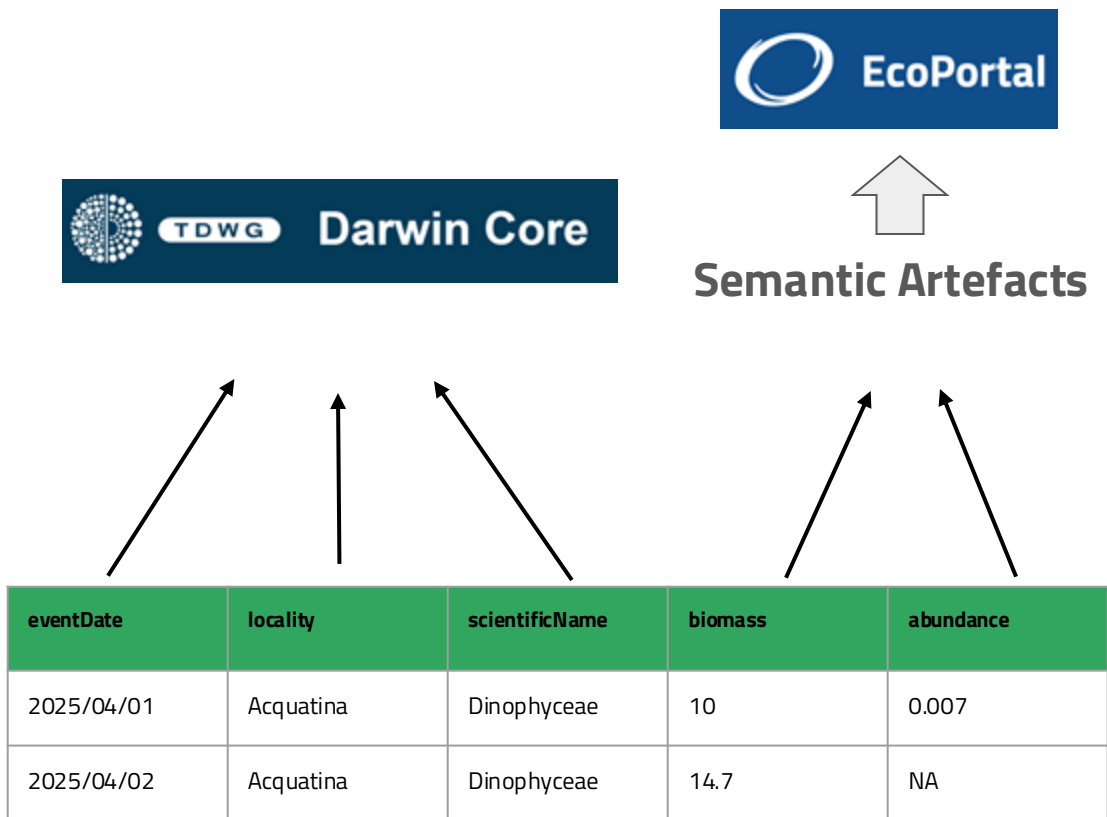
scientificName	traitName	traitValue	traitUnit
Agonum ericeti	Body_length	5.87	mm
Agonum ericeti	Antenna_length	4.2	mm
Agonum gracile	Body_length	4.80	mm



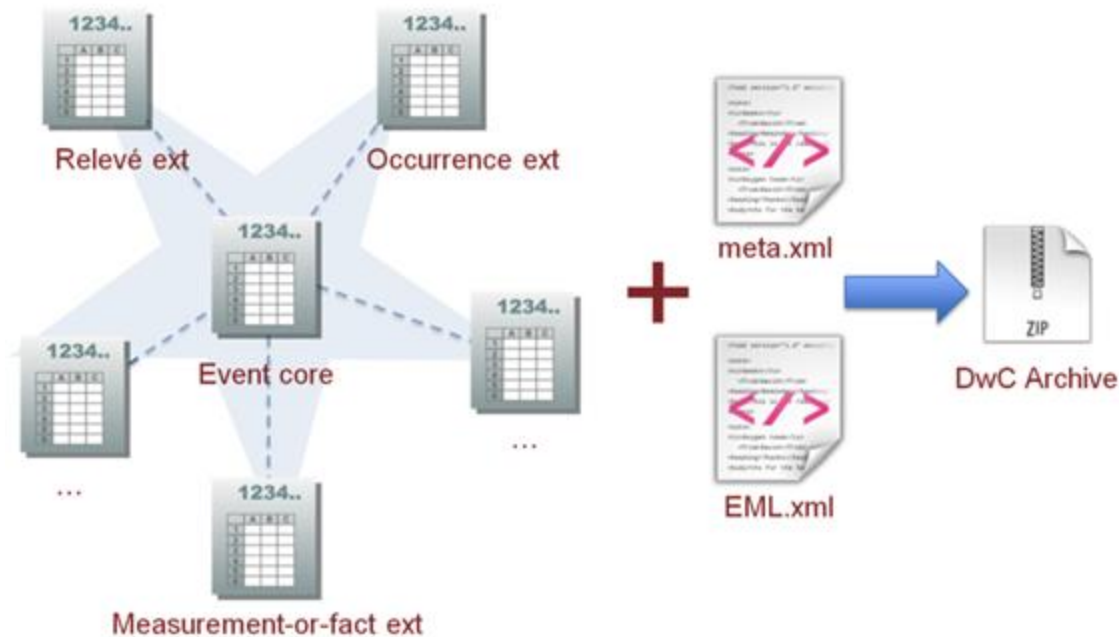
Size	Sex	Life stage	Units
Prosoma length	Female	copepodite	mm
Length	F	Cop 5	micron
Body length	Fem	Cop V	um
Body size	♀	Copepodite	µm
BL	Adult F	C5	cm
Total length		CV	

