EMODnet

European Marine Observation and Data Network

EMODnet Thematic Lot n° V- Biology

CINEA/EMFAF/2022/3.5.2/SI2.895681

Start date of the project: 10/05/2023 (24 months)

Operational Phase

Informative material based on the 4.3.1 outcomes (Workshop for EU Mission Ocean Projects) [D4.1.4]

The European Marine Observation and Data Network (EMODnet) is financed by the European Union under Regulation (EU) No 508/2014 of the European Parliament and of the Council of 15 May 2014 on the European Maritime and Fisheries Fund.

Disclaimer

The information and views set out in this report are those of the author(s) and do not necessarily reflect the official opinion of the CINEA or of the European Commission. Neither the CINEA, nor the European Commission, guarantee the accuracy of the data included in this study. Neither the CINEA, the European Commission nor any person acting on the CINEA's or on the European Commission's behalf may be held responsible for the use which may be made of the information.

Document info

Title [ref]*	4.1.4 Informative material based on the 4.3.1 outcomes (Workshop for EU Mission Ocean Projects)
WP title [ref]*	WP4 Uptake and Outreach
Task [ref]*	Task 5 Contributing static content to dedicated space in the Central Portal
Authors [affiliation]	M. Lipizer, N. Holodkov (OGS), Gert Van Hoey (ILVO), J. Beja (VLIZ)
Dissemination level	Public
Submission date	01/04/2025
Deliverable due date	09/05/2025

Contents

1. Introduction

The recent requirement for EU Mission Ocean funded projects to ensure their data/products are available via EMODnet has identified a need for EMODnet to provide clear information on the data submission pathways and support available.
In late February 2024, EMODnet Ingestion organised the first webinar for these projects and was "designed to inform EU project data management plans, and to optimise the pipeline of marine knowledge from data collection and curation to submission into the EMODnet service". Following this webinar and given the number of requests arriving to the EMODnet Biology Data Management Team, it was decided that a workshop focused on EMODnet Biology's work would help answer some of the questions and create resources that could be used by the projects.
The workshop was aimed at representatives (ideally those responsible for the data management) of recent and ongoing EU Mission Ocean research projects that work with marine biodiversity data. Its objective was to share EMODnet's expertise and experience in marine biodiversity data management to facilitate the publication of data and products from EU-funded projects. It aimed to provide detailed insights into EMODnet Biology's capabilities, demonstrate how EMODnet Biology can support EU projects, and offer specific guidance on publishing projects biodiversity data and products.
Before the workshop a questionnaire designed to collect information on data collected by the projects, on data management practices and on major challenges in terms of data management and sharing was distributed to 59 representatives of EU projects and results were used to adapt the workshop agenda to the participants' needs.
In addition, in order to share information on types of biological data collected and on data availability and data management procedures adopted by recent EU projects, 10 representatives of EU projects were invited to present short pitches (see list of project pitches) and an online poll through Mentimeter was used during the workshop, to collect additional information from the participants.

The workshop was attended by 56 participants, which included 44 representatives from EU projects and from the EU Commission and 13 EMODnet project participants.

The presentations are available in the annexes and the recording of the event are available on the EMODnet YouTube channel:

- Day 1- https://www.youtube.com/watch?v=K77ztSKBSC4
- Day 2- https://www.youtube.com/watch?v=TfYPEtUkk2Y

2. Pre-workshop survey results

The section includes a description of all questions asked in the survey sent to the various EU Mission Ocean funded projects and the results obtained.

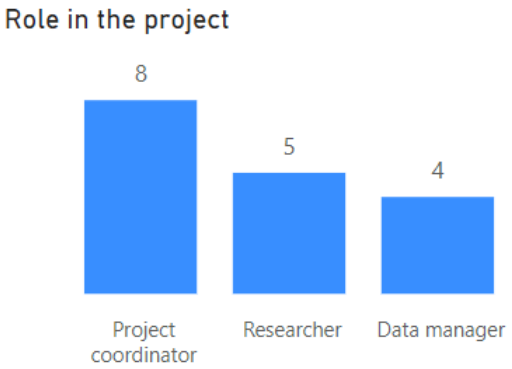
Project Information

Number of projects

16

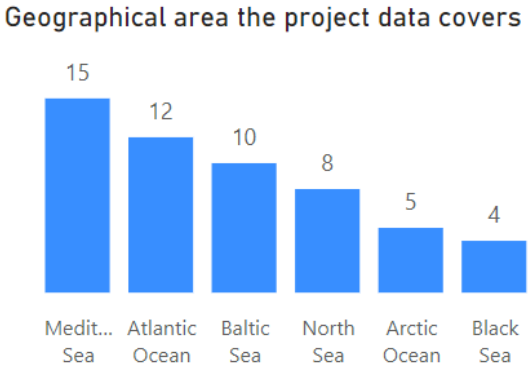
Number of answers

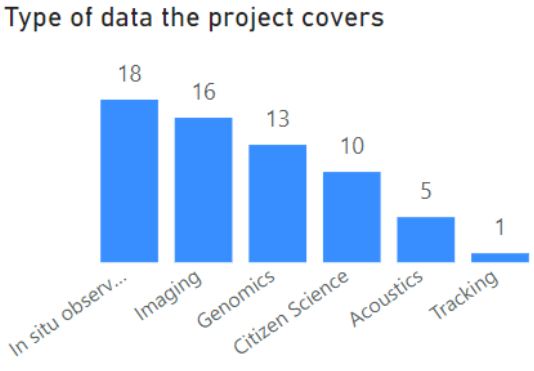
21

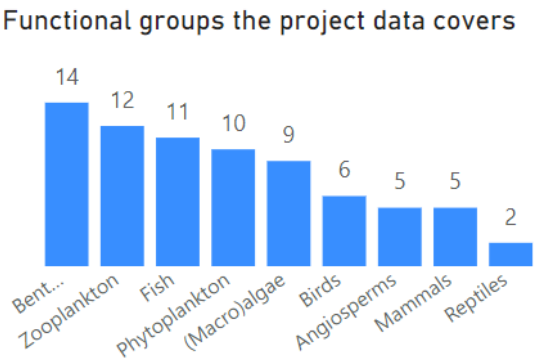


Research and Data Overview

All European marine regions are covered by the projects, with highest participation from the Mediterranean and Atlantic Ocean regions. Most of the data consist of *in-situ* observations, many projects cover more recent data types, such as imaging and genomic data. Data stemming from citizen science activities is also common in the feedback received from the projects. The project's data are covering all functional groups, with benthic (fauna and algae), pelagic (zooplankton and phytoplankton) and fish being the most common groups.



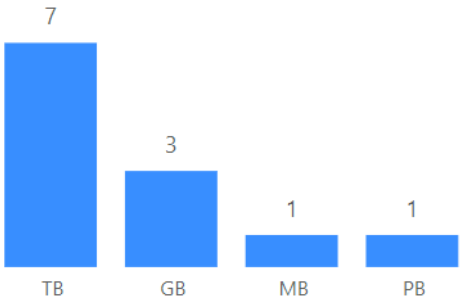




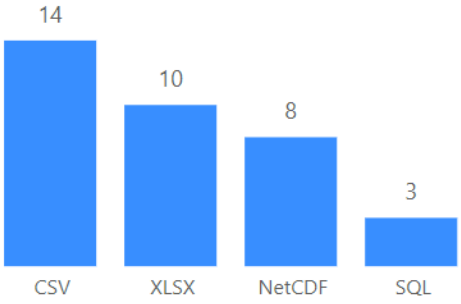
Data size can be quite high for some types of data (terabytes) and data formats are heterogenous, with CSV being the most frequently used.

Data Management Practices

Size of the data the project covers



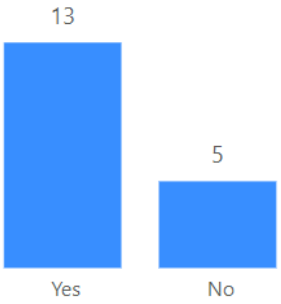
Format of the data the project covers



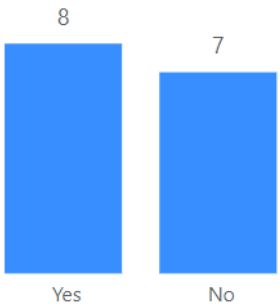
Data Collection and Storage Practices

Most projects have a data management plan, but almost half of the respondents do not use data standards or existing practices.

The project has a Data Management Plan?



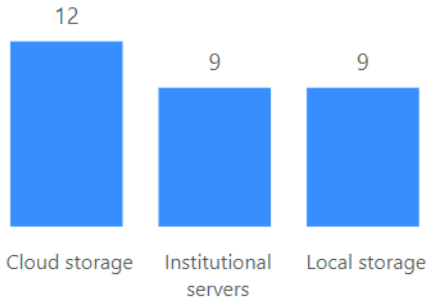
The project uses data standards/practices?



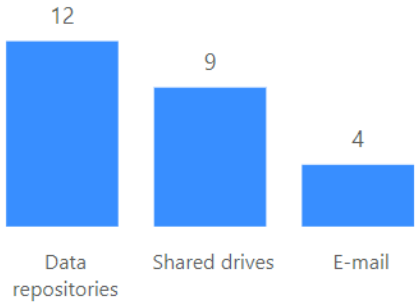
Data Sharing and Accessibility

Data sharing is in many cases still on a local scale (local or institutional servers) instead of cloud storage. Nevertheless, for almost 50% of the projects the data is shared to data repository systems or shared drives.

Storing of the project data

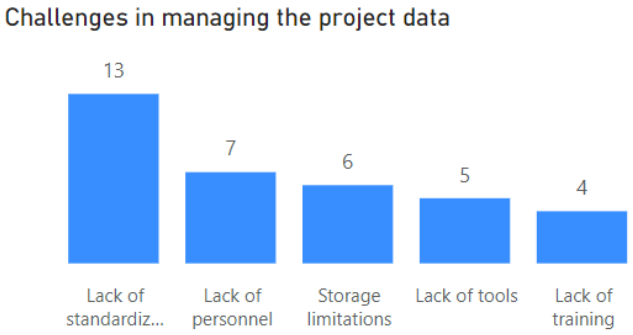


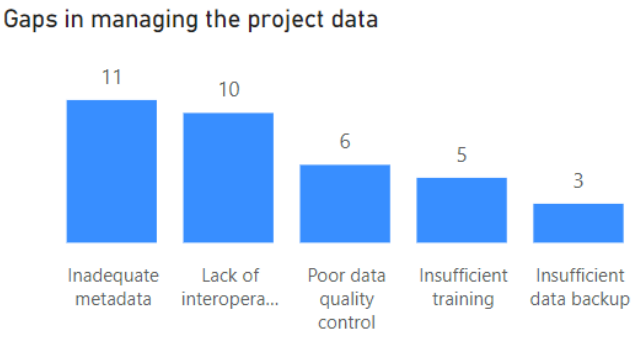
Sharing of the project data



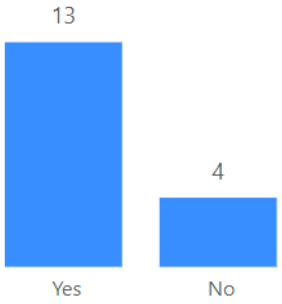
Challenges and Gaps

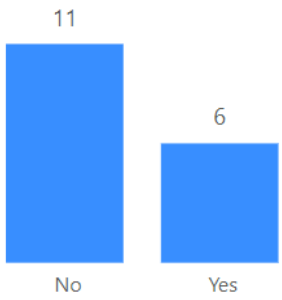
The main challenge identified by the project is the lack of standardized data practices. Lack of resources (personnel and storage) was also a common identified challenge. The main gaps were related to inadequate metadata documentation and lack of interoperability.





Awareness and Use of the EMODnet Data Portal

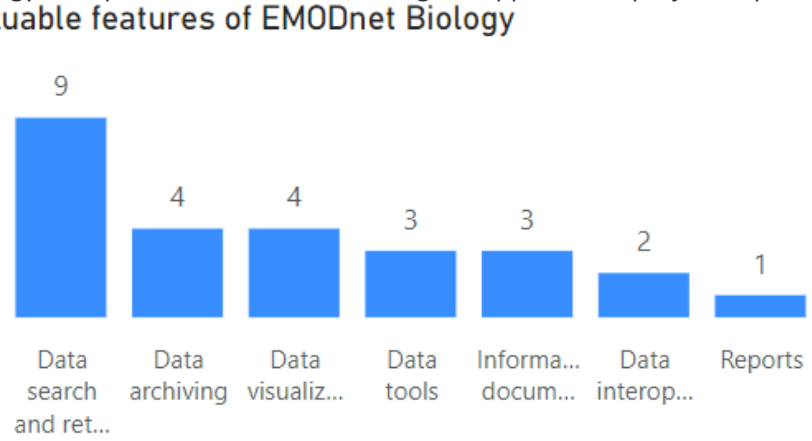




The projects are mostly aware of EMODnet Biology, but most of them have never used data or products published by the thematic lot.

What features or services you find most valuable in EMODnet Biology

According to the results, the EU projects mostly use EMODnet Biology as a data retrieval aggregator. A few indicate it as a system for data archiving and visualization. In particular, the project GES4SEA mentioned that some EMODnet Biology data products can be interesting to support their project implementation.



Suggestions to encourage the use EMODnet Biology:

- Integration with DTO: EMODnet Biology seems to be the mostly widely known repository for biodiversity data (CLIMAREST);
- Availability of complete, harmonized, historical data series until present day from the whole Baltic Sea area (ProtectBaltic);
- Improvement of ease for downloading data (OBAMA-NEXT);
- Improve availability of genomics and microbiome data (BlueRemediomics)
- Availability of freely available abundance data (GES4SEA)
- Development of a literature tracking service, similar to the one developed in GBIF (https://www.gbif.org/literature-tracking) (ANERIS)

How can EMODnet Biology better support data management needs?

The list below captures the feedback given by the workshop participants:

- EMODnet should be entitled to reach out to specific consortia to facilitate data mobilization and improve FAIR data availability. The biggest challenge has been to get people to care/engage with data management;
- Enlarge the partnership to include additional data providers;
- Provide practical support at the consortia level;
- Facilitate to overall data management process;
- Add more functionality for genomics data

In summary, some projects would like to have some more support from the EMODnet Biology in taking up the data management task and the flow of data to the portal, making data management as easy as possible.

Topics needed to be covered in the November workshop?

The online participants were asked what they like to learn from the workshop. In other words, we checked for their expectations. The following three aspects were put forward:

- Data standardization to support interoperability for data ingestion
- Better leverage genomics data
- Data provenance services and literature tracking

3. Workshop agenda

Time	Speaker	Title
10:00	Marina Lipizer	Aim and agenda (10’)
10:10	Zoi Konstantinou -	Short introduction by DG MARE – representing also DG Research (Nicolas Segebarth) (5’)
10:20	Joana Beja	1 - EMODnet Biology (15’)
10:35	Nikola Holodkov	2 - Outcomes from questionnaire (10’)
10:45	All	Q&A (10’)
10:55		3 - Project pitches (15’) BlueRemediomics - Samuel Chaffron (CNRS) Biodiversa MOSTFUN - Laura Garzoli (CNR) AtlantECO, BiOcean5D, BlueCloud2026 - Meike Vogt (ETHZ)
11:10		Coffee break
11:30	Ruben Perez	EMODnet Biology data guidance (35’) - Data management 101: FAIR, licenses, storage, provenance, standards, vocabularies
12:05		Project pitches (15’) Contrast - Steven Brooks (NIVA) B-USEFUL - Julia Polo (UIT) MARHAB - Even Moland (IMR)
12:20	Ville Karvinen	EMODnet Seabed Habitats (20’)
12:40		Q&A (20’)
13:00		End of day 1
Time	**Speaker**	**Title**
10:00	Marina Lipizer	Welcome, aim for 2nd day (5’)
10:05		Project pitches (15’) GES4SEAS - Angel Borja (AZTI) BMD - Niels Raes (NBC)

		PROTECT BALTIC - Kimmo Koivumaki (HELCOM) DIGI4ECO - M. Clavel-Henry (GEOMAR)
10:20	Ruben Perez/Lynn Delgat	EMODnet Biology data submission guidance (45') - EMODnet Bio data flows - DwC & OBIS-env - omics - IPT - Biocheck
11:05		Q&A (10')
11:15	Salvador Fernandez	EMODnet Biology product guidance (15') Preferred output NetCDF CF Compliance COARDS Compliance
11:30		Coffee break
12:00	Marina Lipizer	EMODnet Biology tools and services (15') - Data preparation (training/documentation resources -> courses, tutorials and manuals) - Access data/metadata - Access products
12:15		Q&A/ Participants feedback
13:00		End of day 2

4. Workshop outcomes

This workshop represented the first EMODnet Biology initiative targeting specifically EU Mission Ocean funded research projects and aimed at sharing information on the thematic lot's tools and services and to collect needs and requirements from the EU projects representatives. The workshop followed the 'EMODnet webinar for European Research and Innovation and Mission: Restore our Ocean and Waters projects' held in late February 2024 which provided generic guidance on data submission via the EMODnet Ingestion facility.

As highlighted by the participants, marine biodiversity data are heterogeneous, spanning from traditional microscopy determinations to imaging, genomics and acoustics, which represents a growing challenge for data management and curation.

The workshop provided a comprehensive overview of what EMODnet Biology offers in terms of data management **guidance**, data **resources**, data product **guidance**, **tools** and **services** to facilitate data and products publication according to the **FAIR** (findable, accessible, interoperable, re-usable) principles. The agenda items aimed in particular in providing information on: training resources, data quality control tools, data and metadata standardization to improve interoperability, to address the challenges and gaps identified in the pre-workshop questionnaire.

From the point of view of EU Mission Ocean attending projects, the discussion highlighted, on one side, the difficulty in addressing data management due to the specific knowledge required, the limited availability of dedicated funds for data curation, the high heterogeneity of "biological data" and on the other side the importance of data quality control and metadata availability that is able to provide information on the overall data quality level, indicating whether a peer review was carried out, or if data was collected from citizen science or monitoring efforts.

The importance, yet still overlooked, of data curation has been underlined.

According to the online poll, the major types of marine biodiversity data produced within projects represented at the workshop originate from: observations, genomic and imaging (Fig. 2, top) and data sharing faces several **challenges** such as: data restrictions, lack of human or technical resources for data management, difficulties in data formatting according to standards (e.g Darwin Core) and in using vocabularies (Fig. 2, bottom).

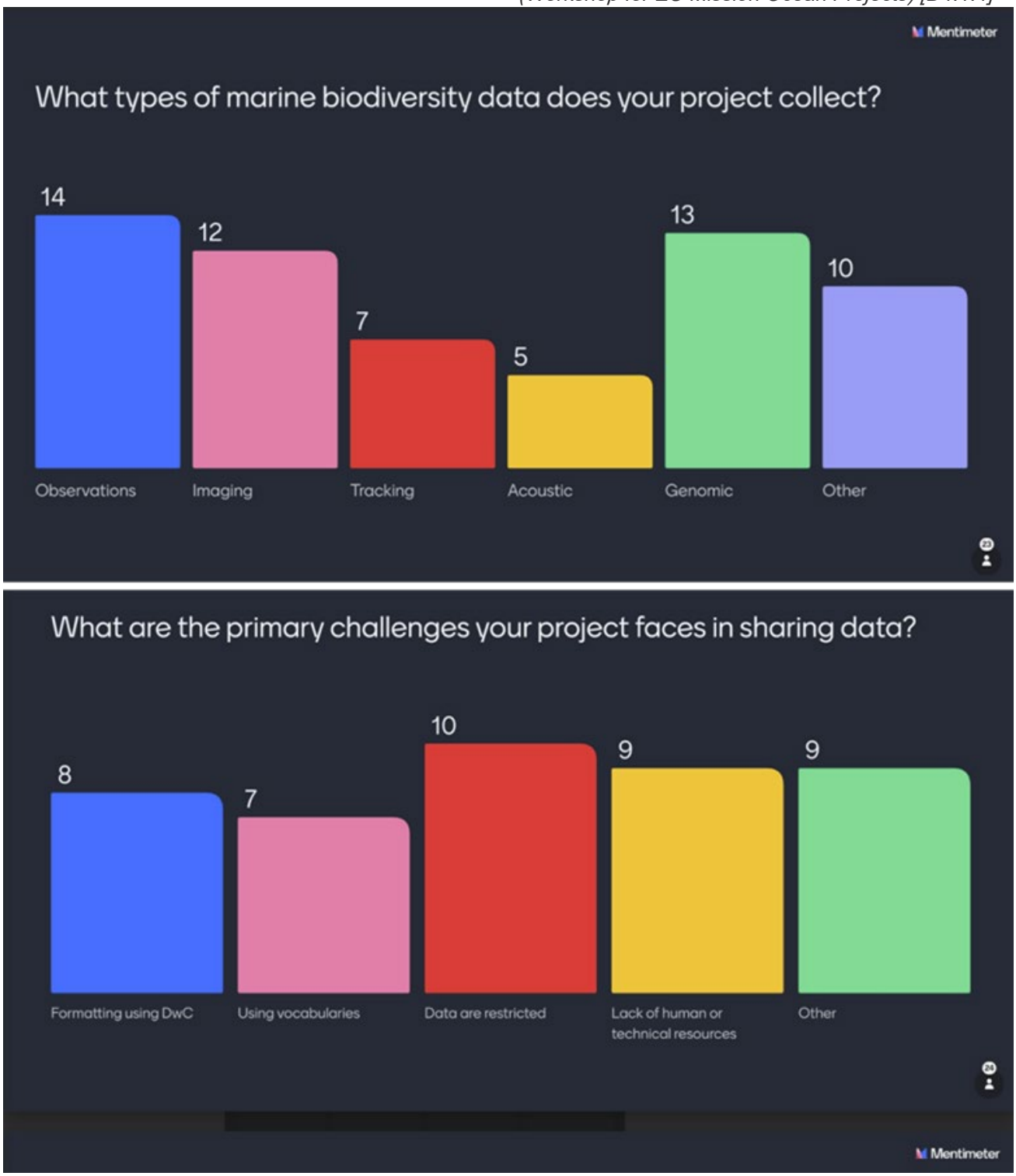


Fig. 2: Results from the online poll collected during the workshop. Types of marine biodiversity data (top) and main challenges in sharing data (bottom).

The main tools, guidelines, tutorials and online courses made available by EMODnet Biology (see presentations in the Annex) were presented to participants and are accessible through:

- https://emodnet.ec.europa.eu/en/tools-guidelines#biotools

- https://emodnet.ec.europa.eu/en/biology#biology-key-services
- classroom.oceanteacher.org/enrol/index.php?id=958

During the workshop several additional information and online resources were provided by the workshop participants and are added below:

- Tutorials on how to work with CF-NetCDF files:
  - in R: https://nordatanet.github.io/NetCDF_in_R_from_beginner_to_pro
  - in Python: https://nordatanet.github.io/NetCDF_in_Python_from_beginner_to_pro
  - https://www.youtube.com/@LukeDataManager
- Support for metabarcoding Data Programme:
  - https://www.gbif.org/metabarcoding
- Resources based on Artificial Intelligence to support dataset standardization:
  - https://www.gbif.org/news/6aw2VFiEHYlqb48w86uKSf/chatipt-system-wins-the-2024-ebbe-nielsen-challenge

5. Conclusion and way forward

Feedback collected after the event confirmed the importance of stronger interactions between the research community collecting data and the major data infrastructures, such as EMODnet Biology, in order to share information on data management systems, tools and services.

