

27-29 May 2024 



imdīs

International conference on Marine Data and Information Systems



MARIS

MEDIN



National
Oceanography
Centre



iImagine



eosc
Blue-Cloud2026



EMODnet

European Marine
Observation and
Data Network

Biology



A Comprehensive Approach to Marine Biodiversity Data Validation: The Biocheck Tool

Session SERVICES: Data services and tools in ocean science

- **Ruben Perez Perez, VLIZ**
- **Joana Beja, VLIZ/EMODnet Biology coordinator**

- 27th to 29th May 2024, IMDIS (Bergen, Norway)



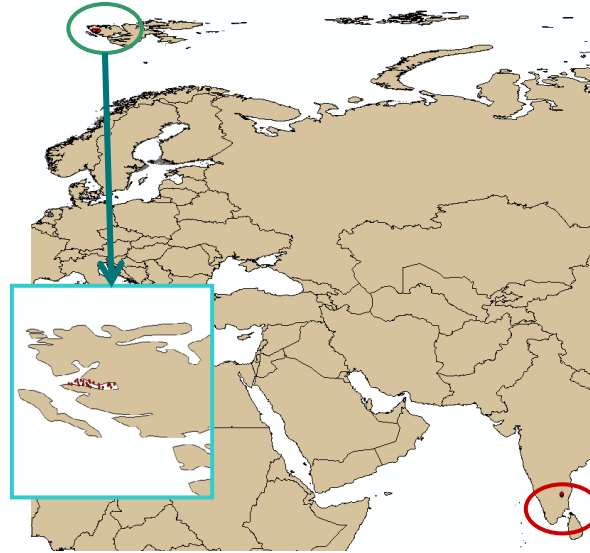
The European Marine Observation and Data Network (EMODnet) is financed by the European Union under Regulation (EU) 2021/1139 of the European Parliament and of the Council of 7 July 2021 establishing the European Maritime, Fisheries and Aquaculture Fund.

Rationale

Data errors are ubiquitous

- Human errors
 - Manual data entry
 - Misinterpretation of data
 - Oversight
- Technical issues
 - Integration failures
 - Software errors

“Monitoring in Kongsfjorden area”



Latitude - longitude switched

“Monitoring in Belgian part of the North Sea”



“+” & “-” signs switched

Rationale

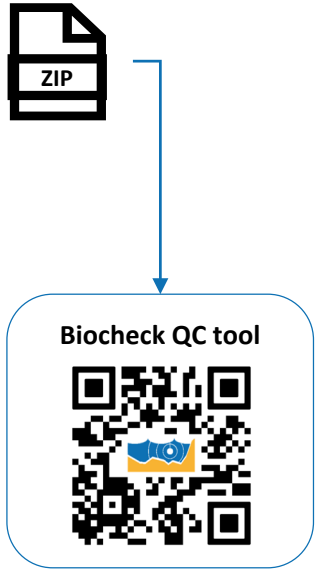
The impact of data errors in science

“In data management, Quality control is the process of identifying and correcting inaccuracies or inconsistencies in data to ensure its accuracy, completeness and reliability.”

| | Species names before quality control | | | | | Species names after quality control | | | | |
|------------------|--------------------------------------|----------------|---------|---------|-------|-------------------------------------|----------------|---------|---------|-------|
| | # Species | # Rare species | H' | $1 - D$ | ES50 | # Species | # Rare species | H' | $1 - D$ | ES50 |
| Rocky shore data | | | | | | | | | | |
| ANE | 219 | 15 | 4.63602 | 0.98777 | 38.11 | 187 | 11 | 4.45772 | 0.98509 | 36.25 |
| Arctic | 646 | 69 | 6.00024 | 0.99666 | 46.33 | 378 | 44 | 5.38261 | 0.99403 | 43.67 |
| Mediterranean | 1,120 | 238 | 5.74091 | 0.99342 | 43.35 | 834 | 159 | 5.49015 | 0.99105 | 41.74 |
| North Sea | 251 | 29 | 4.50662 | 0.98424 | 35.89 | 163 | 25 | 3.95956 | 0.97469 | 30.14 |

The Biocheck tool

Quality controlling biodiversity data



The Biocheck tool

Quality controlling biodiversity data



Data Overview

Biocheck QC tool



BIOFUN1 1

event type: Cruise

BF1A01 14

event type: Sample

measurement types: Gear, sampling net horizontal opening, sampling net mesh :