Our approach

- Temporal, spatial and taxonomic information are described using **Darwin Core**.
- Other information about measurement using **semantic artefacts** (e.g. glossaries, thesauri, ontologies).



Our approach



We don't use the entire Darwin Core because of the heterogeneity of our data. Most of them don't fit in the data schema.

Measurements Or Facts extension is not a good option for our data because several measurement are taken at an individual level. This will results in very long and unmanageable datasets.

PTOC	LTER-MC	40°48.5'N, 14°15'E					
PTOC	Cell Abundance (Cells ml ⁻¹)						
PTOC				26/01/1995	09/02/1995	01/03/1995	12/04/1995
PTOC	1	D	Achnanthes brevipes	0,00	0,00	0,00	0,00
PTOC	1	D	Achnanthes spp.	0,00	0,00	0,00	0,00
	1	D	Amphora spp.	0,00	0,00	0,00	0,00
	1	D	Asterionellopsis glacialis	8,50	775,20	2,72	0,00
	1	D	Asteromphalus cleveanus	0,00	0,00	0,00	0,00
	1	D	Asteromphalus spp.	0,00	0,00	0,00	0,00

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Depth water [m]	A. senarius [H]	A. arachne [H]	A. heptactis [H]	C. concinnus [H]	C. granii [H]	D. pumila [H]	D. brightwellii [H]	F. doliolus [H]	G. striata [H]	L. mediterraneus [H]	L. undulatum [H]	N. longissima [H]	P. indica [H]	P. calcar-avis [H]
0		400			400			6000	200	5000			600	4600
4		800	200		400	800	600	2400	400	6200	800		400	200
8	200	1400		400	1600	600		6000	400	3000	1800		400	4400

LifeWatch Italy Data Portal



DATA

+

JOBS

SUBMIT DATA

-

Dataset

Portal



Drop files to attach them to the item, or [browse](#)

General information

Person organisation

Coverage

License

Project Detail

Methods

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.

Data Curation and Validation

Drop files to attach them to the item, or **browse**



LW-ITA Data Schema based on DarwinCore and controlled vocabularies

scientificName	samplingProtocol	bodyLenght
<i>Idotea balthica</i>	hand operated ekman grab	4.5



scientificName
<i>Idotea balthica</i>



Taxonomic check

Validated data



Data Curation and Validation

Target Data

Select Domain

Select Environment

Select Taxonomy

Matching summary result

1

Select Actions

2

Confirm Actions

File name: e4cae767-c998-494c-a371-732bcd2f0bb2 - Dataset: Macrozoobenthos data collected in the Acquatina lagoon, Apulia, Italy

This matching resulted in **6%** compliance in Italian Backbone.
The remaining **94% (no match)** have this matching: 42% in CoL, 1% in WFO, 99% in WoRMS.