Among the major challenges to be addressed: data curation is a fundamental yet overlooked phase of the scientific process and it should be more clearly accounted for with dedicated resources and expertise. Data curation is currently in place in the case of big data creators/publishers such as Ecotaxa, or big consortia such as Tara Oceans but is not sufficiently tackled by most projects.

This first EMODnet Biology initiative dedicated specifically to EU Mission Ocean research projects allowed to collect needs and requirements in the area of data management and curation. Feedback collected after the event indicated the need for both: general overview of EMODnet Biology tools and services as well as specific data management training, dedicated in particular to early career and researcher-dominated audiences.

The next steps to improve data FAIRness should address three major aspects:

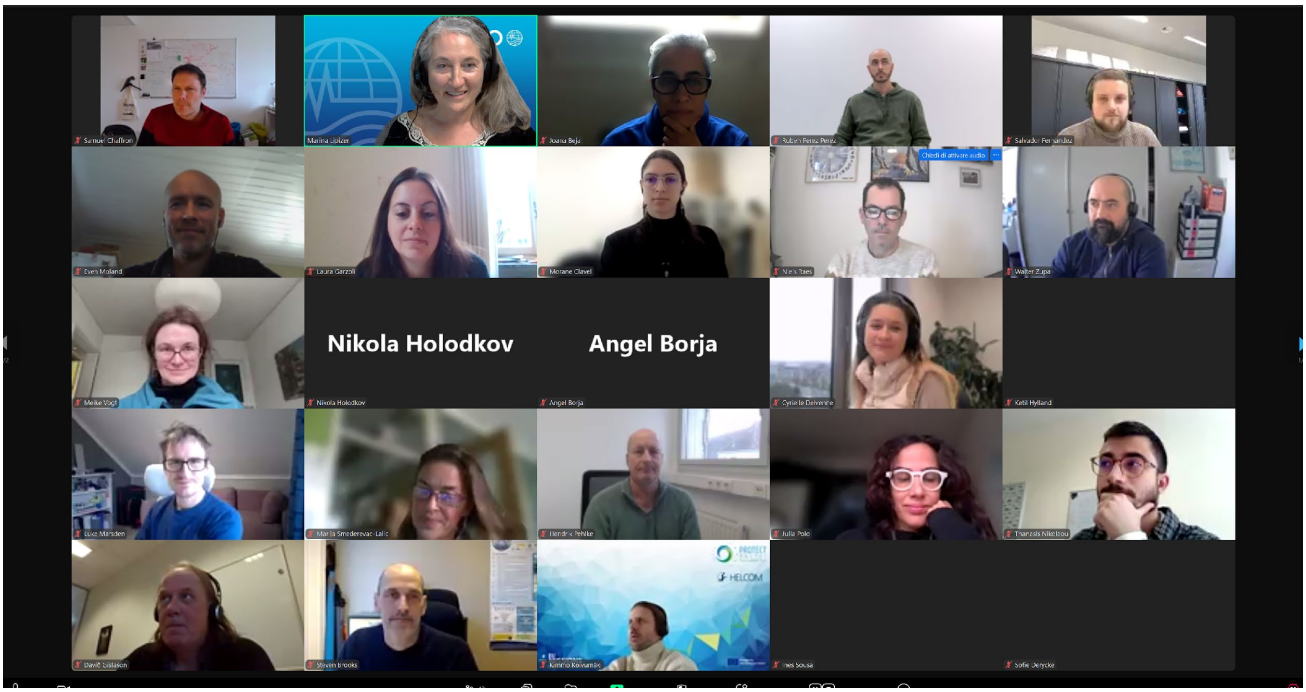
1. a governance aspect which involves dedicating specific funding and capacity building initiatives to support data management and curation;
2. capacity building initiatives to promote the organization of specific hands-on data curation;
3. ICT development to support data management, exploiting the use of Artificial Intelligence to facilitate data management and Quality Control.

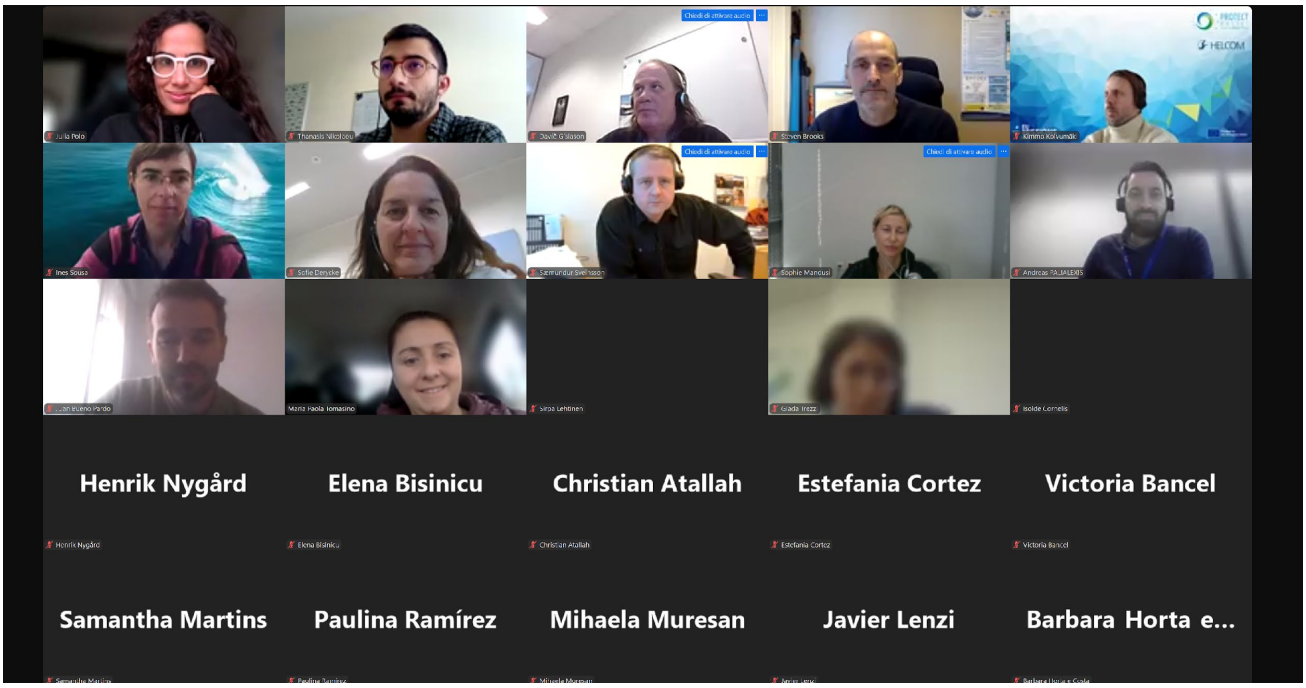
Additionally, EMODnet Biology is committed to providing support to these projects by organising further workshops (online or in-person) and data management support via the data management team at VLIZ. Conversions with the European GBIF nodes will also take place and if possible a collaboration between EMODnet Biology and these nodes will be established, to ensure that both initiatives can provide the adequate data management support for the EU Mission Ocean funded projects.

6. List of project pitches

- BlueCloud2026, AtlantECO, BiOcean5D
- BlueRemediomics
- Biodiversa + MoSTFun
- MARHAB
- B-USEFUL
- CONTRAST
- GES4SEAS
- DIG4ECO
- PROTECT BALTIC
- Biodiversity Meets Data

7. Photos of the event





8. Workshop participants

Organization	Project
SZN	DIGIECO
EMBL-EBI	MGnify; Blue-Cloud 2026
Institut de la Mer Villefranche	
AZTI	GES4SEAS
NIVA	CONTRAST
Universidade de Vigo	ACTNOW
CNRS	BlueRemediomics
GEOMAR	
ILVO	OptiFish and MarineBeacon, eDNAqua-Plan
CCB	PROTECT BALTIC
Stockholm University	Baltic Health Index
Istanbul University	
REA_EC	REA
CNR-IRSA	Biodiversa + MoSTFun
Matís	
CCMAR	MARHAB
Institute of Marine Research, Bergen	Contrast
HELCOM	PROTECT BALTIC
European Commission DG MARE	
Jade University of Applied Sciences	
Aristotle University of Thessaloniki, Greece	
Norwegian Meteorological Institute	Arctic Passion
NIVA	CONTRAST
Institute of Marine Research, Norway	MARHAB

Organization	Project
NIRD - GeoEcoMar, Romania	Marco Bolo
University of the Aegean	
CINEA	Project officers for EU funded projects
European Research Executive Agency (REA)	Project advisor for EU funded projects
AWI	CoastCARB
University of Tromso	B-USEFUL
Instituto de Ciencias del Mar (ICM-CSIC)	DIGI4ECO
Naturalis Biodiversity Center	Biodiversity Meets Data
EMBRC ERIC	eDNAqua-Plan
Institute for Multidisciplinary Research, University of Belgrade	Danube4all
Aarhus University	OBAMA-NEXT
Matis	BioProtect
CIIMAR	DTO-Bioflow
TECNOAMBIENTE SLU	ONDEP
ETH Zurich	
Alfred Wegener Institute Bremerhaven	
Fondazione COISPA	B-useful
Syke	EMODnet Biology/EMODnet Seabed Habitats
VLIZ	EMODnet Biology
NIMRD	EMODnet Biology
Marine Biological Association	EMODnet Biology
OGS	EMODnet Biology

9. Annex: presentations

EMODnet
European Marine Observation and Data Network

Biology

EMODnet Biology workshop: supporting marine biodiversity data management

Online, November 19 - 20, 2024. Time: 10.00 - 13.00 CET

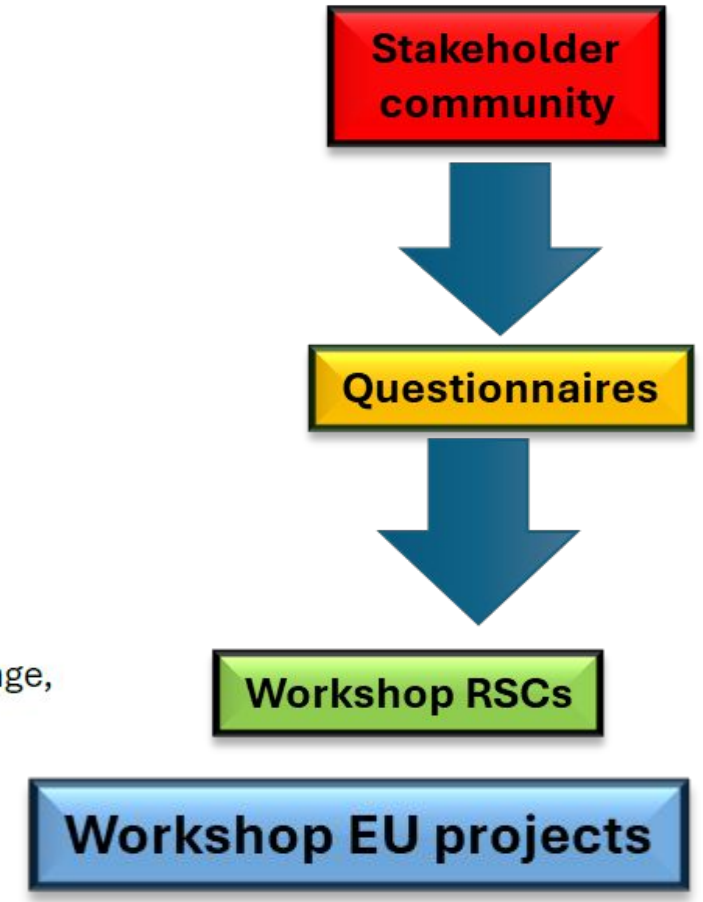
Marina Lipizer OGS/EMODnet Biology WP4 “Uptake and Outreach” leader

19th and 20th November 2024, Online

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.

Day 1:

Time	Speaker	Title
10:00	Marina Lipizer	Aim and agenda (10’)
10:10	Zoi Konstantinou	Short introduction by DG MARE (5’)
10:15	Nicolas Segebarth	Short introduction by DG Research (5’)
10:20	Joana Beja	EMODnet Biology (15’)
10:35	Nikola Holodkov	Outcomes from questionnaire (10’)
10:45	All	Q&A (10’)
10:55		Selected project pitches (15’)
11:10		*Coffee break*
11:30	Ruben Perez	EMODnet Biology data guidance (35’) ○ Data management 101: FAIR, licenses, storage, provenance, standards, vocabularies
12:05		Selected project pitches (15’)
12:20	Ville Karvinen	EMODnet Seabed Habitats (20’)
12:40		Q&A (20’)
13:00		End of day 1



Day 2:

Time	Speaker	Title
10:00	Marina Lipizer	Welcome, aim for 2nd day (5’)
10:05		Selected project pitches (15’)
10:20	Ruben Perez/Lynn Delgat	EMODnet Biology data resources (45’) ○ EMODnet Bio data flows ○ DwC & OBIS-env ○ omics ○ IPT ○ Biocheck
11:05		Q&A (10’)
11:15	Salvador Fernandez	EMODnet Biology product guidance (15’) ○ Preferred output ○ NetCDF ○ CF Compliance ○ COARDS Compliance
11:30		Coffee break
12:00	Marina Lipizer	EMODnet Biology tools and services (15’) ○ Data preparation (training/documentation resources -> courses, tutorials and manuals) ○ Access data/metadata ○ Access products
12:15		Q&A/ Participants feedback
13:00		End of day 2

EMODnet
European Marine Observation and Data Network

Biology

EMODnet Biology overview

Workshop for EU funded projects

Joana Beja, VLIZ/EMODnet Biology coordinator

19th and 20th November 2024, Online

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.

What is EMODnet?

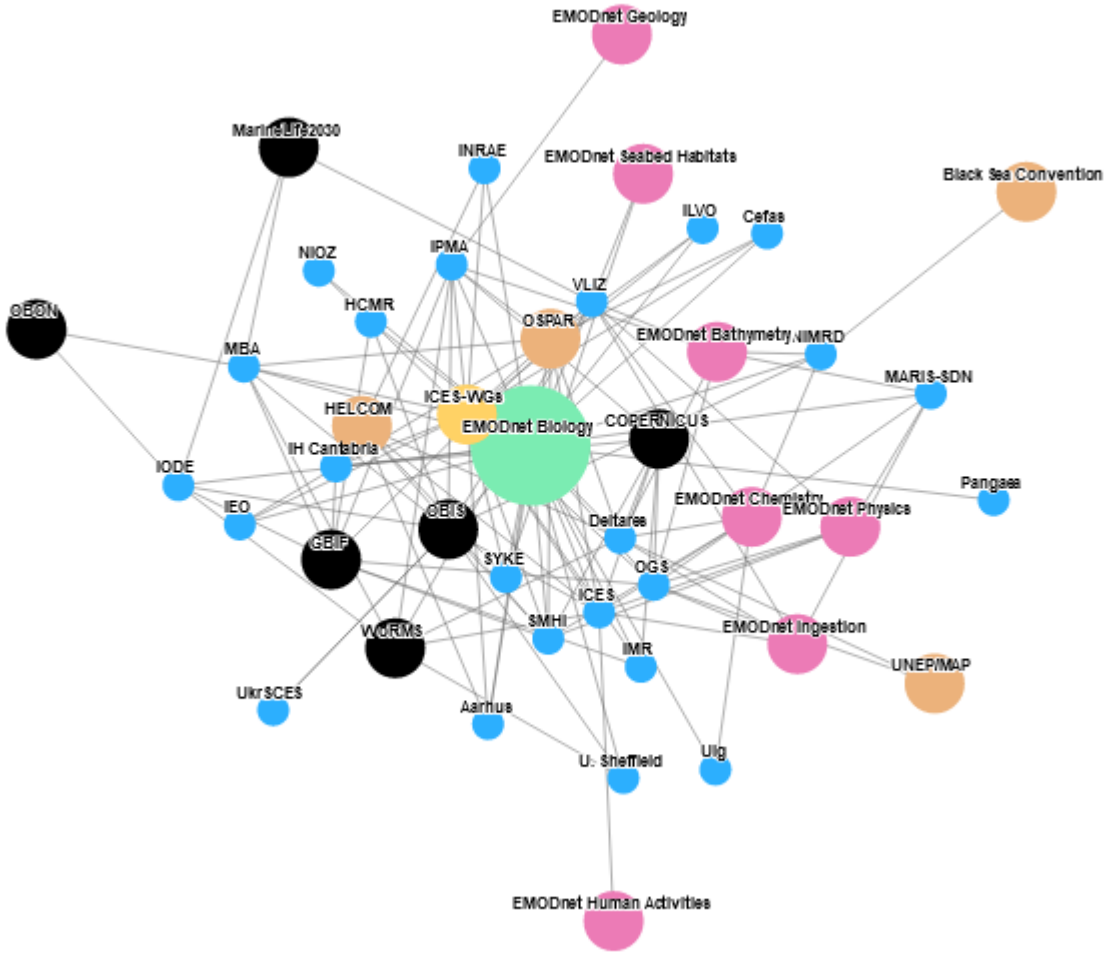
(https://emodnet.ec.europa.eu/en)

- EU service for in situ marine data
- 7 thematic domains: Bathymetry, **Biology**, Chemistry, Geology, Human Activities, Physics, Seabed Habitats (+ Ingestion service)
- Network with more than 120 partner organisations and 30 Associated Partners
- Provides free, unrestricted access to in situ marine data and data products
- Adheres to INSPIRE and FAIR data principles and webservices follow OGC standards

What do we do in EMODnet Biology?

Overview

- Consortium with 24 partners
- Publish marine biodiversity data
- Create biodiversity data products
- Provide training and create guidelines for data providers and product creators
- Publish informative material
- Support the EU member states, EU institutions and Regional Sea Conventions with their work



What do we do in EMODnet Biology?

Data Publication

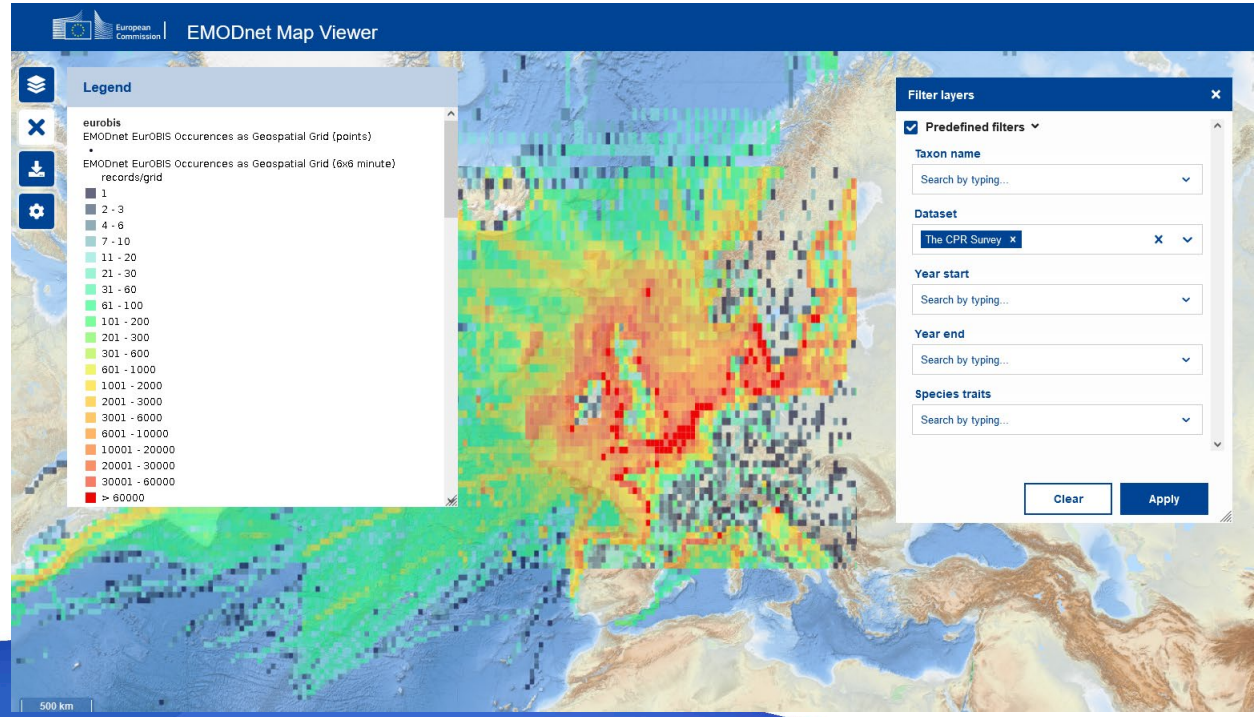
- Focus on European Seas but we accept data from anywhere in the world's oceans
- Target primarily data from 9 sub-themes

- Data types: Observations, acoustic, tracking, imaging, ... and **genomics**

What do we do in EMODnet Biology?