BIOFUN1_BF1A01_249 156

measurement types: Abundance, ObservedIndividualCount, Wet Weigh

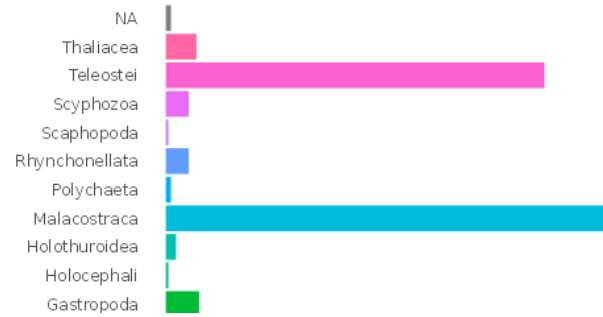
BIOFUN1_BF1A15_390 1

measurement types: Abundance, ,

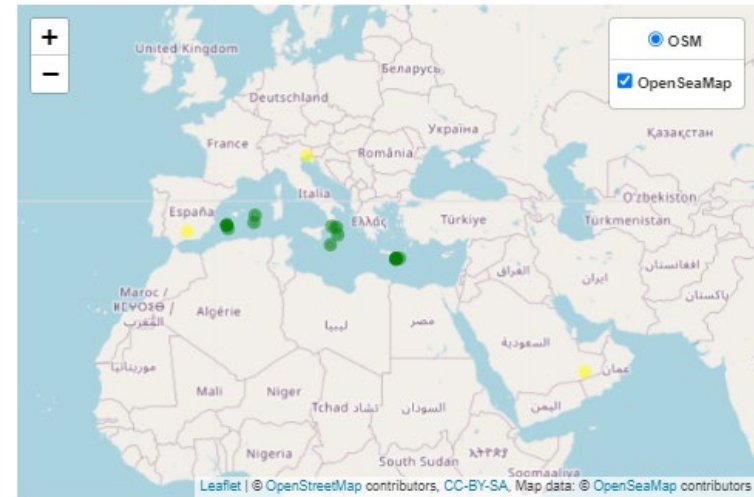
Taxonomic cover of the dataset

Select kingdom

Animalia



Geographical cover of the dataset



The Biocheck tool

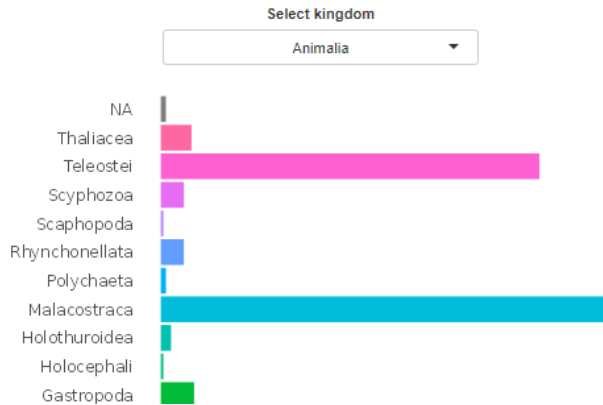
Quality controlling biodiversity data



Data Overview

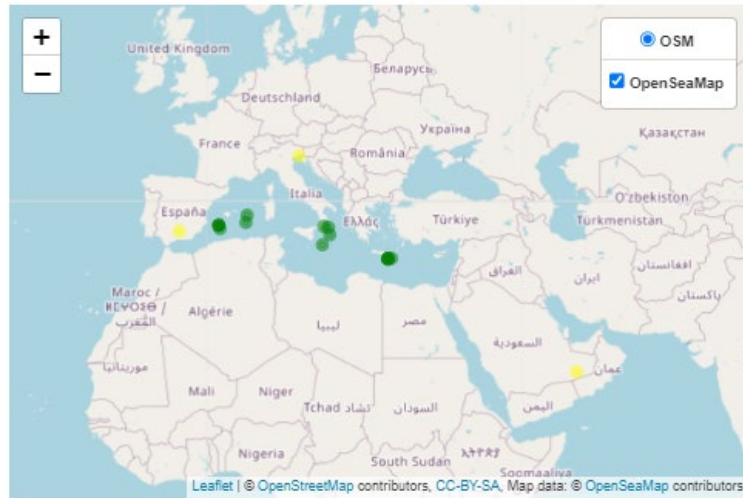
BIOFUN1 1
 event type: Cruise
 BF1A01 14
 event type: Sample
 measurement types: Gear, sampling net horizontal opening, sampling net mesh :
 BIOFUN1_BF1A01_249 156
 measurement types: Abundance, ObservedIndividualCount, Wet Weigh
 BIOFUN1_BF1A15_390 1
 measurement types: Abundance, ,

Taxonomic cover of the dataset

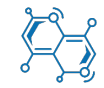


Quality Checks

Geographical cover of the dataset



- Format and integrity
- Taxonomy
- Biogeography
- Temporal coverage
- Sampling methodology
- Biotic quantifications and facts
- Biometrics data
- Environmental parameters



The Biocheck tool

Usage specs

- ***Scope***
- ***Expected data format***
- ***Usage platform***
- ***Usage input data***



- Marine biodiversity occurrence datasets
- Darwin Core (DwC) OBIS
 - Event
 - Occurrence
 - Extended measurements of facts
- [Web app](#)
- [EMODnetBiocheck R package](#)
- IPT resource URL
- DwC-archive file
- DwC tables

The Biocheck tool

Usage specs



Dataset input box

LifeWatch & EMODnet Biology QC tool

Data overview

Issues found

Issues on map

Invalid Event Records

Invalid Occurrence Records

Invalid eMoF Records

OBIS Event Hierarchy tree

About

Welcome to the LifeWatch and EMODnet Biology data quality control tool. This tool allows you to assess to what extent a dataset published on an IPT or a DwC-A file meets the EMODnet Biology Data Quality Criteria. Provide the URL to the main page of the IPT resource in the field 'Link to IPT resource' or upload a [zipped DwC archive](#) under File: 'Upload DwC archive', and click on load. This may take a few minutes for large (+100k occurrence records) datasets. Then navigate through the different tabs above to get detailed QC information, or press the download button to resolve the issues later. The About tab lists which QC checks are carried out, and how the tables and error messages must be interpreted.