No Match
N. records: 13892/14735

Match with warning
N. records: 0/14735

0 records to delete

0 requests to send

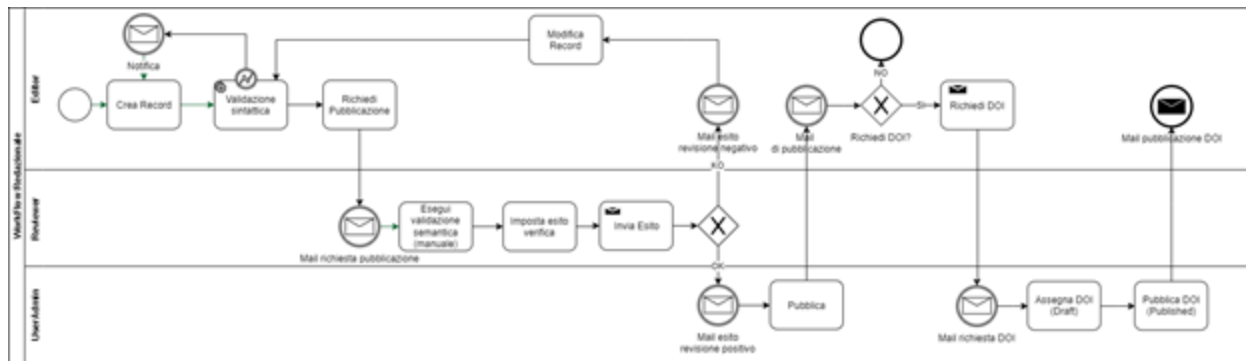
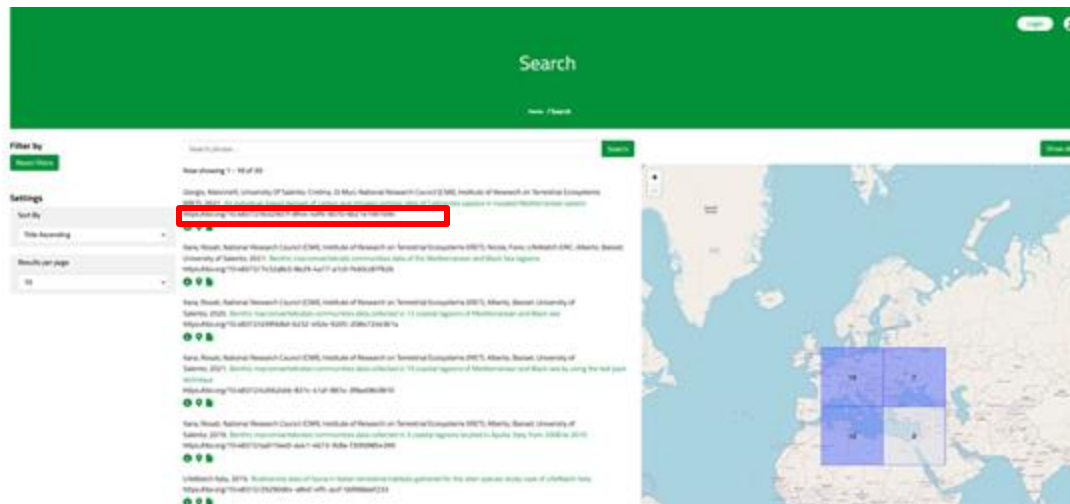
Scientific name	CoL	WFO	WoRMS	Actions
Hydrobia - 778 records			✓	Keep the rows
Idotea balthica - 25 records			✓	Keep the rows
Lekanesphaera hookeri - 10 records			✓	Keep the rows
Lekanesphaera monodi - 142 records			✓	Keep the rows
Leptochelia savignyi - 108 records			✓	Keep the rows

Reviewer



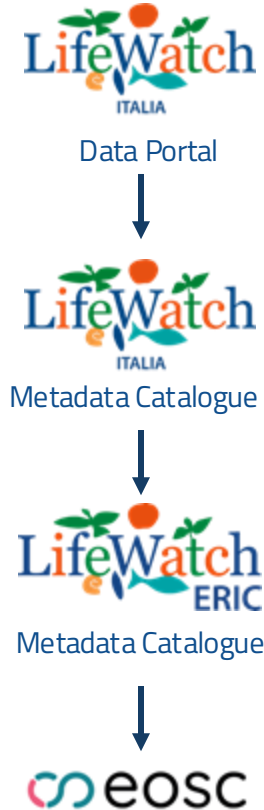
Administrator
Reviewer

After data and metadata validation from reviewers, a DOI can be assigned to their datasets.





Out of LifeWatch



Metadata are harvested in the Metadata Catalogue of LifeWatch Italy and to LifeWatch ERIC one.

The metadata catalogue of LifeWatch ERIC is already a node of EOSC and data will be findable, accessible and discoverable by a wider community.

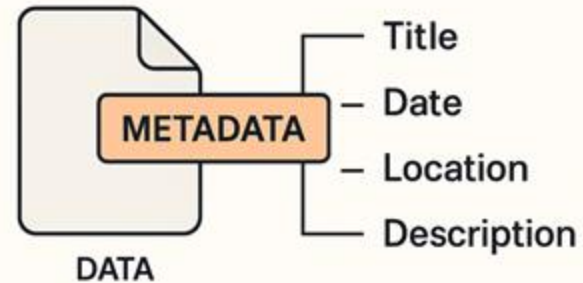
Metadata creation & management



Eftychia Tzafesta

What is metadata?