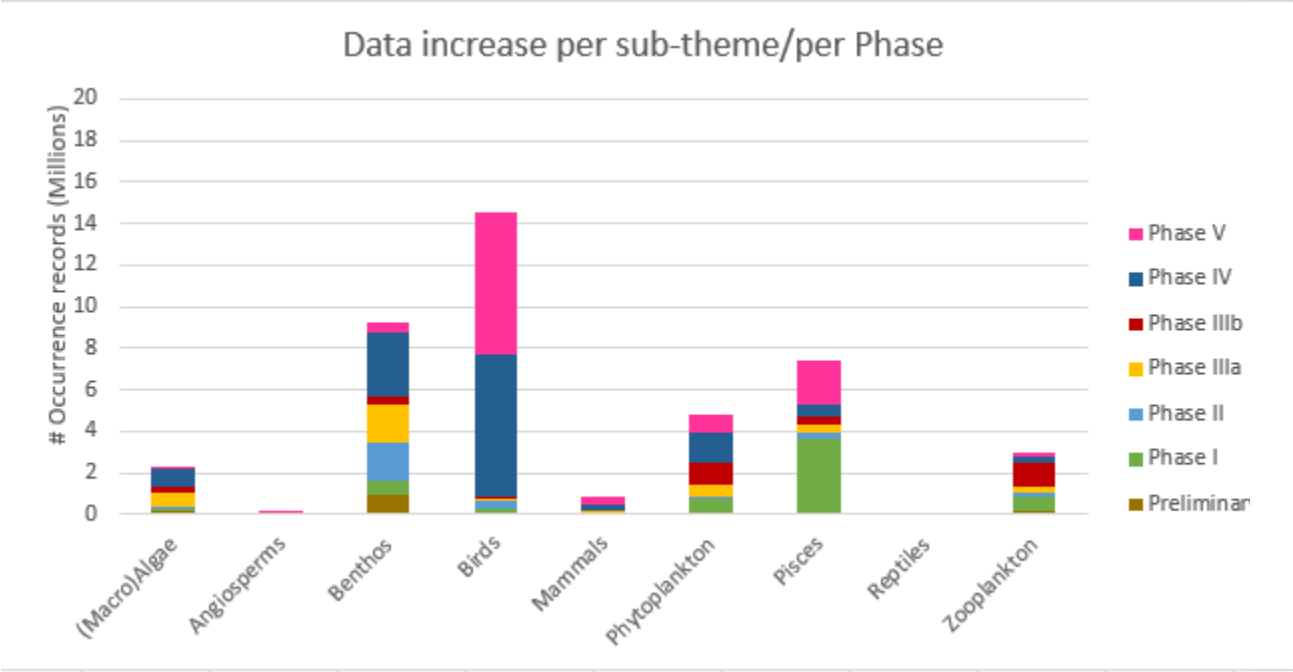
Data Publication

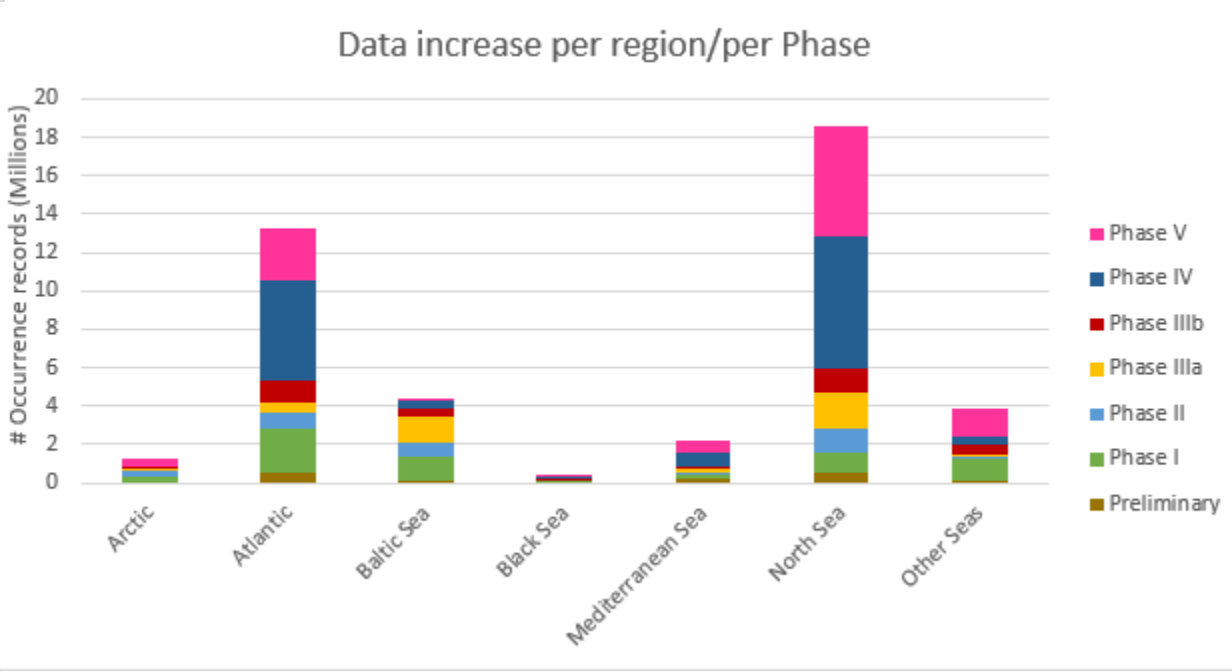
- 1413 datasets with data URLs
- > 41.8M occurrence records + ~ 104M extended Measurements or Facts
- Layer-> Biodiversity records



EMODnet Biology Evolution

Data Publication

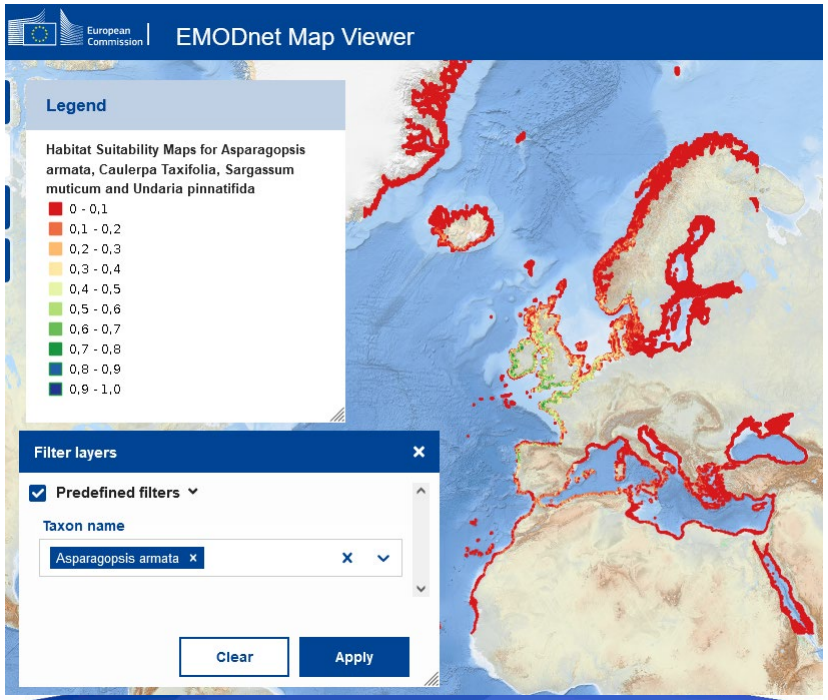


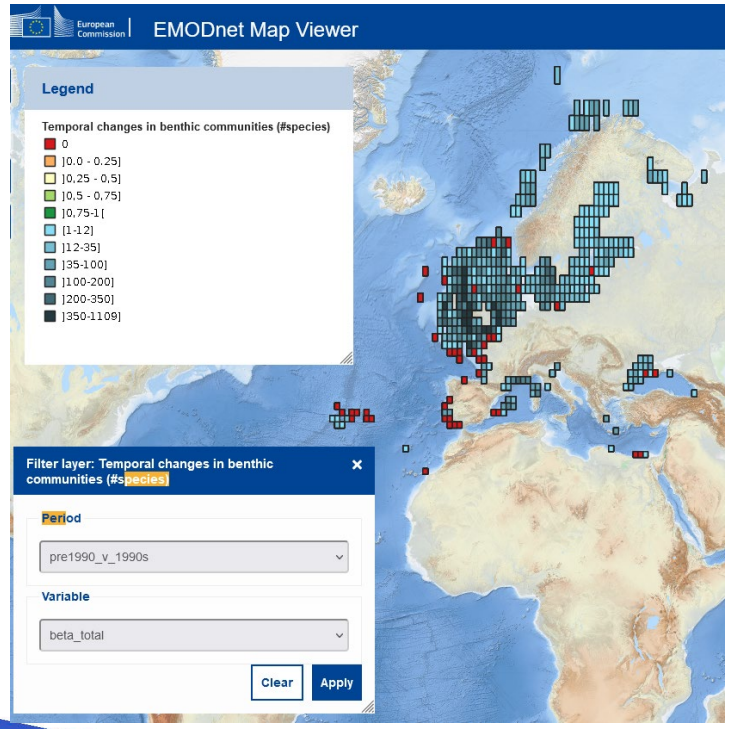


What do we do in EMODnet Biology?

Data Products

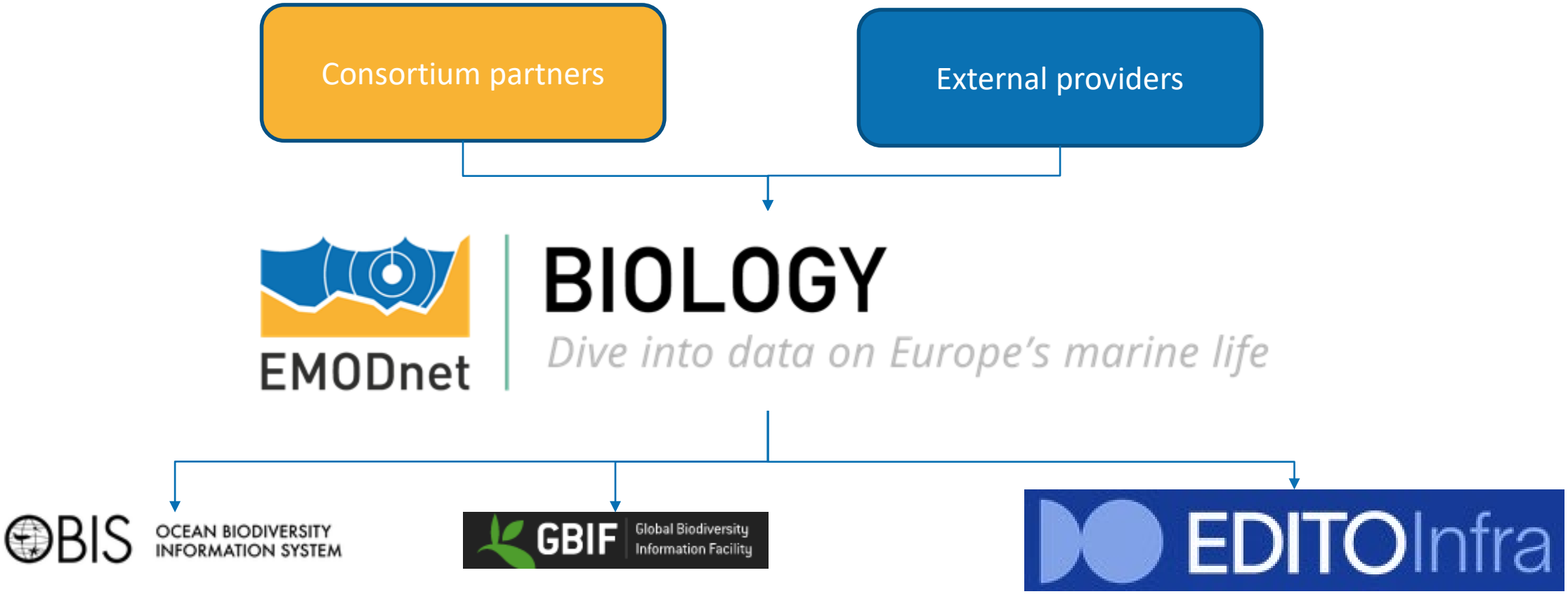
- 38 products (including 3 from external sources)
  - https://emodnet.ec.europa.eu/geonetwork/srv/eng/catalog.search#/home
  - Layers available in the EMODnet viewer: 3 Benthos+ 5 Fish+ 1 Macroalgae + 2 Mammals + 3 Phytoplankton + 4 Zooplankton + 1 other taxa
  - Consortium developed products are documented in GitHub repositories (https://github.com/EMODnet)





Data flows

Data Submission and Publishing



What EMODnet Biology will not do for you

- Follow up on your project's Data Management Plans
- Reformat your data to comply with DwC
- Organise and provide training specifically for your project/data providers
- Reformat your products for publication in the EMODnet viewer

What EMODnet Biology will do

- (Re)publish EU funded project's output (map viewer, catalogue, webservices)
- Publish data (compliant with DwC and QC-ed) that are made available to us
- Ensure your data are shared with other European and international initiatives (OBIS, GBIF, EU DTO)
- Assess which products are relevant for our users and make those visible through the EMODnet viewer
- Publish and maintain resources to be used by the projects (for data and products)
- Continue developing tools and services for data providers and users
- Host and provide technical support for your IPT instance

Biology

EMODnet Biology Workshop - Survey Results

Workshop for EU funded projects

Nikola Holodkov OGS/EMODnet Biology WP4

19th and 20th November 2024, Online

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.

Projects participating

Number of answers

21

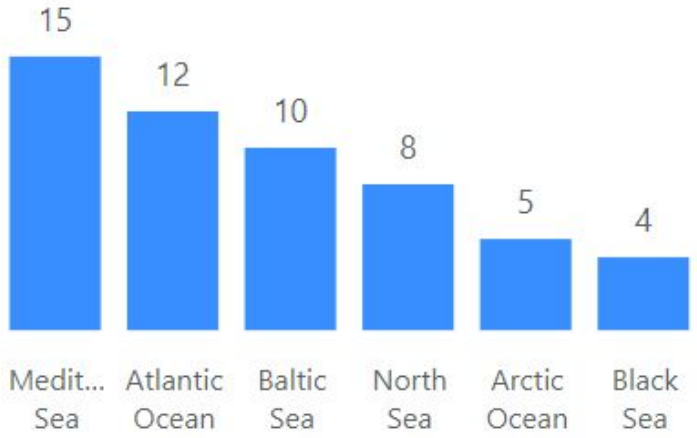
Number of projects

16

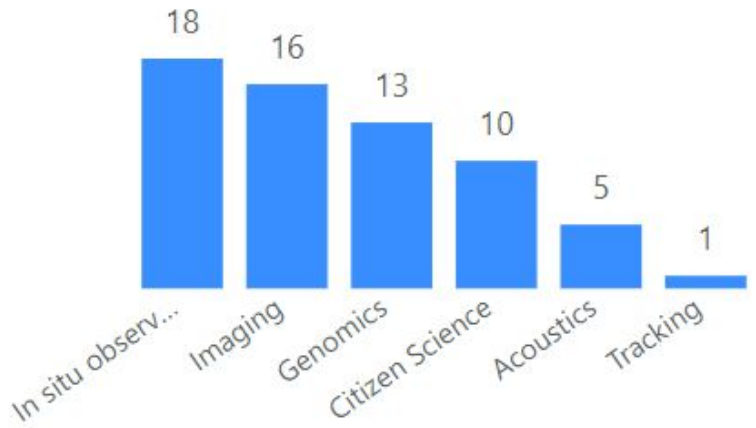
Role in the project

Research and Data Overview

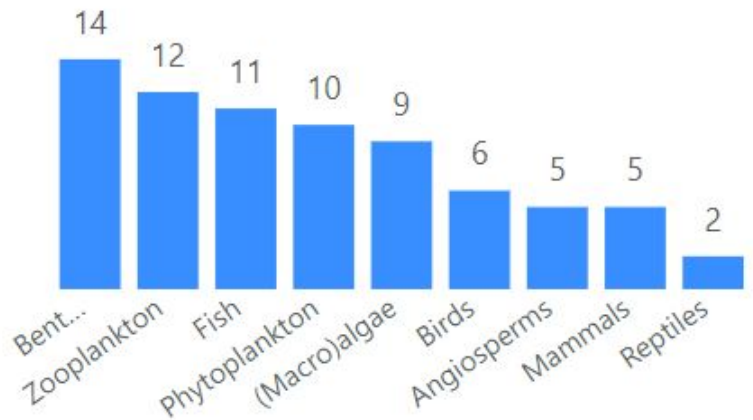
Geographical area the project data covers