To share a report you can use 'http://rshiny.lifewatch.be/BioCheck/?baseurl=' + <URL to your IPT> + '&resource=' + <shortname of the resource>. So for example http://rshiny.lifewatch.be/BioCheck/?baseurl=http://ipt.vliz.be/training&resource=biofun_2009

URL File

Link to IPT resource

Add event hierarchy tree?



This service is powered by LifeWatch Belgium [Learn more»](#)

The Biocheck tool

Usage specs



Data overview tabs

LifeWatch & EMODnet Biology QC tool

Data overview

Issues found

Issues on map

Invalid Event Records

Invalid Occurrence Records

Invalid eMoF Records

OBIS Event Hierarchy tree

About

URL

File

Link to IPT resource

Add event hierarchy tree?

no

Welcome to the LifeWatch and EMODnet Biology data quality control tool. This tool allows you to assess to what extent a dataset published on an IPT or a DwC-A file meets the EMODnet Biology Data Quality Criteria. Provide the URL to the main page of the IPT resource in the field 'Link to IPT resource' or upload a [zipped DwC archive](#) under File: 'Upload DwC archive', and click on load. This may take a few minutes for large (+100k occurrence records) datasets. Then navigate through the different tabs above to get detailed QC information, or press the download button to resolve the issues later. The About tab lists which QC checks are carried out, and how the tables and error messages must be interpreted.

To share a report you can use 'http://rshiny.lifewatch.be/BioCheck/?baseurl=' + <URL to your IPT> + '&resource=' + <shortname of the resource>. So for example http://rshiny.lifewatch.be/BioCheck/?baseurl=http://ipt.vliz.be/training&resource=biofun_2009



This service is powered by LifeWatch Belgium [Learn more»](#)

The Biocheck tool

Usage specs

Quality checks tabs

LifeWatch & EMODnet Biology QC tool

Data overview

Issues found

Issues on map

Invalid Event Records

Invalid Occurrence Records

Invalid eMoF Records

OBIS Event Hierarchy tree

About

URL

File

Link to IPT resource

Add event hierarchy tree?

Welcome to the LifeWatch and EMODnet Biology data quality control tool. This tool allows you to assess to what extent a dataset published on an IPT or a DwC-A file meets the EMODnet Biology Data Quality Criteria. Provide the URL to the main page of the IPT resource in the field 'Link to IPT resource' or upload a [zipped DwC archive](#) under File: 'Upload DwC archive', and click on load. This may take a few minutes for large (+100k occurrence records) datasets. Then navigate through the different tabs above to get detailed QC information, or press the download button to resolve the issues later. The About tab lists which QC checks are carried out, and how the tables and error messages must be interpreted.

To share a report you can use 'http://rshiny.lifewatch.be/BioCheck/?baseurl=' + <URL to your IPT> + '&resource=' + <shortname of the resource>. So for example http://rshiny.lifewatch.be/BioCheck/?baseurl=http://ipt.vliz.be/training&resource=biofun_2009



This service is powered by LifeWatch Belgium [Learn more»](#)

The Biocheck tool

Usage specs

About tab

LifeWatch & EMODnet Biology QC tool

Data overview Issues found Issues on map Invalid Event Records Invalid Occurrence Records Invalid eMoF Records

OBIS Event Hierarchy tree

About

URL

File

Link to IPT resource

<http://ipt.vliz.be/training/a>

Add event hierarchy tree?

no

Load

This tool is based on the EMODnetBioCheck R package (<https://github.com/EMODnet/EMODnetBiocheck>) that was created using the obistools package (<https://github.com/iobis/obistools>) and it is available from the LifeWatch services at <http://rshiny.lifewatch.be/BioCheck/>.

- [BioCheck tool Video TUTORIAL](#) ✓

The Biocheck tool performs a detailed Quality Control on [OBIS-env datasets](#) and [occurrence core dataset](#). It provides a way for a visual exploration of the dataset and highlights potential issues running the following checks:

- [Integrity Checks](#)
- [Format Checks](#) ✓
- [Visual Checks](#)



EMODnet

European Marine
Observation and
Data Network

- <https://emodnet.ec.europa.eu/en/biology>
- bio@emodnet.eu



EMODnet



European Marine
Observation and
Data Network

emodnet.ec.europa.eu

Stay up-to-date with
the latest news



27-29 May 2024 



imdis

International conference on Marine Data and Information Systems



MARIS

MEDIN



National
Oceanography
Centre



iImagine



eosc
Blue-Cloud2026