Metadata are the “data about data”, the descriptors of data.



Questions metadata should answer

Questions	Examples
Who	Who collected the data? Who processed the data? Who funded the research that led to the data? Who is the primary contact for the data? Who is permitted to use the data?
What	What data were collected? What questions or hypotheses were being addressed? What QA/QC procedures were used? What data gaps exist? What do the variable labels mean? What do the codes mean? What limitations do the data have? What software do I need to read the data?
When	When were the data collected? When were the data processed and analyzed? When were or will the data be made available?
Where	Where were the data collected? Where were the data processed and analyzed? Where are the data stored?
Why	Why were the data collected? Why were specific collection, processing and analytical methods employed?
How	How were the data collected and processed? How precisely and accurately were the data values measured? How can the data be accessed? How are the data organized and structured? How should the data be cited?

Why metadata matters?

Metadata is the key to making data useful

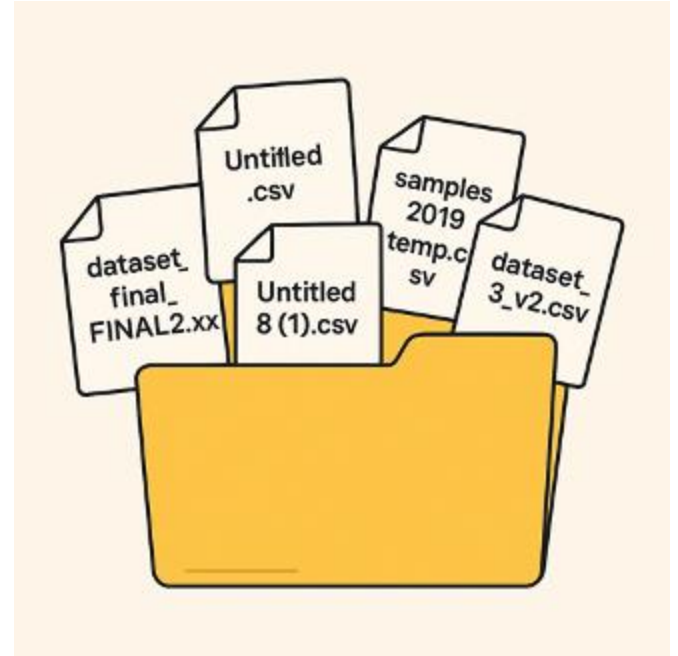
- Enables data discovery and access
 - Without metadata, data is invisible in a sea of files.
- Supports data interpretation and reuse
 - Helps others (and your future self!) understand and trust the dataset.
- Ensures transparency and reproducibility
 - Especially important in scientific research and long-term monitoring.
- Facilitates data sharing and integration
 - Metadata makes it possible to combine datasets across disciplines, regions, and time.



“Metadata turns data into information.”

What happens without metadata?

Without metadata, even well-collected data becomes unusable, unfindable, and untrustworthy.



What is the value to data users?

Metadata helps users to:

-  **Search** for relevant datasets
-  **Find** datasets by topic or location
-  **Evaluate** if the dataset fits their needs
-  **Access** and **understand** how to use the data
-  **Interpret** columns, units, values, etc.

Why use metadata standards?

- **Ensure consistency** – everyone uses the same terms in the same way
- **Enable interoperability** – metadata can be understood across systems
- **Support machine-readability** – software can parse and interpret metadata
- **Improve data discovery** – helps users find and access data easily
- **Facilitate reuse** – clear and structured metadata
- **Save time** – reduce confusion and prevent duplicated effort