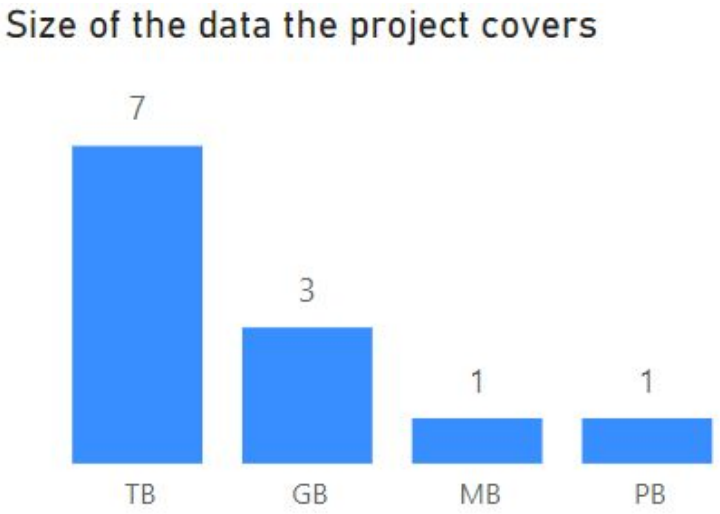
Type of data the project covers

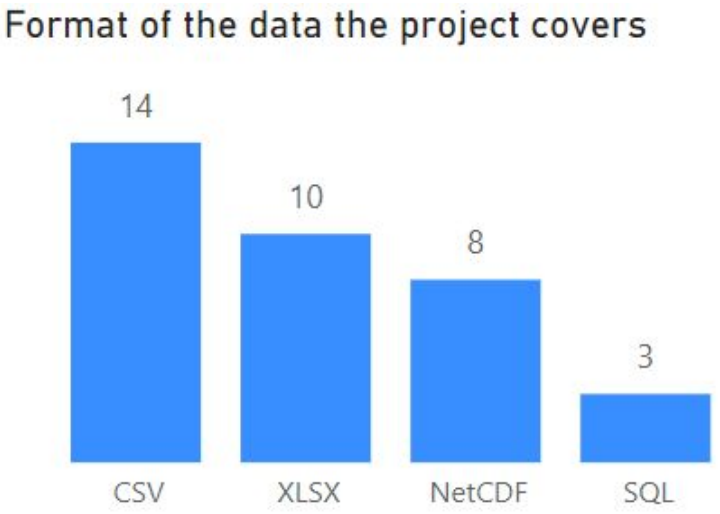


Functional groups the project data covers

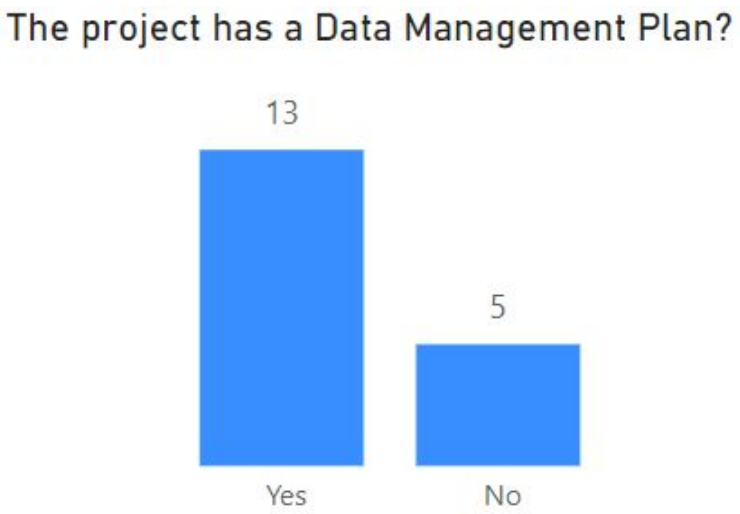


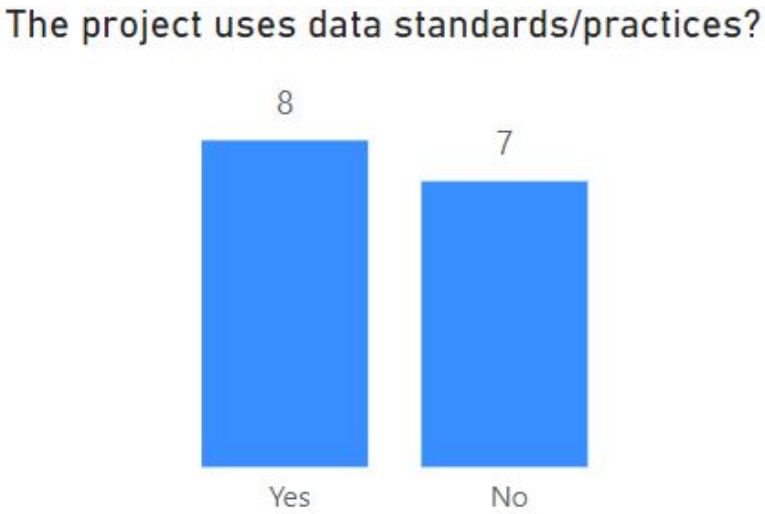
Data Collection and Storage Practices



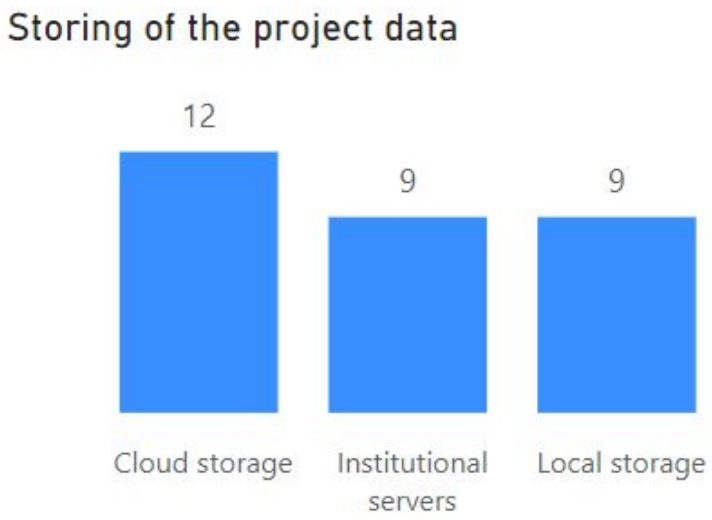


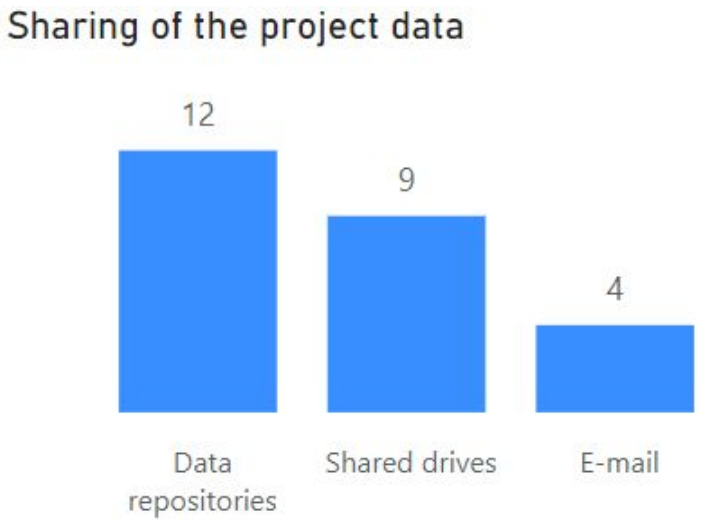
Data Management Practices



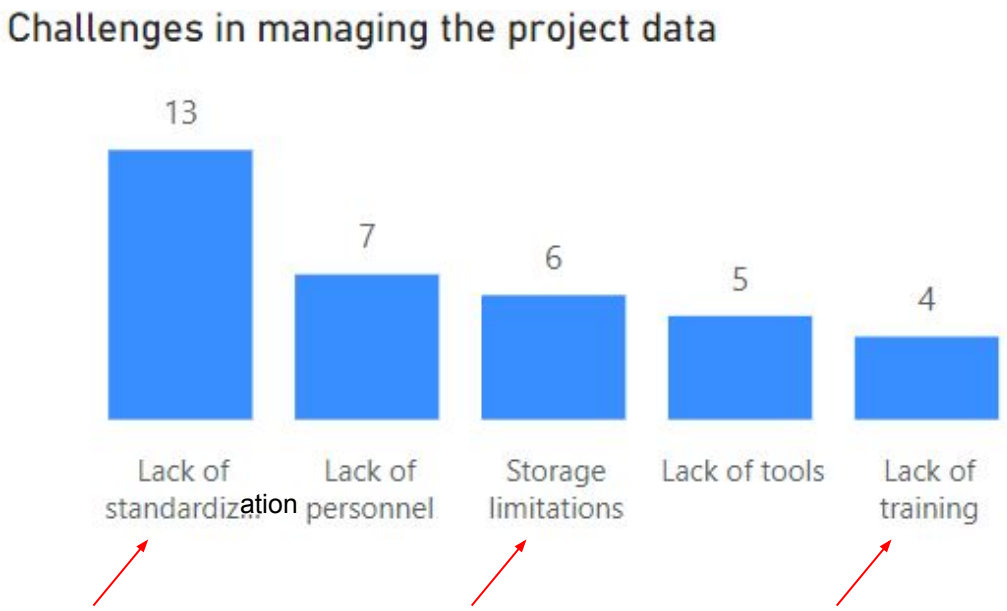


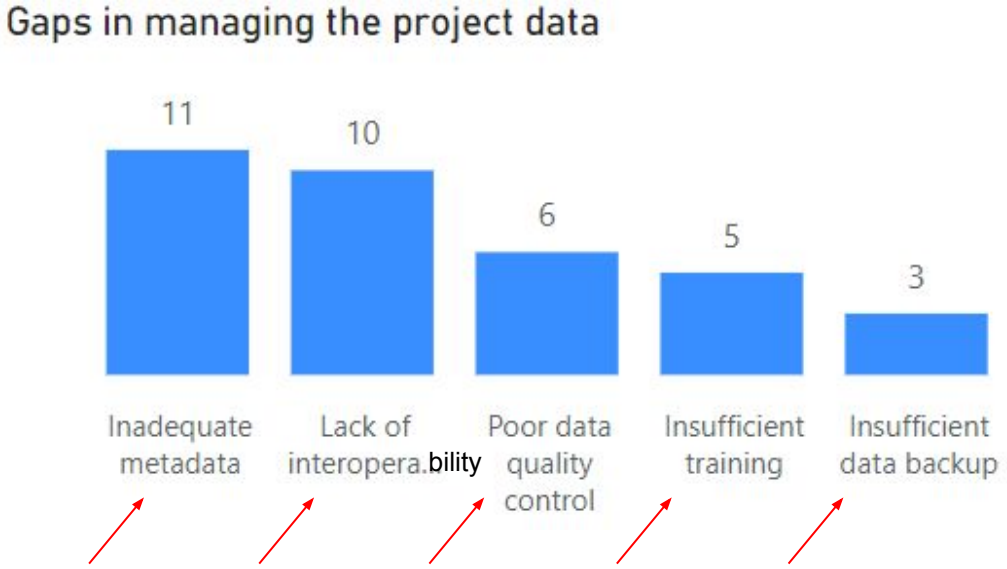
Data Sharing and Accessibility



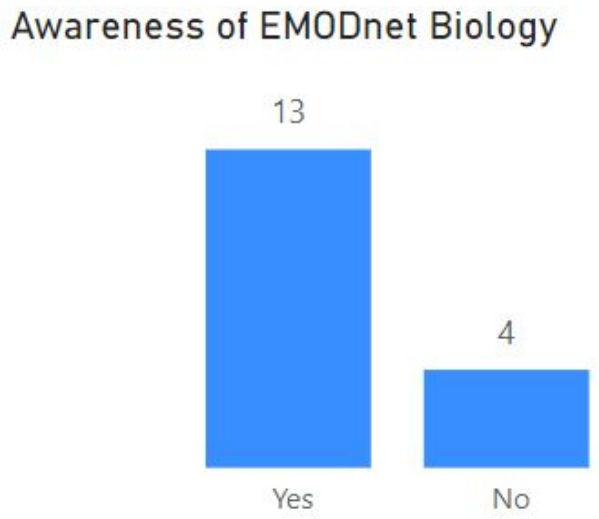


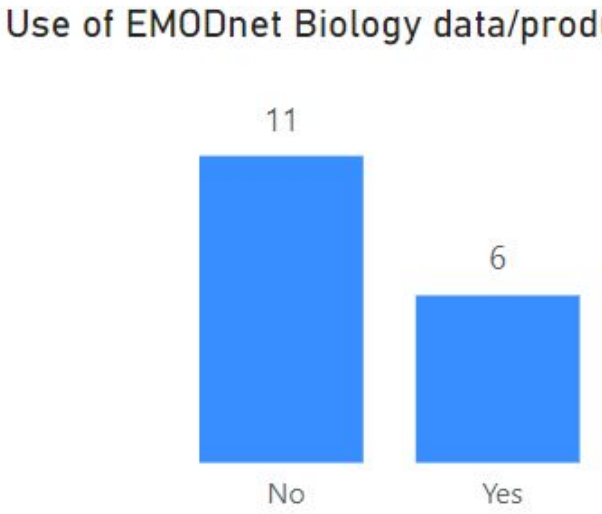
Challenges and Gaps

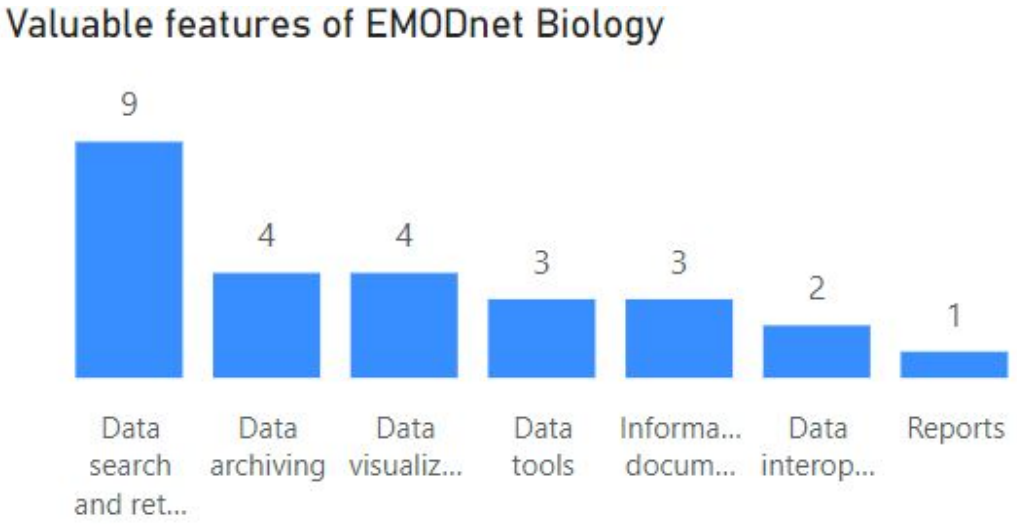




Awareness and Use of EMODnet Data Portal







Thank you for your attention!

Join at menti.com | use code 25 24 19 8

Nikola Holodkov (nholodkov@ogs.it)
OGS/EMODnet Biology WP4

Biology

Project pitches - Part 1

BlueRemediomics

Biodiversa MOSTFUN

AtlantECO, BiOcean5D, BlueCloud2026

EMODnet Biology Workshop

19th and 20th November 2024, Online

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.

Project name: BlueRemediomics

Project coordinator: Rob Finn (EMBL-EBI) and Chris Bowler (CNRS)

Presented by: Samuel Chaffron (CNRS)

Aim of the project:
Harnessing the Marine Microbiome for Novel Sustainable Biogenics and Ecosystem Services

Type of biological data collected: Microbiome sequencing data (metaBGT)

Geographical area: Global Ocean

Data availability, management, connection with EMODnet (access restrictions, standard protocols for data management in place, awareness of standards,…)

- Freely available public microbiome datasets at MGnify (https://www.ebi.ac.uk/metagenomics/)
- Data available via website or JSON:API at https://www.ebi.ac.uk/metagenomics/api/v1/

Any other information

- On-going global ocean presence/absence and abundance profiling of marine genomes
- Submission to EMODnet as netCDF ?

Project name:

biodiversa+
European Biodiversity Partnership

MOSTFUN
Monitoring strategies for aquatic fungi biodiversity

EMODnet
European Marine Observation and Data Network

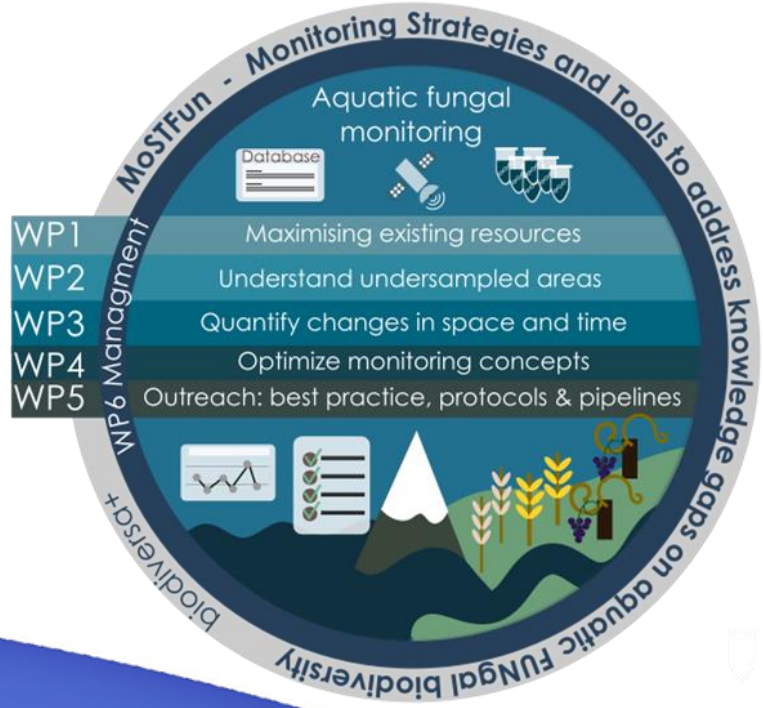
Project coordinator: Andreas Bruder SUPSI University of Applied Sciences and Arts of Southern Switzerland (CH)

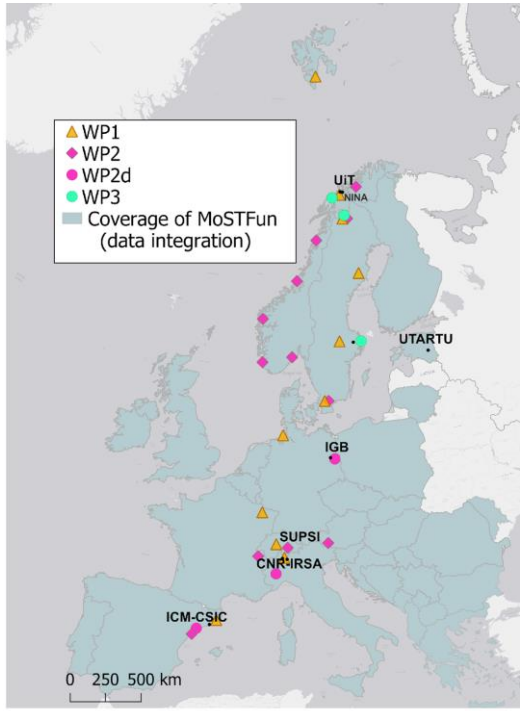
Presented by: Laura Garzoli CNR-IRSA National Research Council, Water Research Institute of (IT)

Aim of the project:

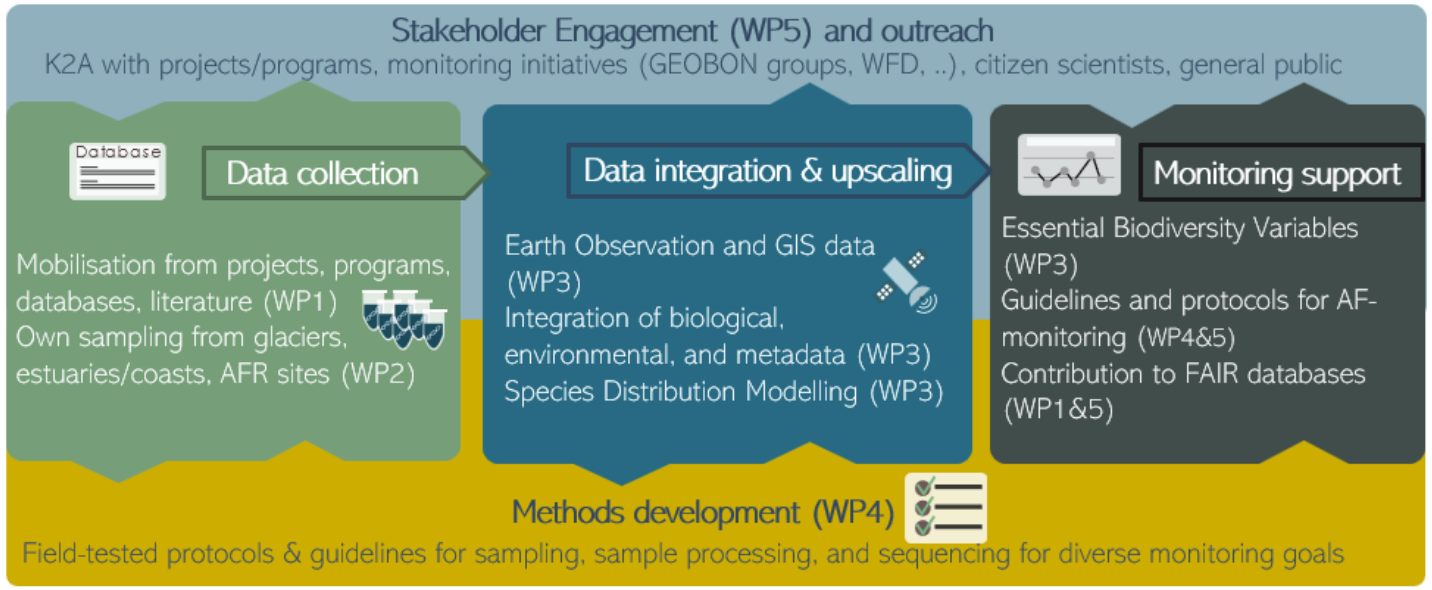
Type of biological data collected:
Water, Sediment, Leaves

Geographical area: EU





Data availability, management, connection with EMODnet (access restrictions, standard protocols for data management in place, awareness of standards,…)



Essential Biodiversity Variables (EBVs):

- Genetic composition
- Species populations
- Species traits
- Community composition
- Ecosystem function
- Ecosystem structure

Any other information: Please visit our website https://mostfun.eu/

Project name: AtlantECO, BiOcean5D, BlueCloud2026

Project coordinator: S. Pesant (EBI), C. de Vargas (CNRS), D . Schaap (Maris)

Presented by: M. Vogt (WP2 'Ecosystem structure and function'; WP3 'Data to knowledge', WB3 'Ecosystem EOVs')

Aim of the project: 'Atlantic Ocean Ecosystems structure/function/services', 'Coastal-Open Ocean Biodiversity', 'Digital Twin of the Ocean'

Type of biological data collected: 'microscopy (presence-absence, abundance, biomass), omics, imaging, acoustics, bio-prospecting molecules, plastisphere, biologically mediated carbon fluxes, pigments, indicators, OHI, …
Geographical area: Atlantic Ocean (global), European Coastal seas, global

Data availability, management, connection with EMODnet (access restrictions, standard protocols for data management in place, awareness of standards,…)

AtlantECO: DMP, DwC standard headers, WORMS taxonomy, CFC netcdf convention, project wide data collection templates, AtlantECO GeoNode, public data collection published in Zenodo so far, **destination for model data: EMODnet**

BiOcean5D: DMP, DwC standard headers, WORMS and others for taxonomy, CFC netcdf convention, project-specific data hub (access time-limited), project-wide data templates, all raw data to be transferred to long-term storage (ENA, Mgnify, Ecotaxa to EurOBIS); **destination for model data: EMODnet**

BlueCloud2026: DMP, public data in open cloud computing system, CFC netcdf conventions for outputs, web-based access to all data inputs and products, **EMODnet as project partner**

Any other information: Issues with data re-use with added value (biomass)

Biology

EMODnet Biology: Research data management 101

Rubén Pérez Pérez, Flanders Marine Institute (VLIZ)

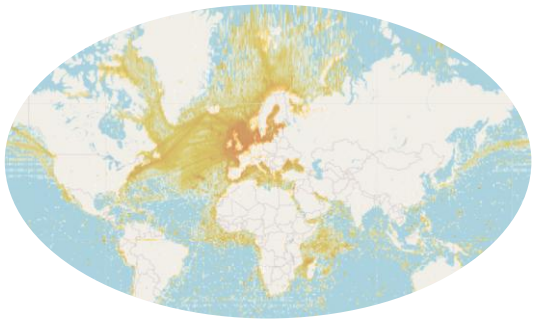
2024-11-19 | EMODnet Biology online workshop

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.

Ruben
the “data manager”

Science Support at the Data Centre

- Data management
- Data products
- Training
- Data networking



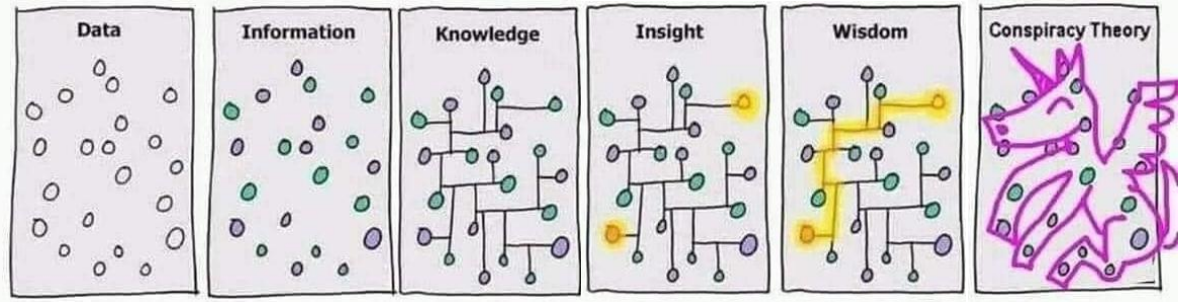
Government/public
Academic
Industry
Citizen

Research data management

Why it is important

FAIR & Open data

Research data cycle



Why it is important

“Data is a precious thing and will last longer than the systems themselves.”

Tim Berners-Lee, inventor of the World Wide Web

RDM

Why it is important

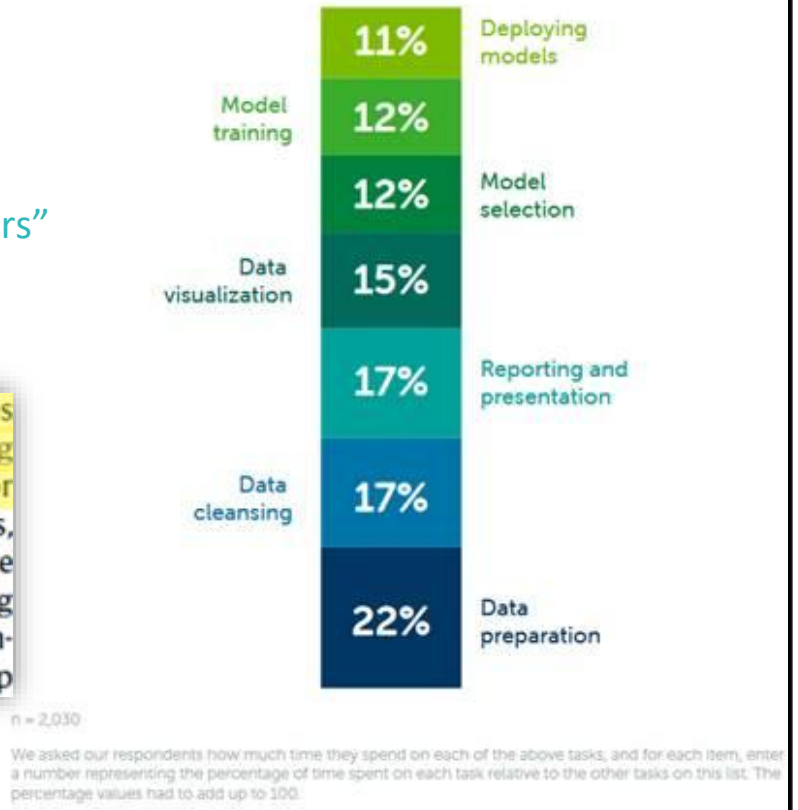
Order, chaos or organized chaos?

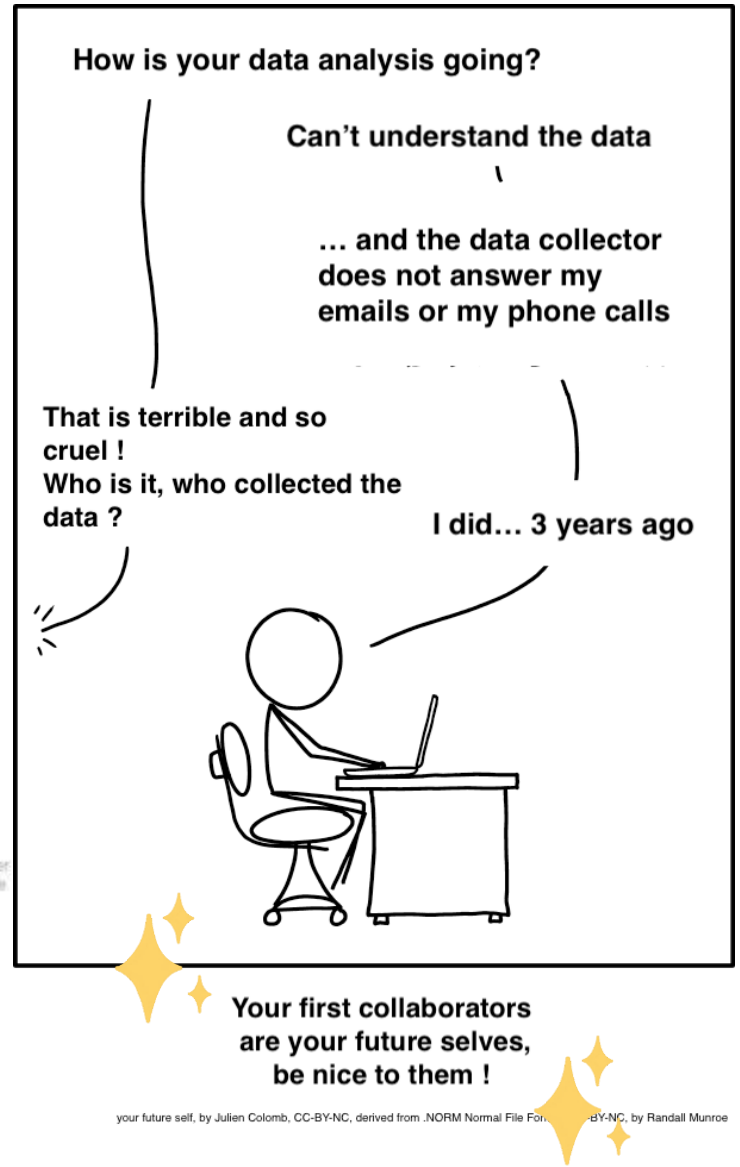
RDM

Why it is important

"We are all data providers and data users"

under the current system. Students in PhD programmes spend up to 80% of their time on 'data munging', fixing formatting and minor mistakes to make data suitable for analysis – wasting time and talent. With 400 such students, that would amount to a monetary waste equivalent to the salaries of 200 full-time employees, at minimum. So, hiring 20 professional data stewards to cut time lost to data wrangling would boost effective research capacity. Many top





More than 70% of researchers have tried and failed to reproduce another scientist's experiments, and more than half have failed to reproduce their own experiments. Those are some of the telling figures that emerged from *Nature*'s survey of 1,576 researchers who took a brief online questionnaire on reproducibility in research.