 *Without standards, metadata becomes messy, unclear, and hard to use.*

FAIR principles



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

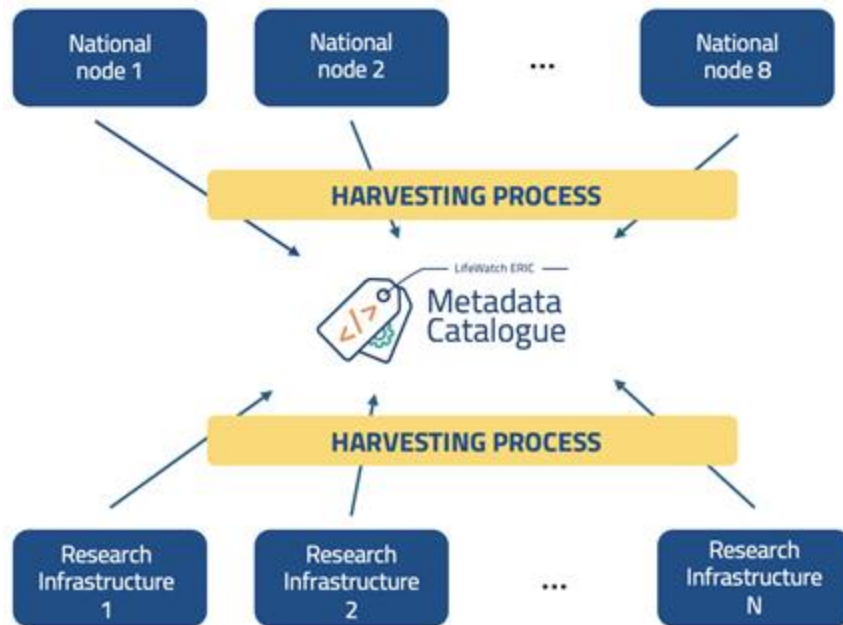


Image: Icons by Freepik from www.flaticon.com and ARDC
<https://conference.eresearch.edu.au/fair-go-new-resources-to-support-fair-data/>

LifeWatch ERIC Metadata Catalogue

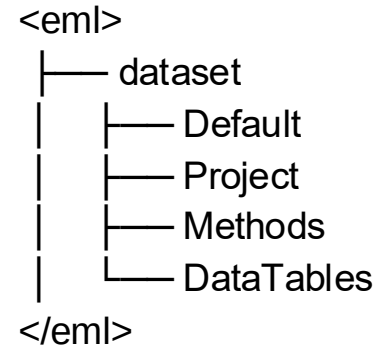
A platform for harvesting (e.g. LW Italy Data Portal), storing and publishing metadata about datasets in biodiversity and ecosystem research.

- Based on a customised **GeoNetwork** platform
- Uses **EML 2.2.0** as one of its supported metadata standards

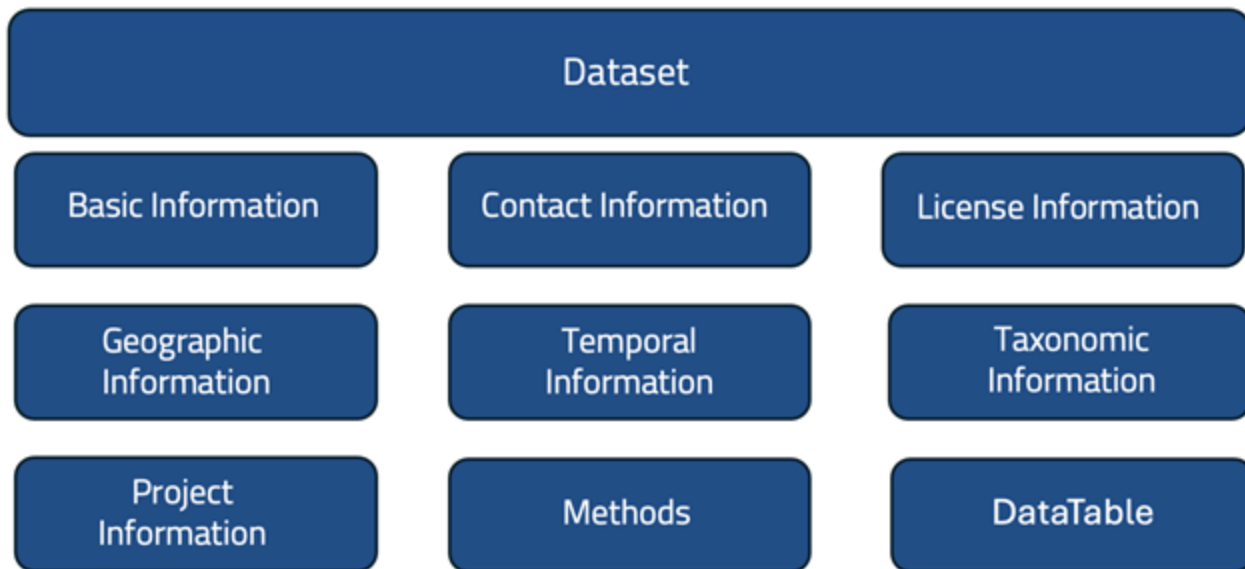


Ecological Metadata Language (EML)

- XML-based metadata standard developed for ecology and environmental science
- Every section contains a set of attributes, e.g. Title, Publication date, Temporal coverage, Taxonomic coverage, etc, to thoroughly describe your data.
- EML is particularly well-suited for ecological datasets because it provides structured fields for describing key research components, such as sampling methods, data collection protocols, equipment and software used.