...ker, M. (2016). 1,500 scientists lift the lid on reproducibility. Nature, 533(7604).https://doi.org/10.1038/533452a
Mons, B. (2020). Invest 5% of research funds in ensuring data are reusable. Nature, 578(7796), 491-491. https://doi.org/10.1038/d41586-020-00505-7
Anaconda (2021). 2021 State of Data Science report. https://www.anaconda.com/resources/whitepapers/state-of-data-science-2021

EMODnet
European Marine Observation and Data Network

RDM

Why it is important

Personal benefits

Work more efficient & organised

More references & credits to your work

Career recognition

Collaborations

Moral obligations

Efficient use of public resources

Data are unique & come with a high cost

Facilitates data finding & re-use

- New research & new insights

Better data leads to better research

- Improved decisions-making
- Increased transparency & trust in science

EMODnet
European Marine Observation and Data Network

RDM

Why it is important

FAIR &
Open data

Research data management

Why it is important

FAIR & Open data

Research data cycle

EMODnet
European Marine Observation and Data Network

FAIR & Open data

Findable

- Rich metadata & available online
- Persistent identifier

Accessible

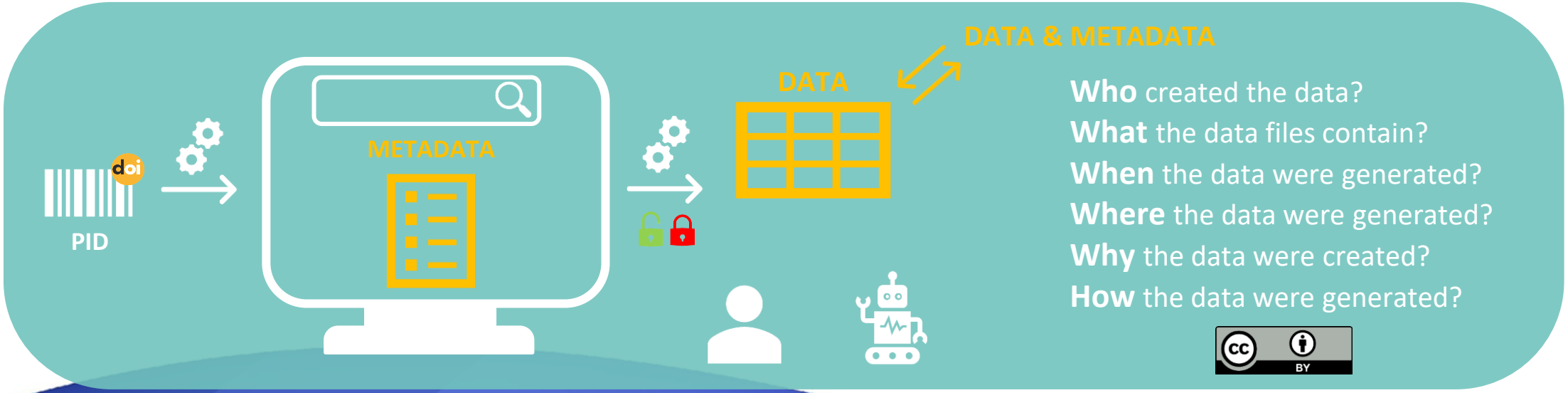
- Retrievable
- Accessible ≠ OPEN
- Authentication & authorisation steps
- Metadata should always be accessible

Interoperable

- Machine readable components
  - Open formats
  - Recognized standards
  - Linked data
  - Integration ready

Reusable

- Data ‘provenance’
- Data usage licence



Ghent University Data Stewards (2020). Knowledge clip: FAIR data principles. Available at: https://youtu.be/2uZxFu9SFi8

EMODnet
European Marine Observation and Data Network

FAIR & Open data

FAIR data

It is a spectrum

≠ Open data

Open data is data that anyone can access, use & share

EMODnet
European Marine Observation and Data Network

Research data cycle

Research data management

Why it is important

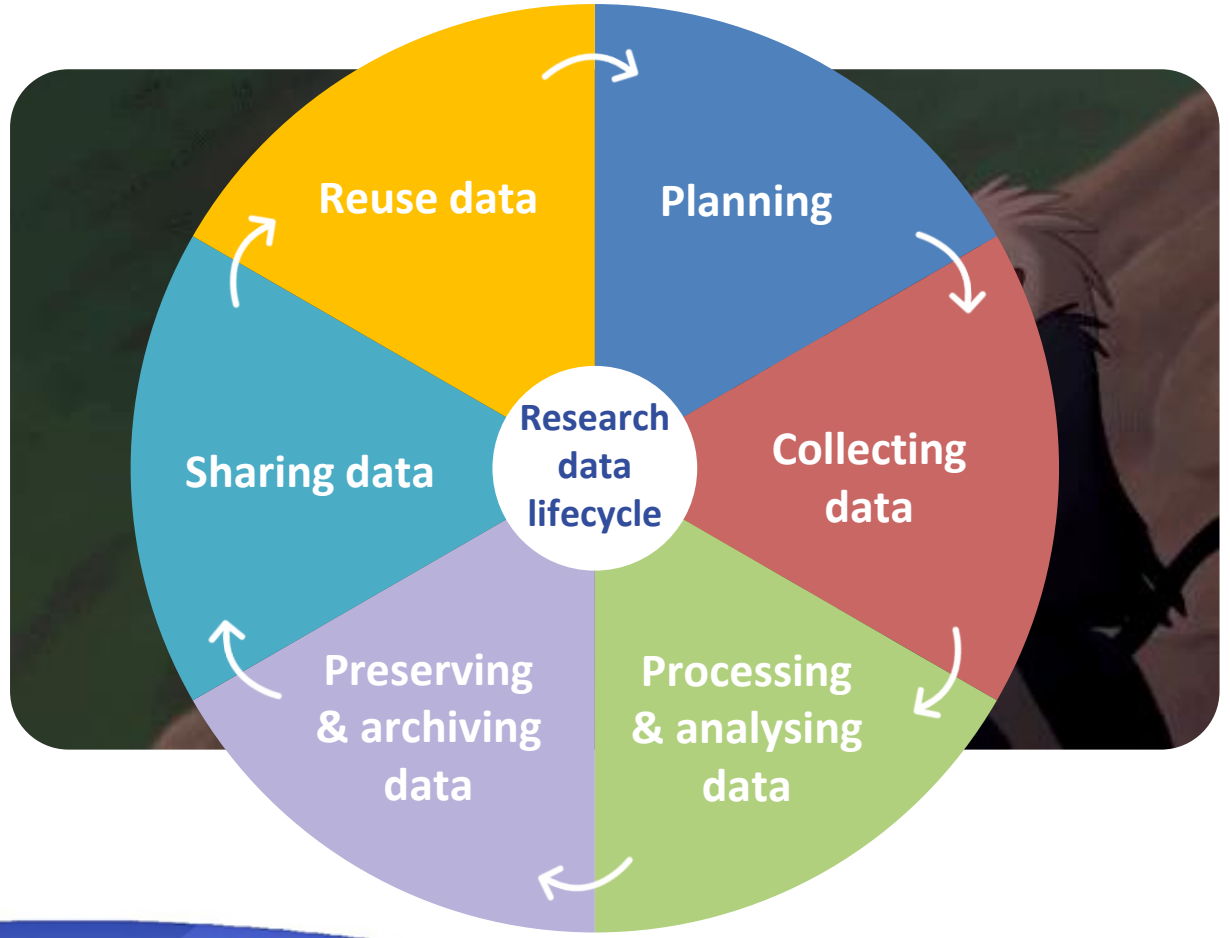
FAIR & Open data

Research data cycle

EMODnet
European Marine Observation and Data Network

Research data cycle

The circle of ~~life~~ data

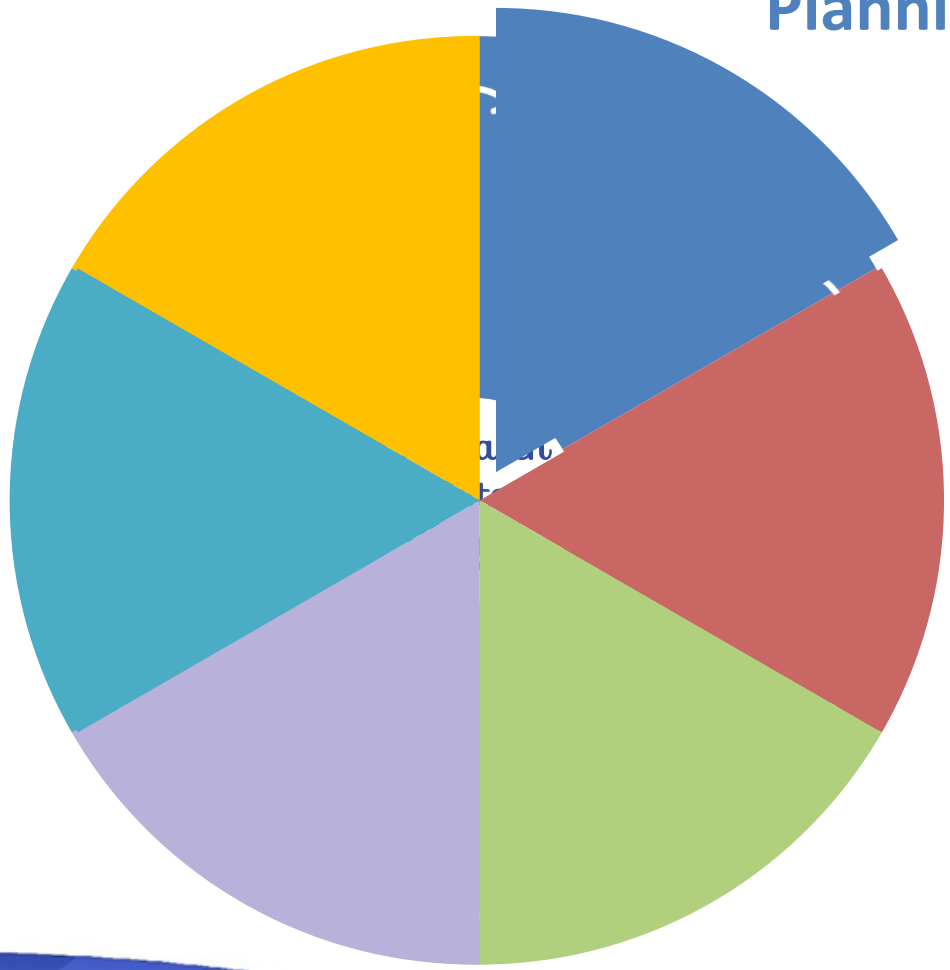


EMODnet
European Marine Observation and Data Network

Research data cycle

RDM in practice!

Planning { - DMP



Research data cycle

Data Management Plan

Planning

What?

- How data will be handled **during & after** a research project
- Formal & “living” document

Why?

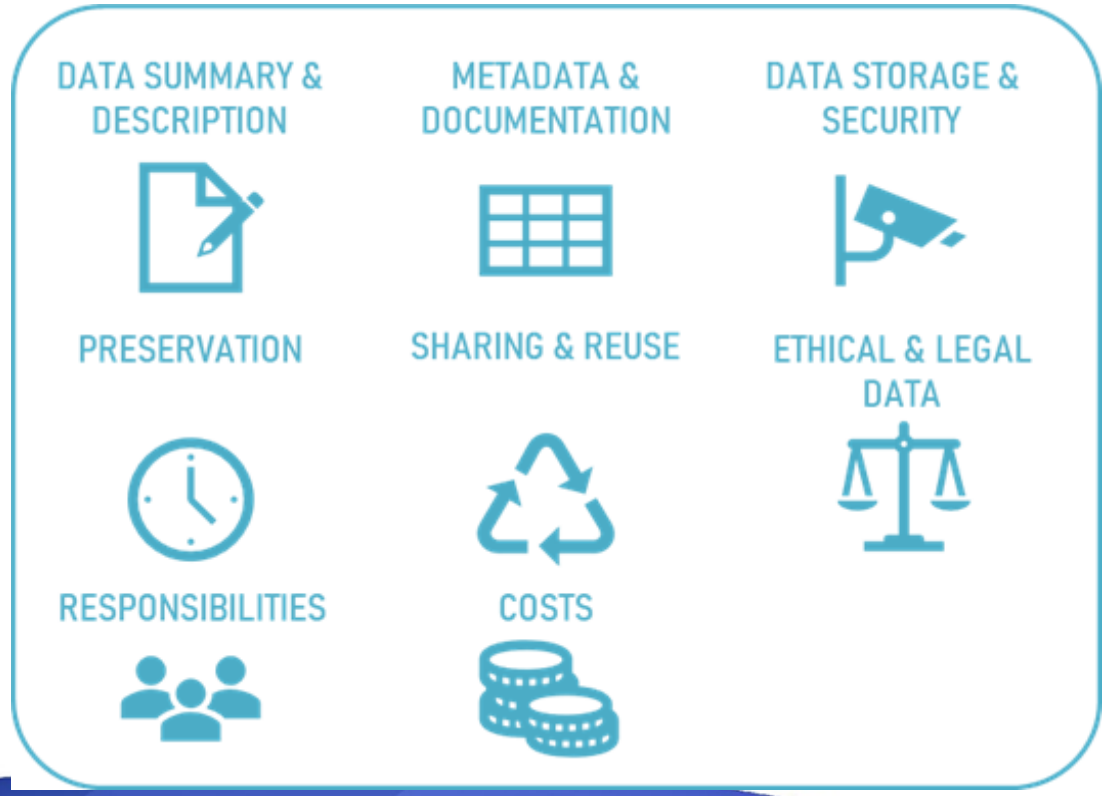
EMODnet
European Marine Observation and Data Network

Research data cycle

Data Management Plan

Planning

Content of DMP

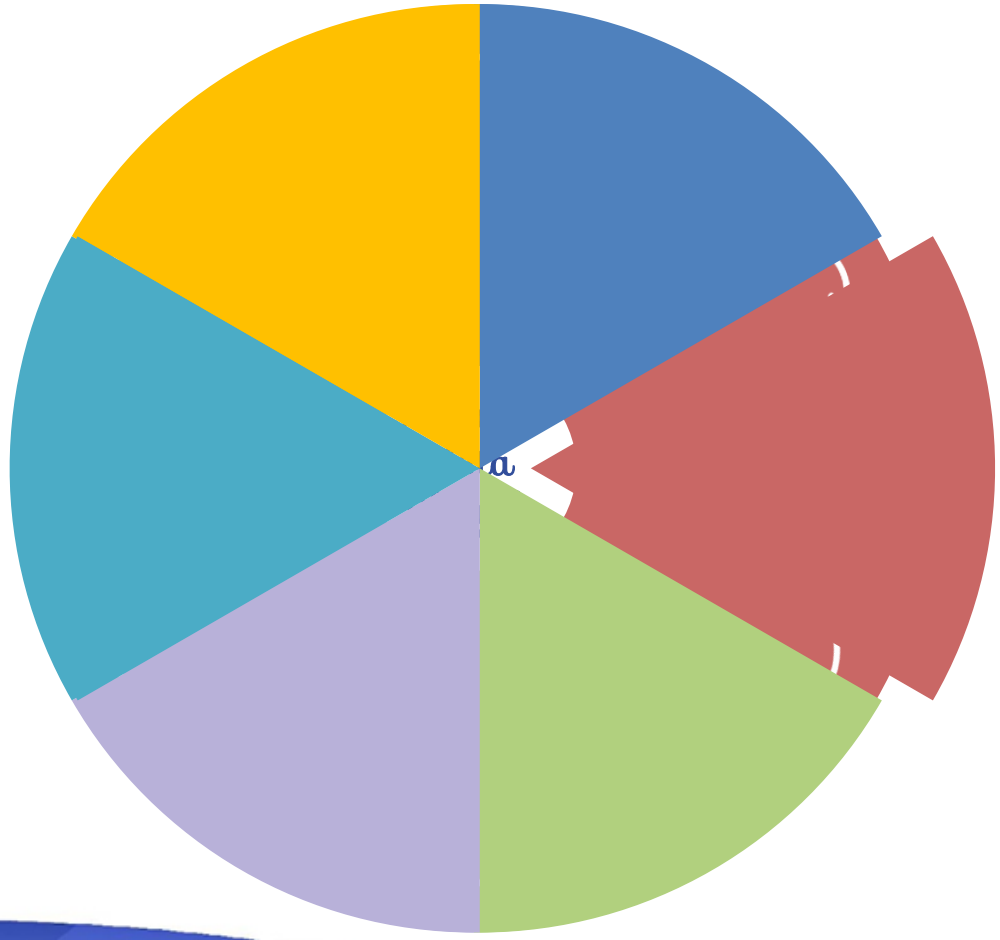


DMP templates

DMPonline.be

EMODnet
European Marine Observation and Data Network

Research data cycle



Collecting data

- Metadata
- Controlled vocabularies
- Standards

Research data cycle

Metadata and Documentation

Collecting

On three levels

- Project (directory structure)
- Files (naming conventions, READme)
- Data standards & vocabularies

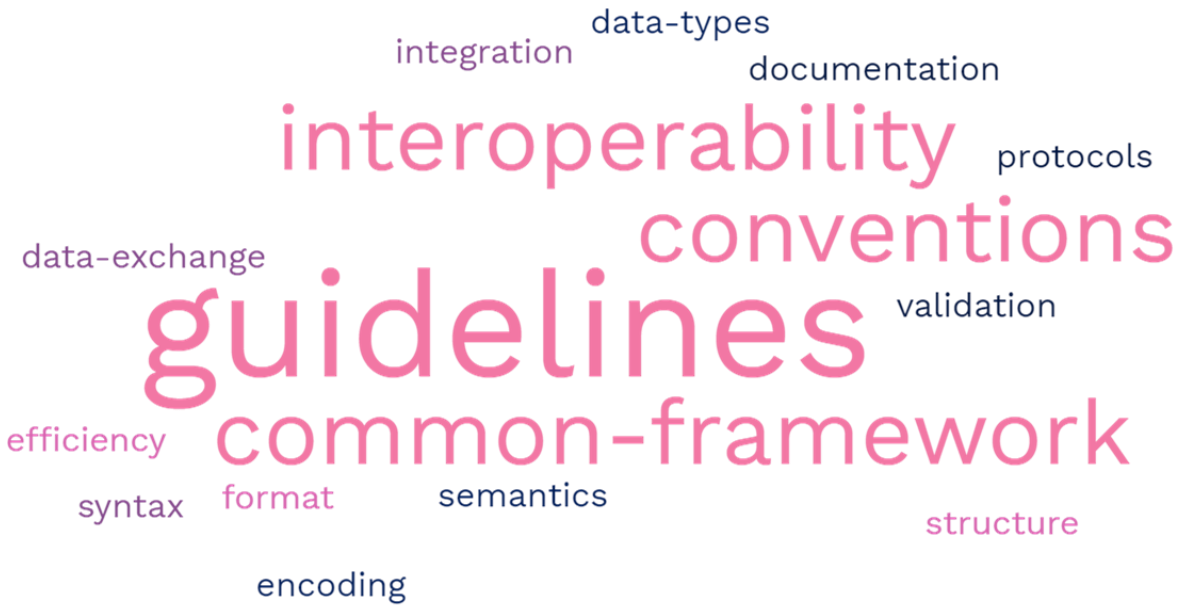
EMODnet
European Marine Observation and Data Network

Research data cycle

Data standards

Collecting

Global & multidisciplinary standards:



"Set of guidelines or rules that specify how data should be structured, formatted, and represented to ensure consistency, interoperability, and efficient data exchange"

ISO

International Organization for Standardisation

EMODnet
European Marine Observation and Data Network

Research data cycle

Data standards

Collecting

Domain specific standards:

DwC

= Darwin Core

EML

= Ecological Metadata Language

eventID	parentEventID	eventDate	decimalLongitude	decimalLatitude
site__1			54.7943	16.9425
zone__1	site__1			
zone__2	site__1			
zone__3	site__1			
quadrat__1	zone__1	2019-01-02		
transect__1	zone__2	2019-01-03		
transect__2	zone__3	2019-01-04		

id	occurrenceID	scientificName
quadrat__1	occ__1	Ulva rigida
quadrat__1	occ__2	Ulva lactuca
transect__1	occ__3	Plantae
transect__1	occ__4	Plantae
transect__2	occ__5	Gracilaria
transect__2	occ__6	Laurencia

- Basic Metadata
- Geographic Coverage
- Taxonomic Coverage
- Temporal Coverage
- Keywords
- Associated Parties
- Project Data
- Sampling Methods
- Citations
- Collection Data
- External links
- Additional Metadata

EMODnet
European Marine Observation and Data Network

Research data cycle

Controlled vocabularies

Collecting

- List of terms where each term means just one thing
- Ensure standardisation

Example Marine Sciences

National Oceanography Centre
British Oceanographic Data Centre BODC

- Biomass

Identifier ↑	Preferred label ↑	Alternative label ↑	Definition ↑
SDBIOL09	Dry weight biomass of biological entity specified elsewhere per unit volume of the water body	WaterDryWtBiom_BE007117	The mass measured after drying at elevated temperatures until a stable mass is reached, of an identified biological object described elsewhere in the metadata occurring in a given volume of any body of salt or fresh water.
SDBIOL07	Ash-free dry weight biomass of biological entity specified elsewhere per unit volume of the water body	WaterAshFreeBiom_BE007117	The mass lost on ignition of an identified biological object described elsewhere in the metadata occurring in a given volume of any body of salt or fresh water.
SDBIOL04	Wet weight biomass of biological entity specified elsewhere per unit volume of the water body	WaterWetWtBiom	The mass as caught of an identified biological object described elsewhere in the metadata occurring in a given volume of any body of salt or fresh water.
SDBIOL12	Biomass as carbon of biological entity specified elsewhere per unit volume of the water body by computation	WaterCarbonBiomassConv	The carbon biomass, calculated from the cell counts using literature conversion factors, of an unspecified biological entity in a given volume of any body of salt or fresh water.

EMODnet
European Marine Observation and Data Network

Research data cycle

Collecting

Taxonomic standard

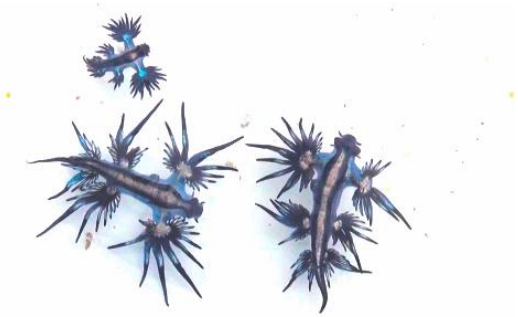
WoRMS provides the most authoritative list of names of all marine species globally, ever published

WoRMS World Register of Marine Species

WoRMS taxon details

Glaucus atlanticus Forster, 1777

AphiaID	140022 (urn:lsid:marinespecies.org:taxname:140022)
Classification	Biota > Animalia (Kingdom) > Mollusca (Phylum) > Gastropoda (Class) > Heterobranchia (Subclass) > Euthyneura (Infraclass) > Ringipleura (Subterclass) > Nudipleura (Superorder) > Nudibranchia (Order) > Cladobranchia (Suborder) > Aeolidioidea (Superfamily) > Glaucidae (Family) > *Glaucus* (Genus) > *Glaucus atlanticus* (Species)
Status	accepted

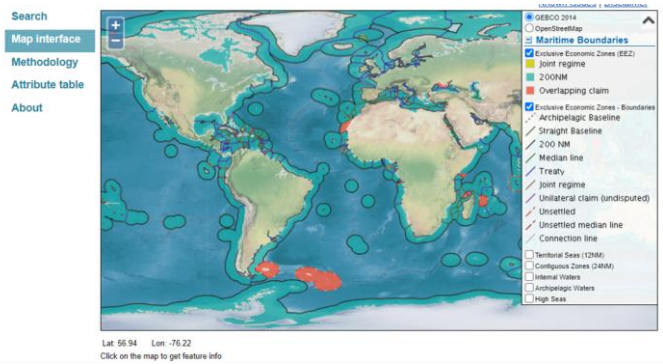


Geographic standard

Standard list of marine georeferenced place names & areas

Marineregions.org

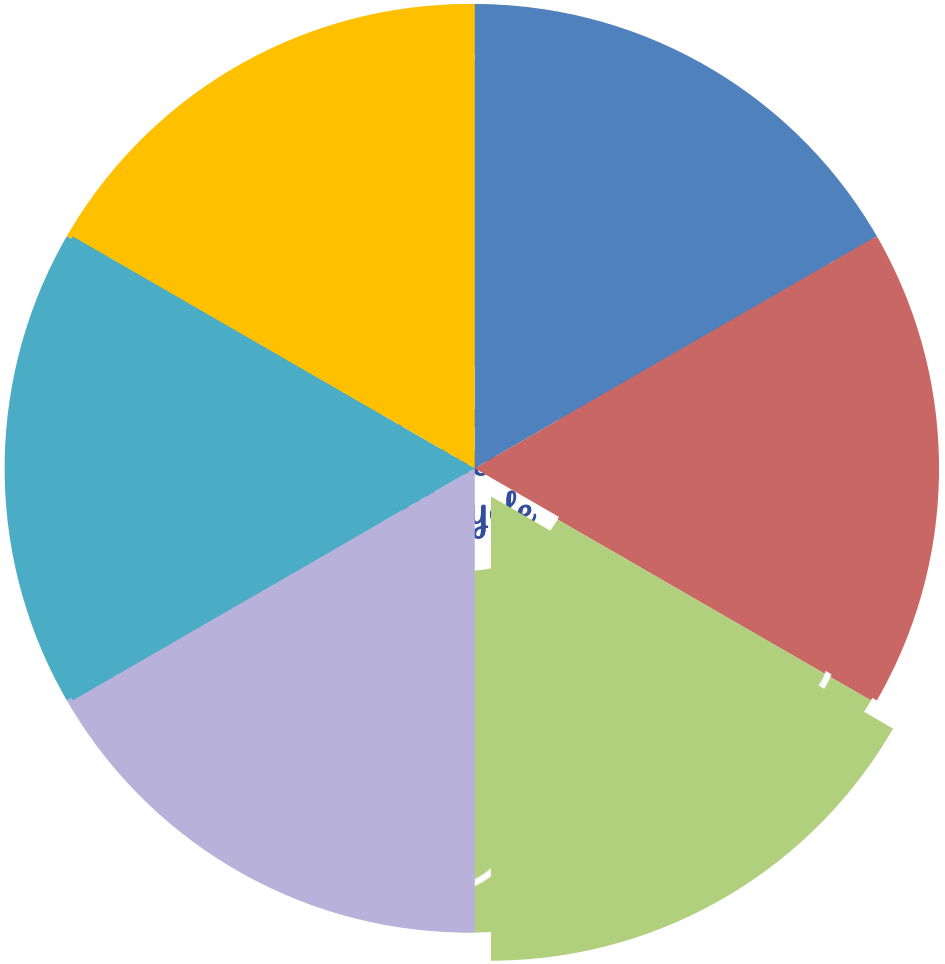
a standard for georeferenced marine names



EMODnet

European Marine Observation and Data Network

Research data cycle



Data curation -
File organization -

Processing & analysing data

EMODnet
European Marine Observation and Data Network

Research data cycle

Data curation

Processing & analysing

Curation steps

Data exploration

Data transformation

Data enrichment

Data validation

Reproducible procedures

Keep raw data intact

Document transformation

Version Control

Document Quality Control procedures

Use Open formats




EMODnet
European Marine Observation and Data Network

Research data cycle

Data curation

Processing & analysing

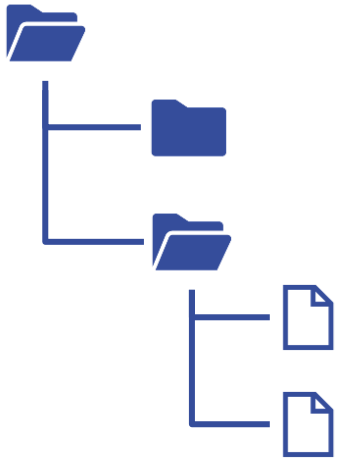
Name	Phone	Birth date	Country
John Smith	445-881-4478	August 12, 1989	Belgium
Fitch, Marie	(876)546-8165	June 15, 72	US
Deere, Alan	+1-189-456-4513	11/12/1965	USA

Name	Phone	Birth date	Country
John Smith	445-881-4478	1989-08-12	Belgium
Marie Fitch	876-546-8165	1972-06-15	USA
Alan Deere	189-456-4513	1965-11-12	USA

EMODnet
European Marine Observation and Data Network

Research data cycle

File naming conventions



Processing & analysing

Recommendations:

Be consistent

Use standards (e.g. YYYYMMDD)

Do not use special characters or spaces

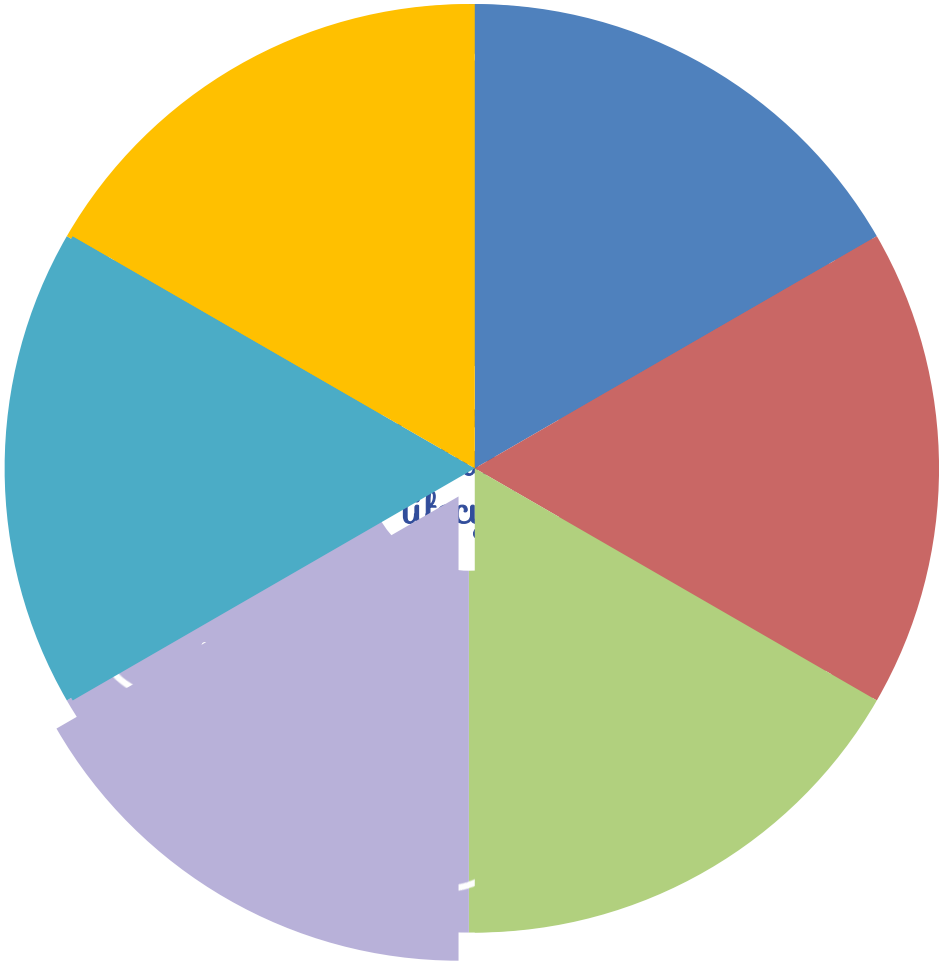
Avoid words like 'draft', 'final'... – use version numbers instead (v01, v02)

...

Examples of files without a naming convention:	Examples of files with a naming convention:
Meeting notes jan 10.doc	20230110_OT_ODM_exercise1_v01.doc
Third_test.xls	20230110_OT_ODM_exercise1_v03.doc
ProjectProposalFirstVersion.doc	20230109_OT_ODM_EvaluationResults.xls
Project-data.xls	20230109_OT_ODM_RDLC.jpg

EMODnet
European Marine Observation and Data Network

Research data cycle



- Data archiving

Preserving & archiving data

Research data cycle

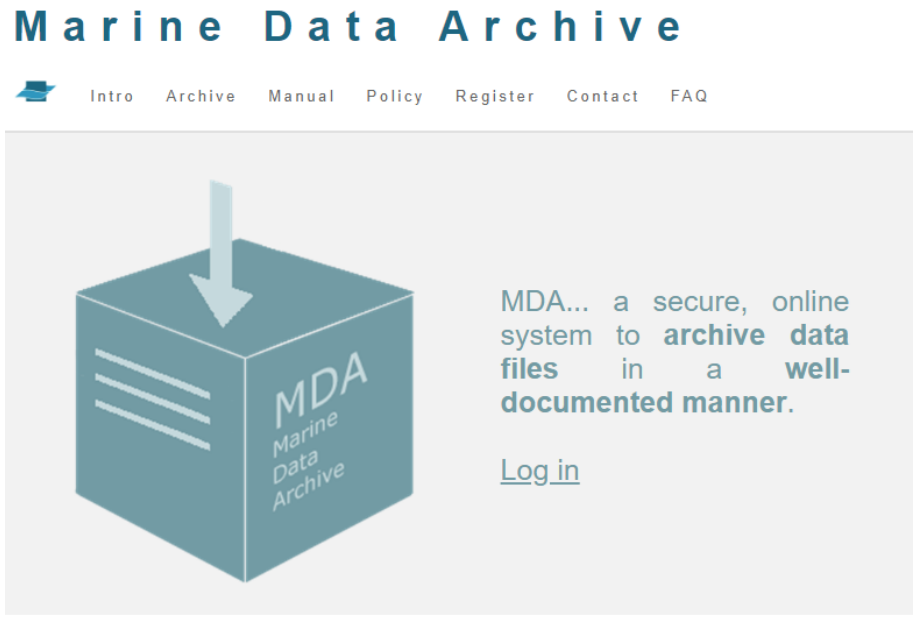
Data archiving

Preserving & archiving

Marine Data Archive - MDA

= trusted data repository for marine, coastal and estuarine research

- **Closed** repository for personal files & projects / collaboration
- **Open** repository for data publication

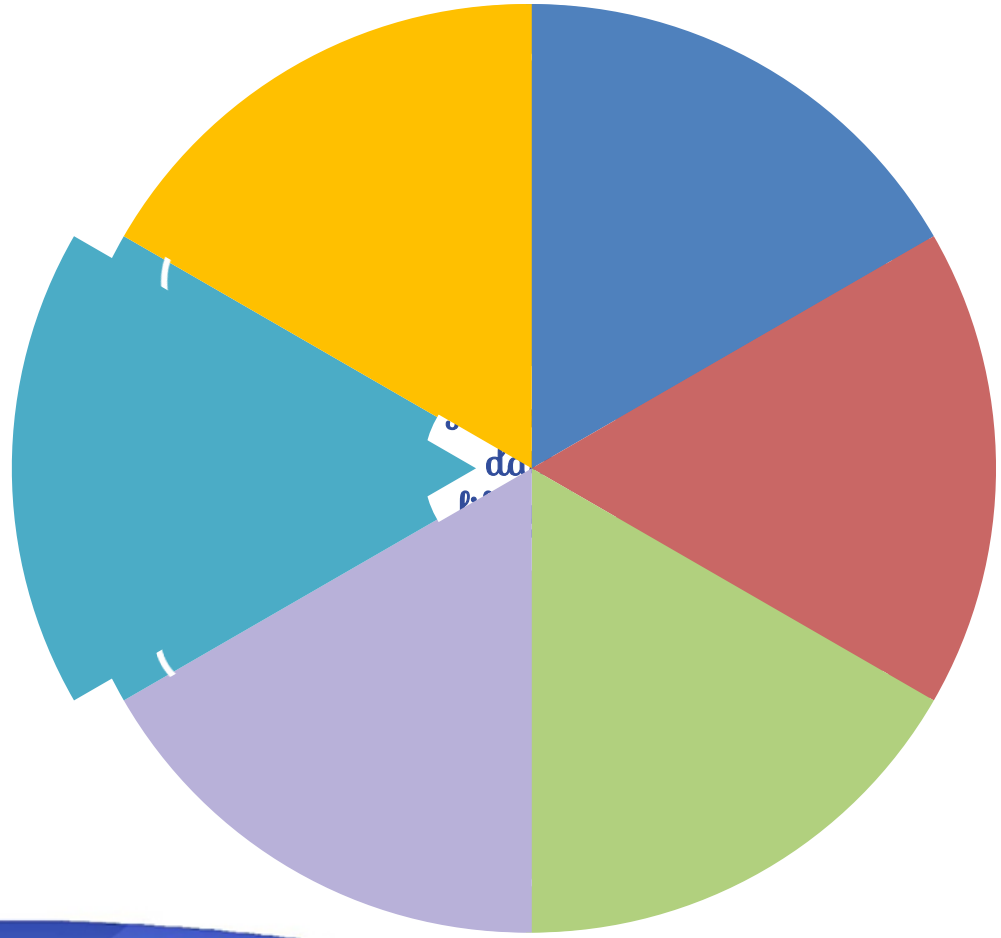


https://mda.vliz.be/

EMODnet
European Marine Observation and Data Network

Research data cycle

- Searchable resources
- IMIS



Sharing data

Research data cycle

Searchable resources

Sharing

Repositories

- **Archiving and sharing**
- **Generic, discipline specific or institutional**

re3data.org
REGISTRY OF RESEARCH DATA REPOSITORIES

Catalogue

- **Description** (rich metadata) **of and link to data**

FAIRsharing.org
standards, databases, policies

Portal

- **Archiving and sharing + interactive tools** (visualisation, combining data, ...)
- **Often thematic**

EMODnet
European Marine Observation and Data Network

Research data cycle

Searchable resources

Sharing

EMODnet

LifeWatch BELGIUM

hcmr ΕΛΚΕΘΕ

med OBIS Ocean Biodiversity Information System

SeaDataNet

VLIZ IMIS

museum Belgian Marine Data Centre

and more ...

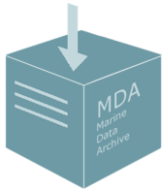
EMODnet European Marine Observation and Data Network

Research data cycle

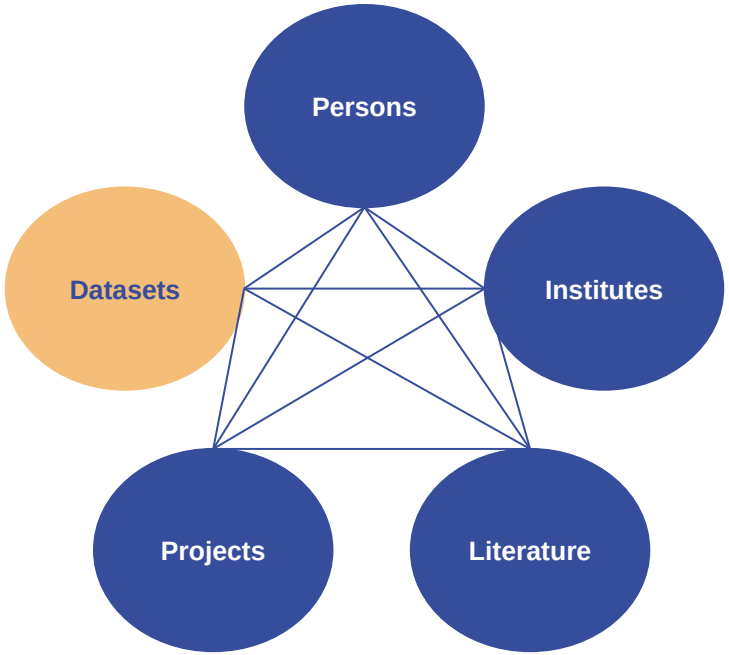
Integrated Marine Information System

Sharing

= catalogue with metadata information about:

- All datasets (open / not open)
- Related to marine and coastal research / topics
- Link to data 

or contact person



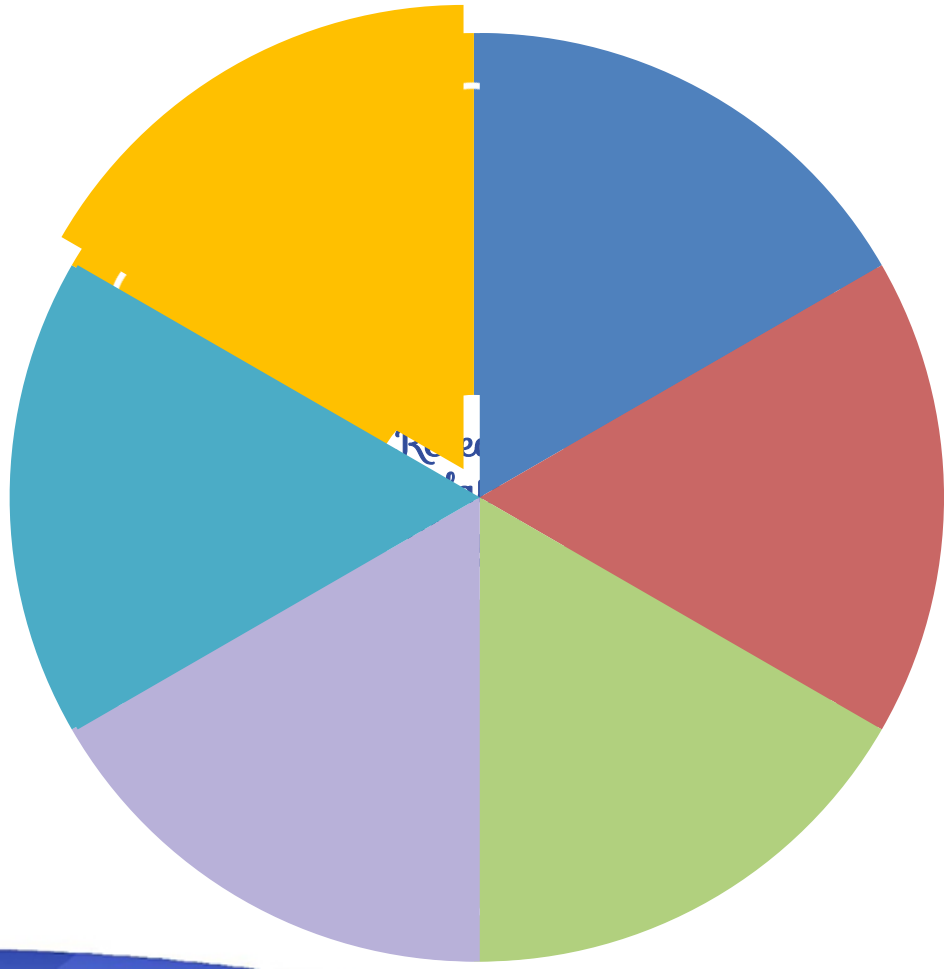
https://www.vliz.be/imis

EMODnet
European Marine Observation and Data Network

Research data cycle

Reusing data

- Provenance
- Rights



Reusing data

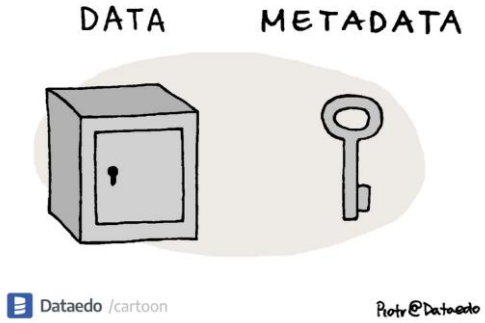
Provenance and documentation

Usage license and credit

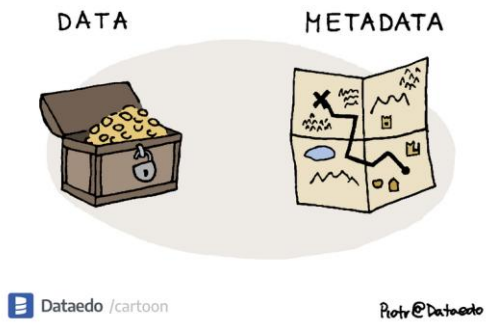
creative commons

EMODnet
European Marine Observation and Data Network

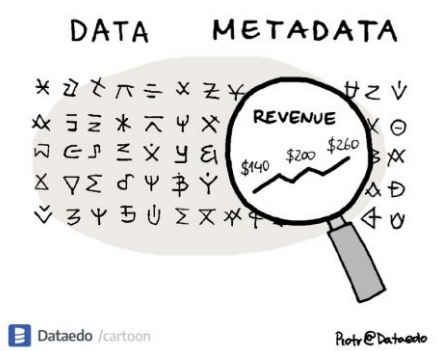
Reusing data



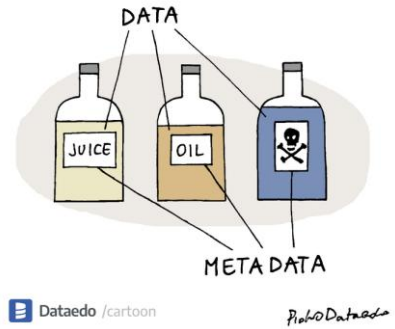
Metadata is the key to data



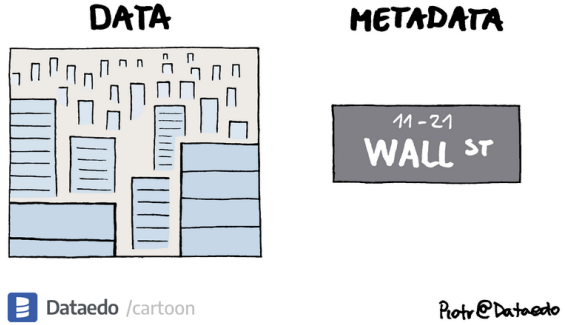
Data is treasure, metadata is a map



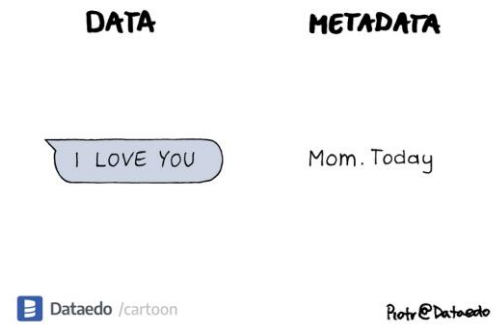
Decoding information



A Matter of Life and Death



How to get to the data?