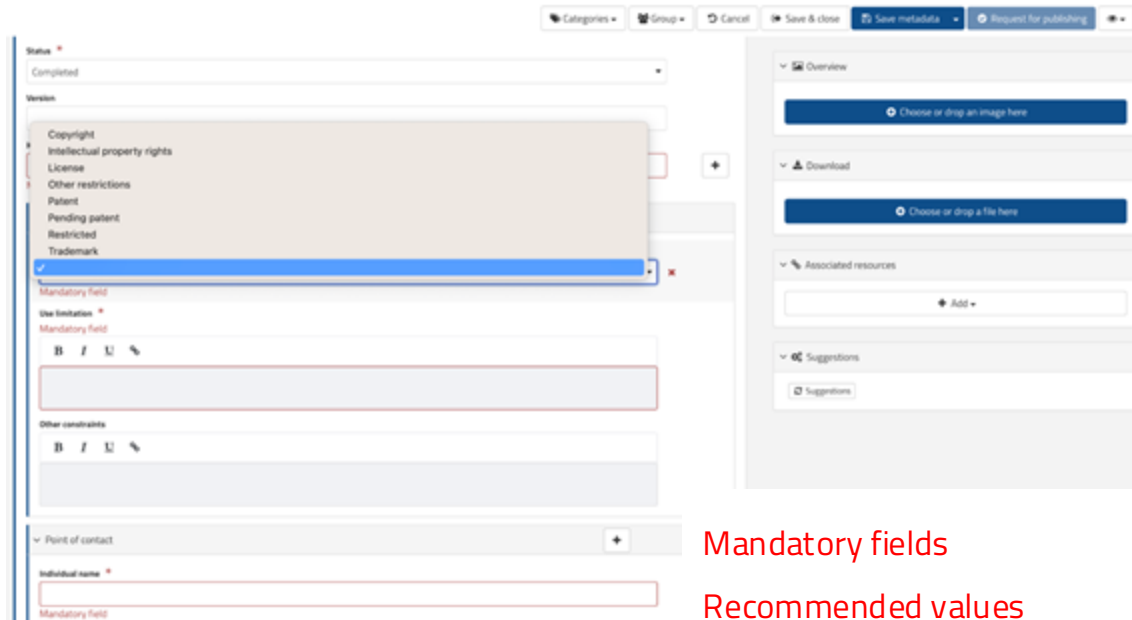


LifeWatch ERIC Metadata Catalogue



LifeWatch ERIC Metadata Catalogue

Metadata validation



The screenshot displays a metadata form with several validation features highlighted in red:

- Mandatory fields:** Indicated by a red asterisk (*) next to the "Status" field (set to "Completed") and the "Individual name" field.
- Recommended values:** A dropdown menu for "Copyright" is open, showing a list of options: "Intellectual property rights", "License", "Other restrictions", "Patent", "Pending patent", "Restricted", and "Trademark". The "Trademark" option is highlighted with a blue bar.
- Controlled vocabularies:** A red box highlights the "Use limitation" field, which contains a list of icons (B, I, U, S) representing different types of limitations.

The form also includes a sidebar on the right with sections for "Overview", "Download", "Associated resources", and "Suggestions". At the top, there are buttons for "Categories", "Group", "Cancel", "Save & close", "Save metadata", and "Request for publishing".

Mandatory fields

Recommended values

Controlled vocabularies

LifeWatch ERIC Metadata Catalogue

```

▼<keywordSet>
  <keyword>Biodiversity</keyword>
  <keyword>Taxonomy</keyword>
  <keyword>Mediterranean basin</keyword>
  <keywordThesaurus>Biodiversity Thesaurus</keywordThesaurus>
</keywordSet>
▼<keywordSet>
  <keyword>Black Sea</keyword>
  <keywordThesaurus>Biogeographical regions</keywordThesaurus>
</keywordSet>
▼<keywordSet>
  <keyword>Invertebrates</keyword>
  <keyword>Lagoon</keyword>
  <keywordThesaurus>eLTER Controlled Lists</keywordThesaurus>
</keywordSet>