Context and Source matters

EMODnet
European Marine Observation and Data Network

Research data cycle

Example IMIS metadata record for dataset: https://vliz.be/en/imis?module=dataset&dasid=4687

Data reuse

IMIS

log in

Publications | Institutes | Persons | Datasets | Projects | Maps

[report an error in this record]

LifeWatch observatory data: zooplankton observations by imaging (ZooScan) in the Belgian Part of the North Sea

Citable as data publication

Flanders Marine Institute (VLIZ), Belgium (2023): LifeWatch observatory data: zooplankton observations in the Belgian Part of the North Sea. https://doi.org/10.14284/584

Download Data

Previous versions (6) view

Contact: data@vliz.be

Access data

EMODnet | EurOBIS | GBIF | OBIS

Archived data

IPT

Also accessible through:
http://rshiny.lifewatch.be/zooscan-data/

Availability: This dataset is licensed under a Creative Commons Attribution 4.0 International License.

Notes: Images are available upon request via LifeWatch Belgium (info@lifewatch.be)

Description

In the framework of the Lifewatch marine observatory a number of fixed stations on the Belgian Part of the North Sea (BPNS) are visited on a monthly or seasonal basis using the RV Simon Stevin. A grid of nine stations covers the coastal zone and are sampled monthly. Eight additional stations, located further at sea, are sampled on a seasonal basis. This dataset contains zooplankton observations in the Belgian Part of the North Sea (BPNS) since 2012. Zooplankton is sampled by vertical WP2 net tows, samples scanned with ZooScanner and identification with plankton analyser software, followed by manual validation.

EMODnet
European Marine Observation and Data Network

EMODnet Seabed Habitats

Ville Karvinen, Coordinator (Syke)

Nov 19th, 2024

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.

Seabed Habitats

EMODnet
European Marine Observation and Data Network

Content

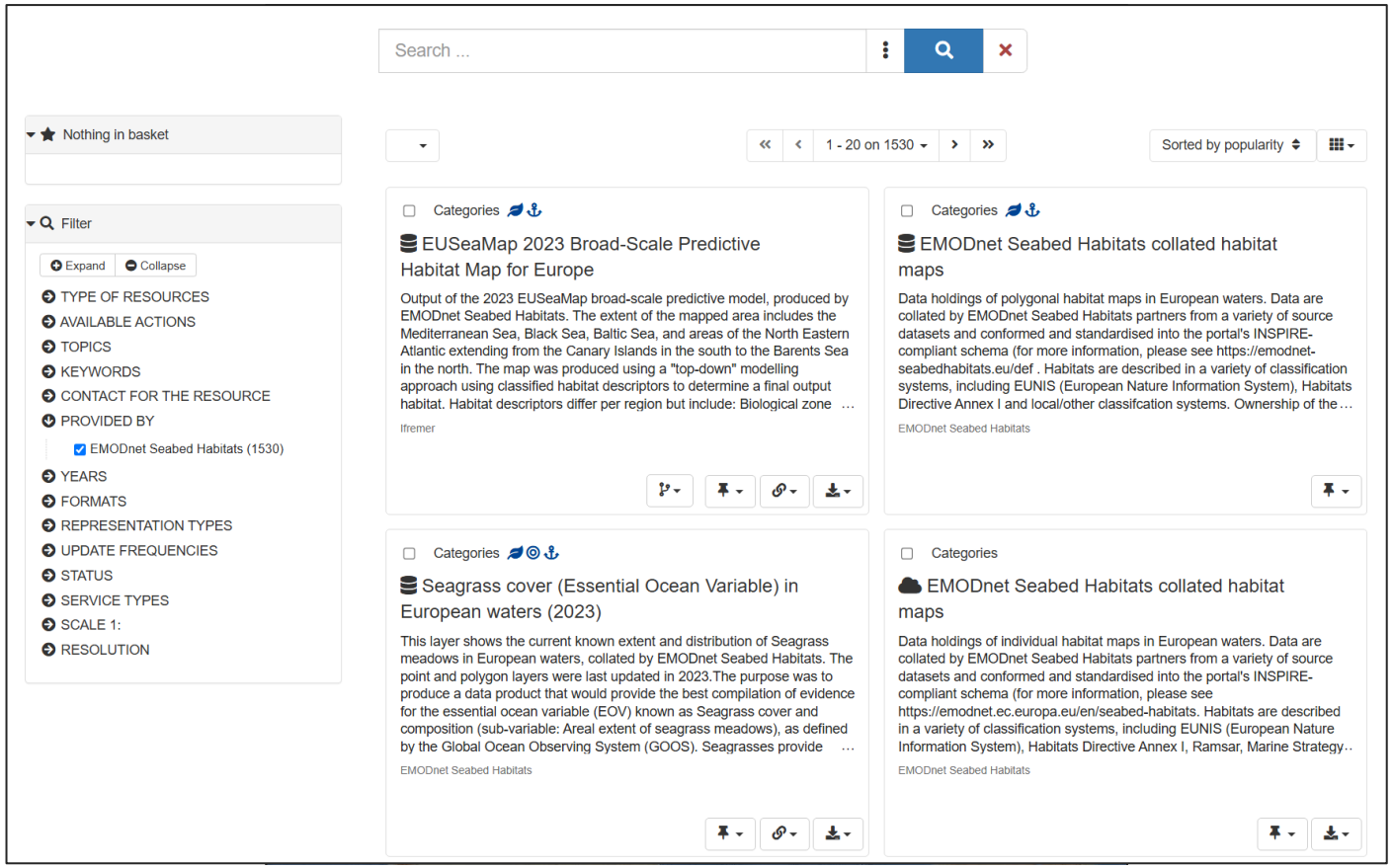
- EMODnet Seabed Habitats Overview
- Catalogue and products
- Workflow, guidelines and recommendations for projects

EMODnet Seabed Habitats

Consortium of 11 partners and 4 subcontractors

EMODnet Seabed Habitats

Catalogue and products



EMODnet Seabed Habitats in the EMODnet products catalogue at https://emodnet.ec.europa.eu/geonetwork/srv/eng/catalog.search

Products available in

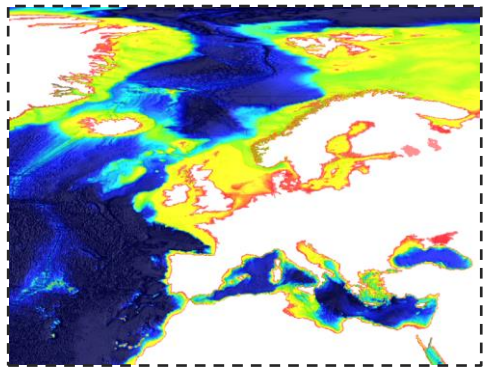
- EMODnet Map Viewer
- EMODnet products catalogue

Main products are

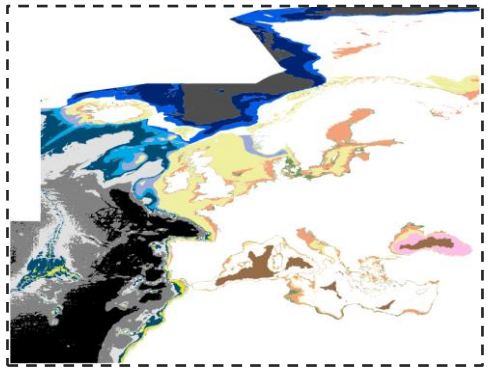
- EUSeaMap
  - CaribbeanSeaMap and CaspianSeaMap
- Survey data, habitat maps and models
- Composite products

EMODnet Seabed Habitats catalogue - EUSeaMap

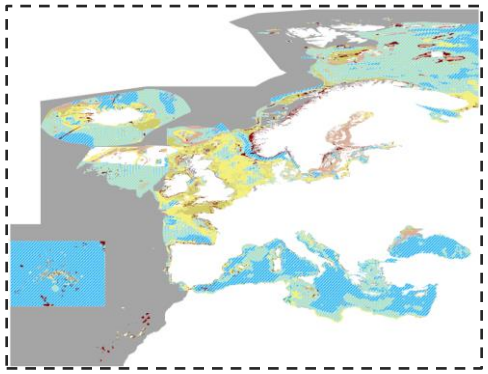
Modelled broadscale seabed habitat map for all European seas



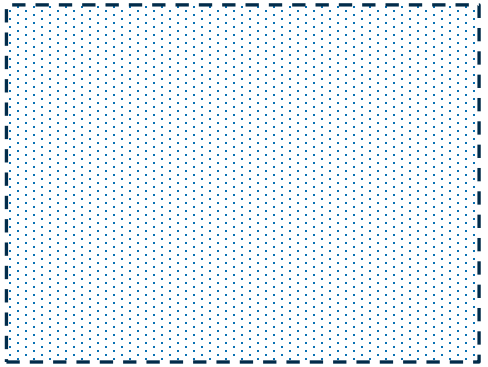
Bathymetry



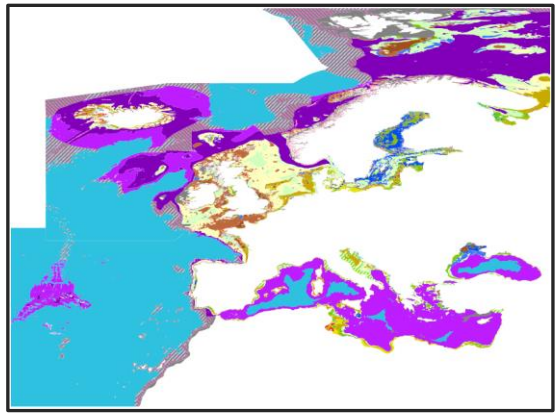
Biological zones



Substrate type

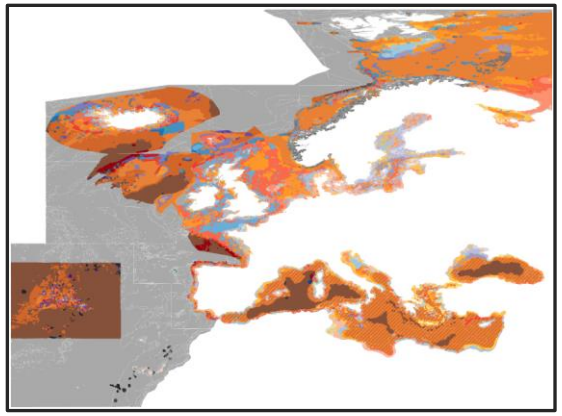


Other background layers



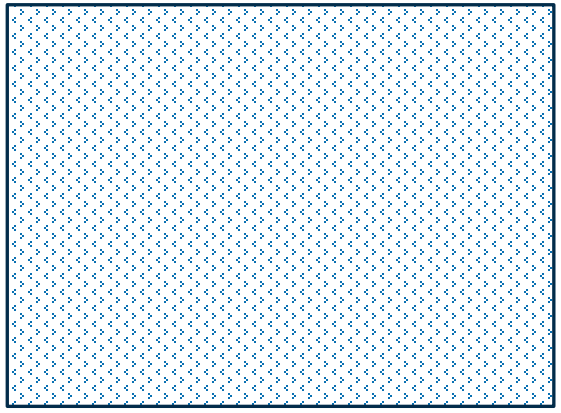
MSFD Benthic Broad Habitat Type

+



EUNIS (2007 & 2019)

+



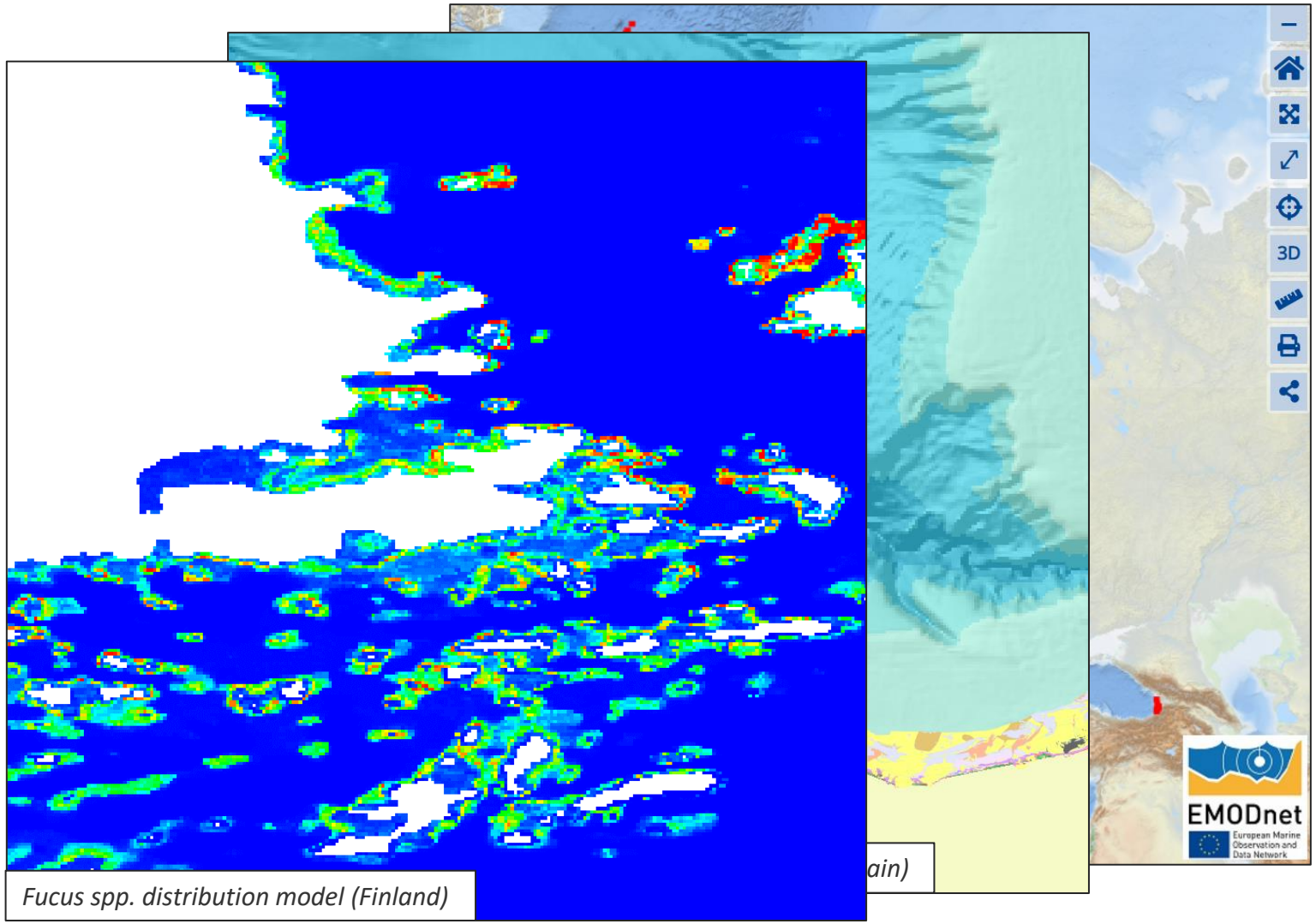
Other classifications

EMODnet Seabed Habitats catalogue - survey data and habitat maps

Habitat maps from national and regional collections as well as from completed EU projects

Survey data and habitat maps

- Point habitat data
- Habitat maps
- Modelled habitat maps

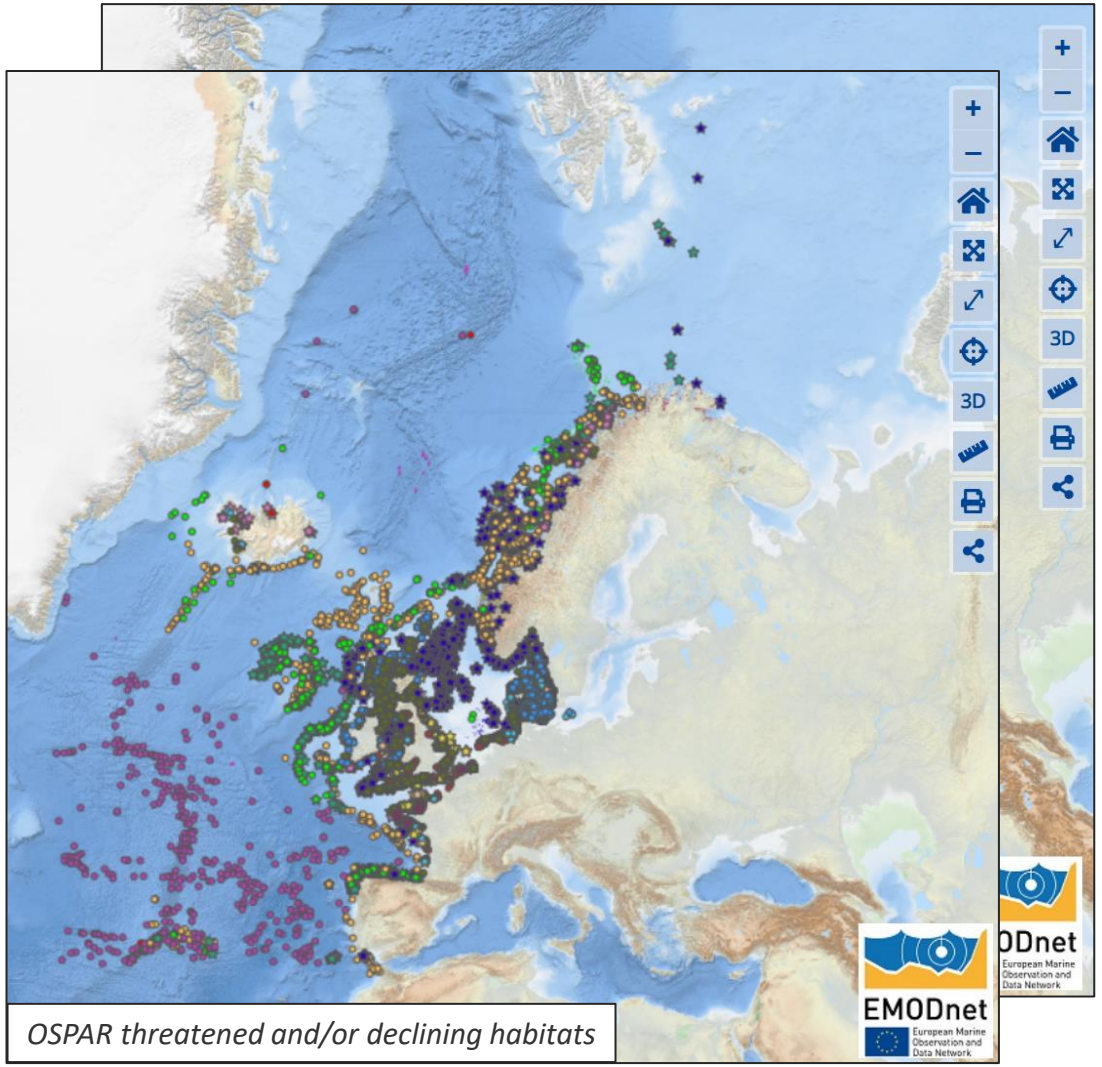


Fucus spp. distribution model (Finland)

EMODnet Seabed Habitats catalogue – Composite products

Combined and harmonized data products produced by ESH

- Essential Ocean Variables (EOVs)
- Essential Fish Habitats (EFHs)
- Coastal Wetlands
- Vulnerable Marine Ecosystems (VMEs)
- Regional Sea Convention specific products

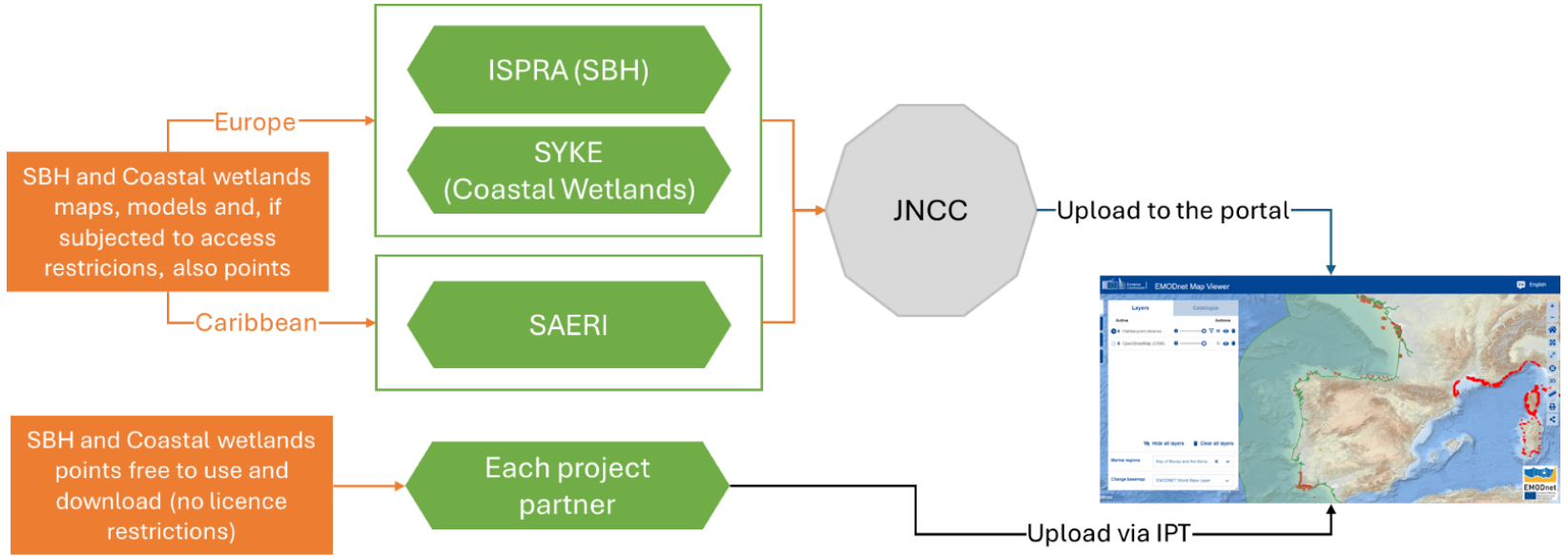


OSPAR threatened and/or declining habitats

EMODnet Seabed Habitats data ingestion process

Workflow from data identification to publishing

- Consortium identifies datasets to be published for each project phase
- Additional datasets identified and included during project
- Data formatted to ESH Data Exchange Format and harmonized to comply with INSPIRE standards
- Metadata formatted and harmonized
- Quality control (e.g. geometry checks for spatial data)
- Collections combined and harmonized into single layers when feasible