```

- Data harmonisation
- Data standardisation
- Data reuse

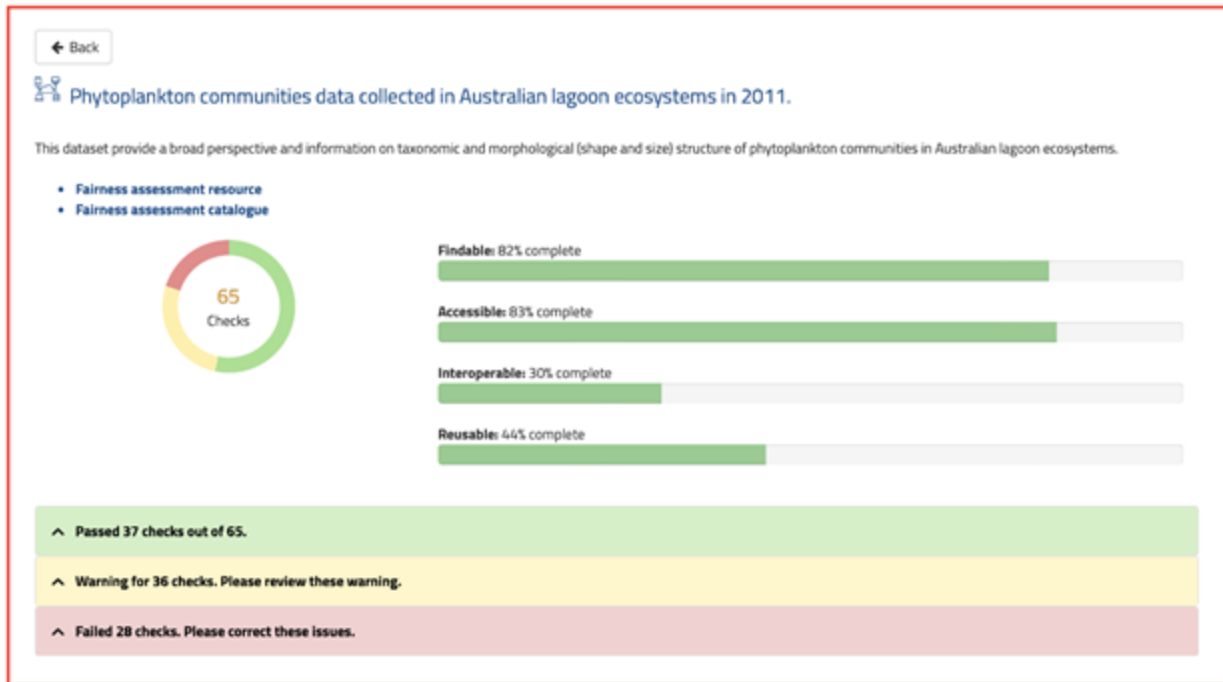
```

▼<attribute>
  <attributeName>phylum</attributeName>
  <attributeLabel>Phylum</attributeLabel>
  <attributeDefinition>The full scientific name of the phylum or division in which the taxon is classified.</attributeDefinition>
  ▼<measurementScale>
    ▼<nominal>
      ▼<nonNumericDomain>
        ▼<textDomain>
          <definition>Any Text</definition>
        </textDomain>
      </nonNumericDomain>
    </nominal>
  </measurementScale>
  ▼<annotation>
    <propertyURI label="has context">http://ecoinformatics.org/oboe/oboe.1.2/oboe-core.owl#hasContext</propertyURI>
    <valueURI label="Phylum">http://rs.tdwg.org/dwc/terms/phylum</valueURI>
  </annotation>
</attribute>

```

LifeWatch ERIC Metadata Catalogue

FAIRness assessment



LifeWatch ERIC Metadata Catalogue

Metadata publication

All changes saved

Categories Groups Cancel Save & close Save metadata Request for publishing

Default Project Methods Data Tables

Identification

Alternate Identifier
50523767-7652-4300-9e99-525a77a70708

Publication Date
02/04/2023

Title
Phytoplankton communities data collected in Maldivian lagoon ecosystems

Short Name

Abstract
B I U
This dataset provides a broad perspective and information on taxonomic and morphological (shape and size) structure of phytoplankton communities in Maldivian lagoon ecosystems.

Dataset Language
en

Dataset Creator

Organization Name
Regional Agency for the Environmental Prevention and Protection

Individual Name

Given 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".

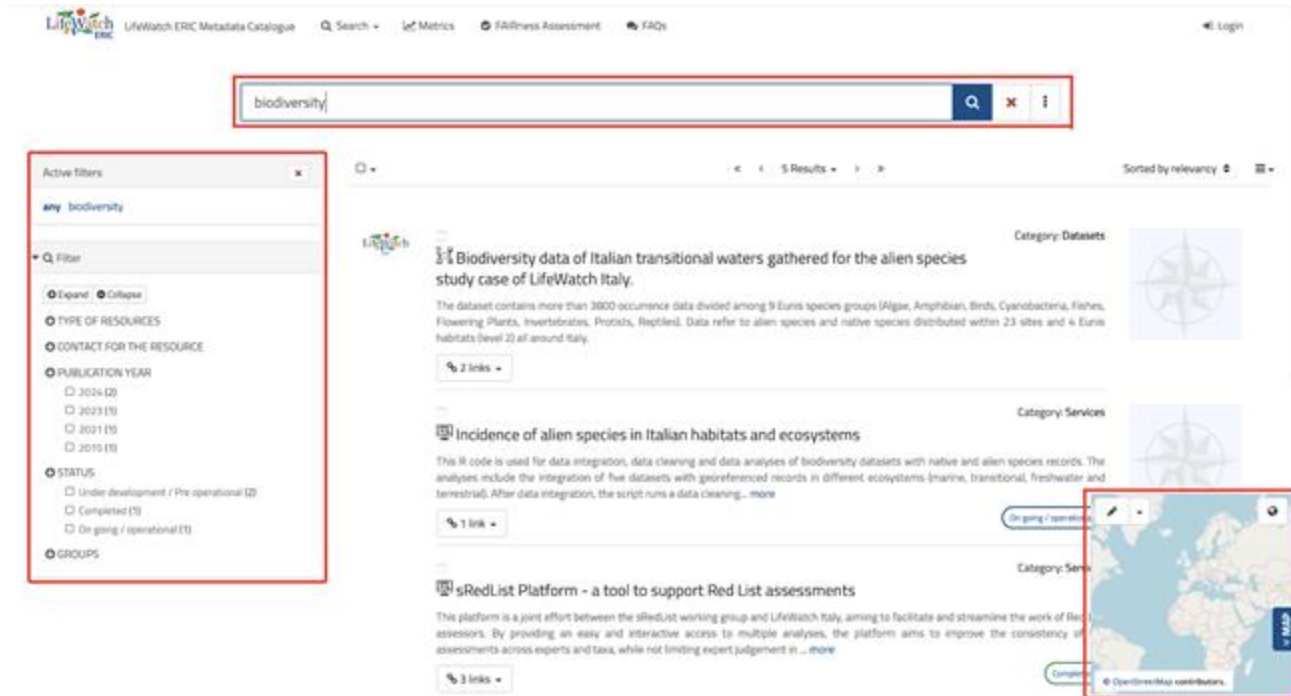
If your metadata record does not have a DOI, you can ask for it right now by selecting the following checkbox. The DOIs are assigned only upon verification and validation of the metadata record.

☒ I want to get a DOI

Cancel Send

LifeWatch ERIC Metadata Catalogue

Discovery and reuse of resources by their metadata



The screenshot displays the LifeWatch ERIC Metadata Catalogue interface. At the top, the LifeWatch ERIC logo is on the left, and navigation links for Search, Metrics, Fairness Assessment, and FAQs are in the center. A user login link is on the right. Below the navigation bar is a search bar containing the text "biodiversity". To the left of the main content area is a sidebar with "Active filters" showing "any biodiversity". Below this is a "Filter" section with expand/collapse buttons and several filter categories: "TYPE OF RESOURCES", "CONTACT FOR THE RESOURCE", "PUBLICATION YEAR" (with checkboxes for 2024 (2), 2023 (1), 2021 (1), and 2010 (1)), "STATUS" (with checkboxes for Under development / Pre operational (2), Completed (1), and On going / operational (1)), and "GROUPS". The main content area shows search results. The first result is titled "Biodiversity data of Italian transitional waters gathered for the alien species study case of LifeWatch Italy." and includes a description of the dataset and a "2 links" button. The second result is titled "Incidence of alien species in Italian habitats and ecosystems" and includes a description of the R code and a "1 link" button. The third result is titled "sRedList Platform - a tool to support Red List assessments" and includes a description of the platform and a "3 links" button. On the right side of the results, there are category filters for "Category: Datasets" and "Category: Services", each with a compass icon. At the bottom right, there is a map of Europe with a red box highlighting a specific area, and a "Map" button.

Team



Ilaria Rosati

CNR-IRET



Lucia Vaira

LifeWatch ERIC



Davide Raho

CNR-IRET



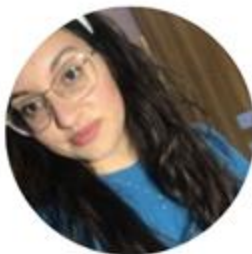
Francesco De Leo

CNR-IRET



Cristina Di Muri

CNR-IRET



Martina Pulieri

University of Salento



Mariantonietta La Marra

University of Salento



Eftychia Tzafesta

LifeWatch ERIC



Majid Fiaz

LifeWatch ERIC

THANK YOU!

martina.pulieri@unisalento.it

eftychia.tzafesta@lifewatch.eu