Consortium members to format their data/metadata and send to regional data coordinator / upload freely accessible point data via IPT by:	Regional / Thematic coordinator to send Qced data to JNCC by:	JNCC to upload data to portal by
1. 22 January 2024	1. 29 March 2024	1. 30 April 2024
2. 10 July 2024	2. 02 September 2024	2. 27 September 2024
3. 20 January 2025	3. 31 March 2025	3. 28 April 2025
4. 10 July 2025	4. 01 September 2025	4. 24 September 2025

EMODnet Seabed Habitats Phase V (2023–2025) data submission timetable for seabed habitat maps, points and models

EMODnet recommendations for EU research projects

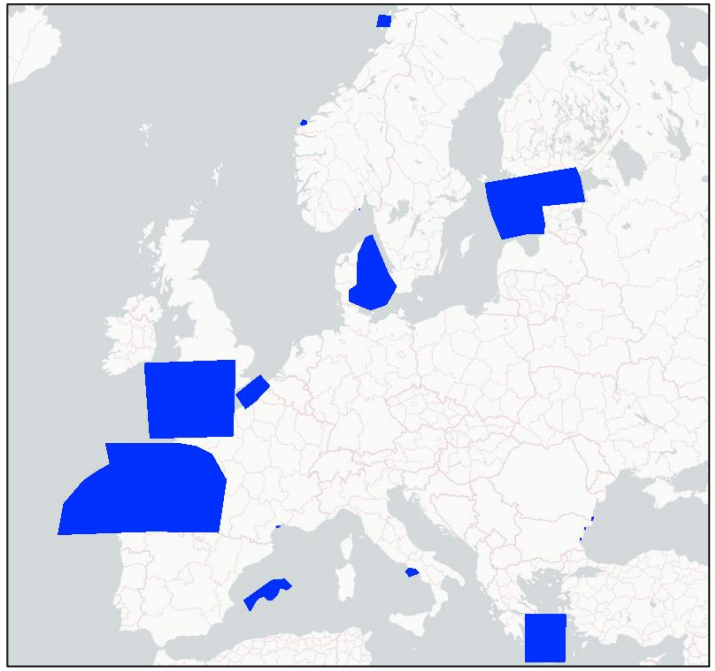
General guidelines

- Formulate a Data Management Plan
- Adopt FAIR data principles
- Engage EMODnet early in the project
  - EMODnet Data Ingestion can assist in planning data flows, integration, ingestion, archiving etc.
- For more information, see: https://www.emodnet-ingestion.eu/guidelines/guidelines-for-european-projects-data-submissions
- If you have questions regarding habitat data, please reach out to EMODnet Seabed Habitats: emodnetseabedhabitats@jncc.gov.uk (general data related issues) / ville.karvinen@syke.fi (coordinator)

EMODnet Seabed Habitats and EU projects' data

Case OBAMA-NEXT

- EU Horizon project with 12 learning sites around Europe
- Wide variety of methods used in both benthic and pelagic research



Challenge: How to effectively collate, standardize, QC and publish such a wide variety of data?
Proposed solution: Work together with the Data Manager and Learning Sites to specify data types and the timetable, engage with EMODnet Data Ingestion and the thematic lots

Let’s work together to ensure smooth and seamless data flows!

If you have habitat data questions, please contact
emodnetseabedhabitats@jncc.gov.uk (general data related issues)
ville.karvinen@syke.fi (coordinator)

Biology

Project pitches - Part 2

CONTRAST

B-USEFUL

MARHAB

EMODnet Biology Workshop

19th and 20th November 2024, Online

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.

Project name: CONTRAST

Project coordinator: Steven Brooks (NIVA)

Presented by: Steven Brooks

Aim of the project: To investigate the impacts of chemicals of emerging concern on the marine environment

Type of biological data collected: Biological effects data, biodiversity in sediments and water column

Geographical area: North Sea, North-East Atlantic, Mediterranean Sea

Data availability, management, connection with EMODnet (access restrictions, standard protocols for data management in place, awareness of standards,…)

- Data management plan in place
- FAIR principles*
- Standards in place, include the use of common vocabularies to code, for example, keywords, data disciplines and measured parameters.

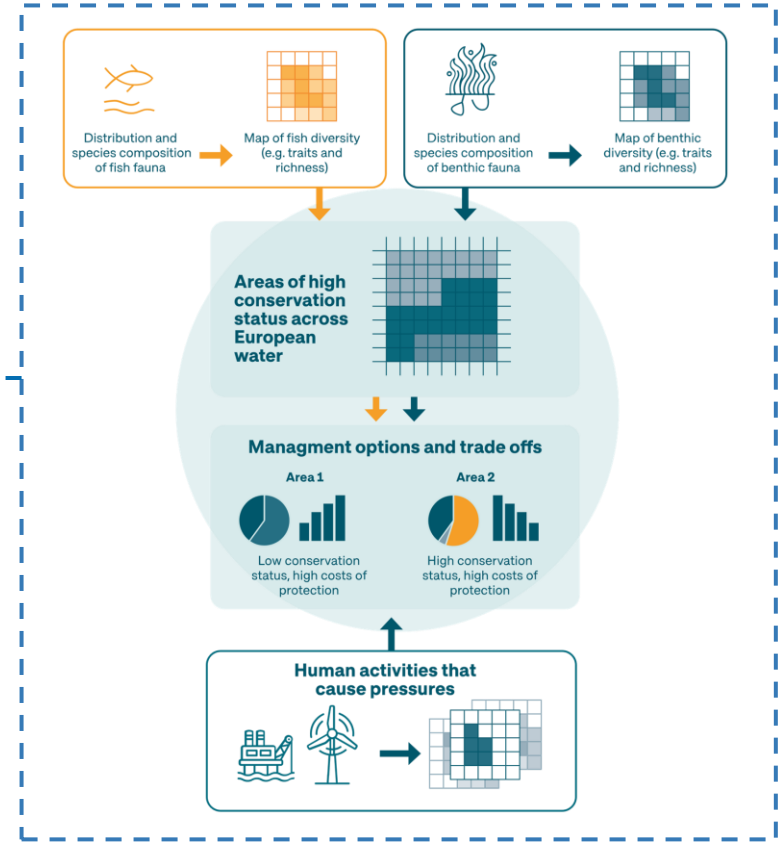
Main deliverable type: Data (data sets, microdata)		
Type of data	Characteristics	Possible formats
Chemical and physicochemical data	Concentration of CECs and legacy contaminants in different matrices, physicochemical properties such as temperature, salinity, oxygen, suspended particle matter (SPM), dissolved organic matter (DOM)	Recommended: .csv or .tab Accepted: .xls or .xlsx
Biological Data	Information regarding living organisms, encompassing genomics (other omics data), physiological parameters, and other related aspects, e.g. field, lab, mesocosm. Biological effects data and biodiversity (eDNA, proteomics, visual ID data)	Recommended: .csv or .tab Accepted: .xls or .xlsx
Computational Data	Data produced by computer simulations, modelling, algorithms, or software applications, e.g., model outputs	Recommended: NetCDF

* To ensure harmonised data management, guidelines and templates for structuring data and metadata will be provided throughout the CONTRAST project. These resources will enhance the findability, accessibility, interoperability, and reusability of the data, in alignment with FAIR

Any other information

- **COORDINATED BY** Technical University of Denmark (DTU, Martin Lindegren).
- **AIM OF THE PROJECT:** Operationalize **biodiversity** and **ecosystem services** indicators through end-users co-developed decision-support tool.
- **TYPE OF BIOLOGICAL DATA COLLECTED:** benthic, demersal and pelagic fish and invertebrate communities' data.

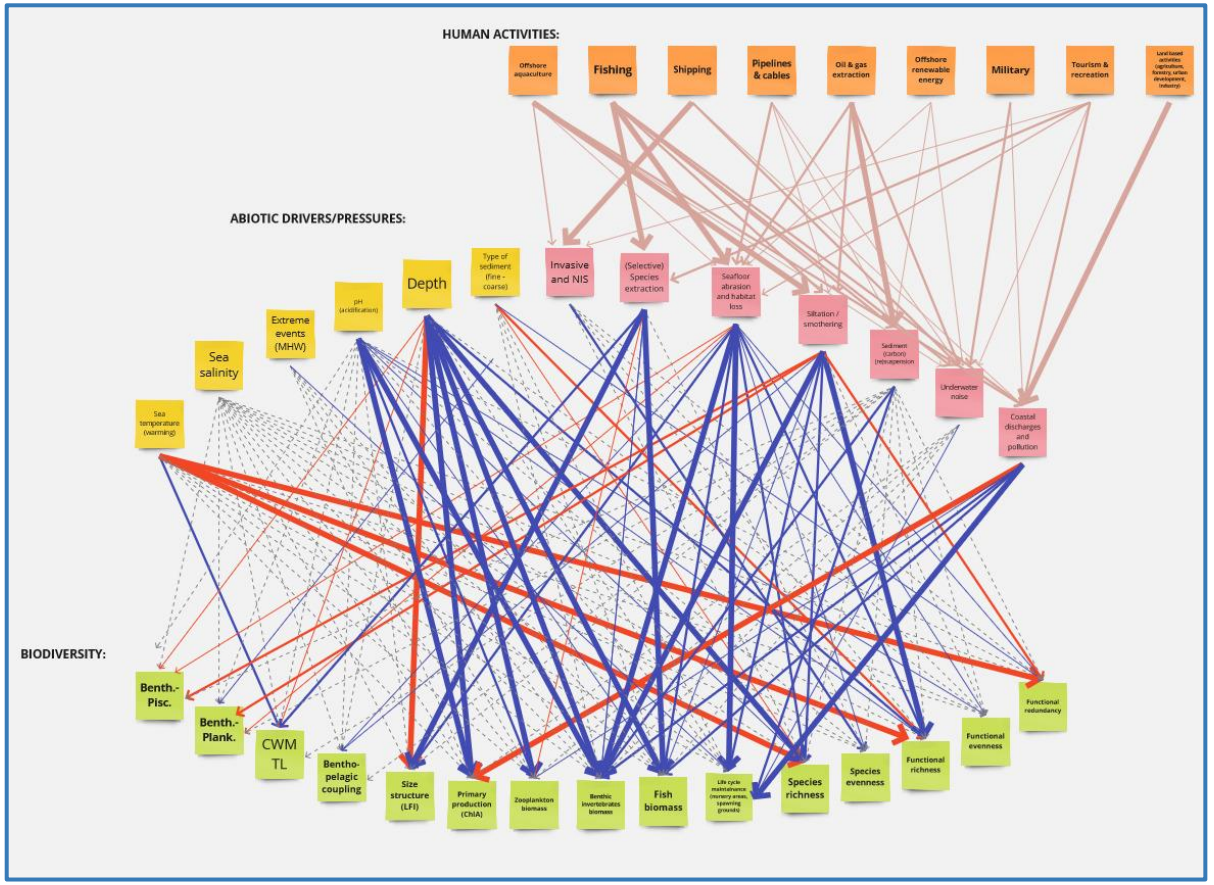
FOR EUROPEAN REGIONAL SEAS (Mediterranean, NE Atlantic, Barents and Norwegian seas, North sea) IDENTIFY AREAS OF PARTICULAR INTEREST IN TERMS OF BIODIVERSITY AND ECOSYSTEM SERVICES AND PROVIDE USEFUL TOOLS FOR CONSERVATION AND MANAGEMENT.



CONNECTION TO EMODnet:

- **At project level, use and generation of data: climatological, biological, etc**

- **At WP level: Links Between Ecosystem Services, Biodiversity And Pressures.**



Project name: Improving marine habitat status by considering ecosystem dynamics (MARHAB)

Project coordinator: Institute of Marine Research, Norway

Presented by: Even Moland (coordinator)

Type of biological data collected: Coastal fish assemblages (BRUVs), marine animal movement (acoustic telemetry, satellite telemetry), genetics, function (predator presence and size structure)

Geographical area: Skagerrak and Kattegat

MarHab
Horizon Europe 2024–2027

MarHab
Horizon Europe 2024–2027

Top predators

Mesopredators and grazers

Habitats

Trophic links essential to habitat health

MARHAB actions

Evaluate conservation status along protection-effectiveness gradients

Define *Favourable Conservation Status* and develop monitoring

Assess shortcomings of marine protection

Communicate solutions

Reconcile marine conservation with fisheries management

Healthy, resilient and productive ecosystems

INSTITUTE OF MARINE RESEARCH

Data availability, management, connection with EMODnet (access restrictions, standard protocols for data management in place, awareness of standards):

All data will be shared and made available according to the project DMP. Telemetry data will be shared with European tracking network.

Any other information: Follow MARHAB project on LinkedIn and visit marhab.eu

MarHab
Horizon Europe 2024-2027

Biology

Project pitches - Part 3

GES4SEAS

Biodiversity Meets Data - BMD

PROTECT BALTIC

DIGI4ECO

EMODnet Biology Workshop

19th and 20th November 2024, Online

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.

Project name: Achieving Good Environmental Status for maintaining ecosystem services, by assessing integrated impacts of cumulative pressures (GES4SEAS) www.ges4seas.eu 2022-2026

Project coordinator: Angel Borja (AZTI)

Aim of the project: Guide marine governance processes on minimizing **human pressures** and their impacts on coastal and **marine biodiversity** and ecosystem functioning, while maintaining the **sustainable delivery of ecosystem services**, through the development of an innovative **toolbox**, tested, validated and demonstrated, in the context of an adaptive **ecosystem-based management**

Type of biological data collected: No new data collected, but data on human activities, pressures and ecosystem components collated from open data sources and the 20 partners, and presented as GIS layers, plus additional data from local monitoring

Geographical area: From 11 Learnings Sites: 9 covering the 4 regional seas, 1 in French Polynesia and 1 in whole Europe

Data availability, management, connection with EMODnet (access restrictions, standard protocols for data management in place, awareness of standards,…)

- Look at the updated data management plan (Deliverable 5.5): https://www.ges4seas.eu/resources/#2024
- Data will be available in supplementary material of papers, and in our Zenodo repository: https://zenodo.org/communities/ges4seas/records?q=&l=list&p=1&s=10&sort=newest

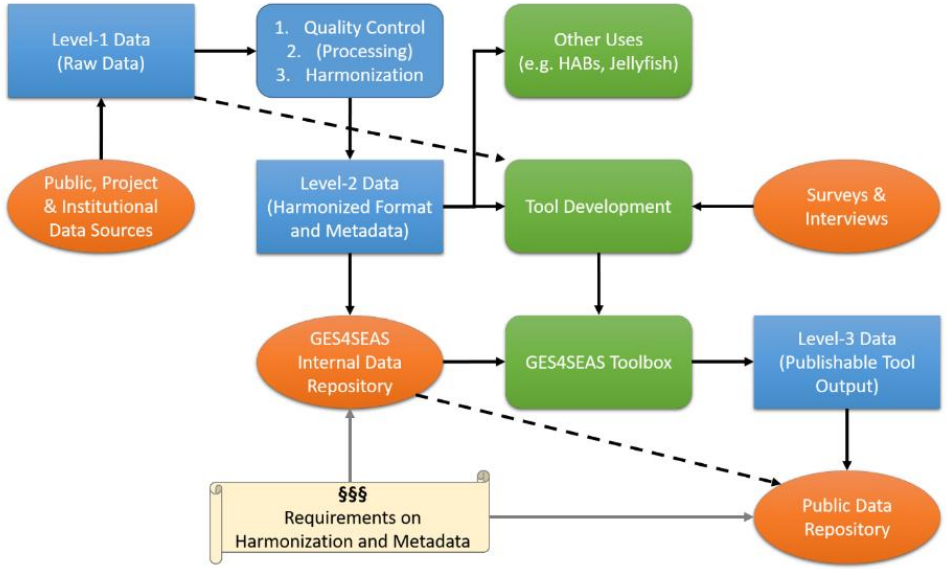


Figure 1. Schematic illustration of data-flow (black arrows) within the GES4SEAS project. Dashed arrows symbolize partial data-flows. HABs: Harmful Algal Blooms.

Any other information:

- Borja, A., 2023. A dataset and template for assessing the ecological status of marine sediments and waters, based on microbial taxa. Gigabyte, 2023: 10.46471/gigabyte.46486. A list of 1,974 microbial taxa with ecological groups to assess the status
- https://ambi.azti.es A list of 11,950 taxa with ecological groups to assess benthic status. Provided to WoRMs
- Borja et al., submitted. Marine Biodiversity and Environmental Data: An AI-Ready, Open Dataset from the long term (1995–2023) Basque Country Monitoring Network. *Frontiers in Ocean Sustainability.* Includes water, sediment, biota, phytoplankton, macroalgae, macroinvertebrates and fish from estuaries and coasts

Project name: BMD Biodiversity Meets Data

EMODnet European Marine Observation and Data Network

Project coordinator: Niels Raes

Presented by: Niels Raes

Aim of the project: Develop a Single Access Point to biodiversity data and analyses tools for managers of natural resources and policy makers across Europe

Type of biological data collected: eDNA / High-throughput biodiversity monitoring data

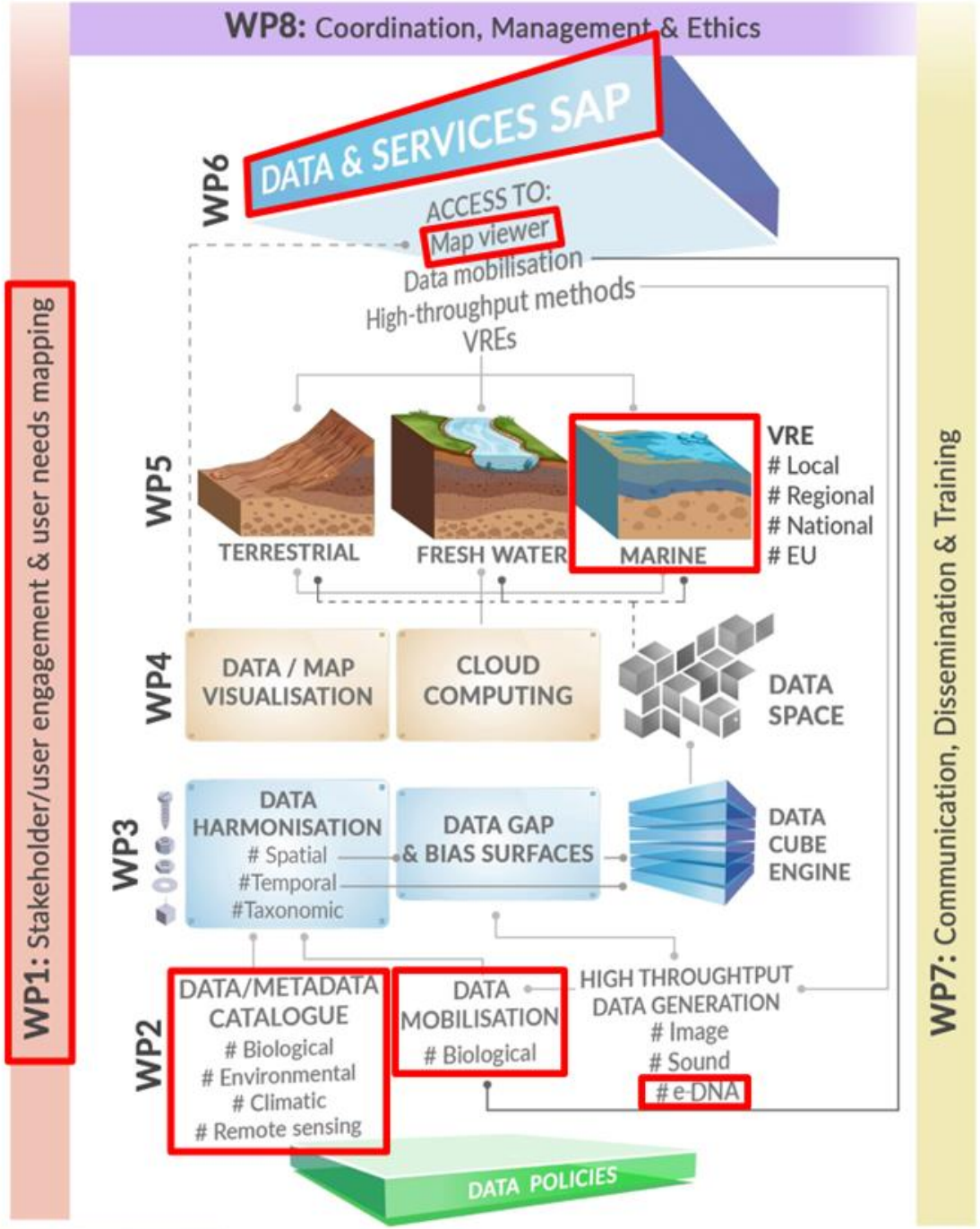
Geographical area: Europe/Global

BMD
Biodiversity Meets Data

Niels Raes –
niels.raes@naturalis.nl

BMD is a **Single Access Point (SAP)** that provides managers of natural resources and policy makers with access to:

- **High-throughput biodiversity monitoring** tools including image and sound capture, and e-DNA sampling combined with **AI taxon identification services**;
- A suite of co-designed **Virtual Research Environments (VREs)** for the terrestrial, freshwater and **marine** realms, including tools for:
  a) **biodiversity monitoring**,
  b) identification and analysis of **drivers of change**, and
  c) projections of **land cover and climate change** on the distribution of species and habitats.
- **Web-GIS Map viewer** to explore the data and VRE results.
- Stakeholder/user engagement & user needs mapping at all stages of the SAP development (WP1).
- **Start March 2025.**



Project name: PROTECT BALTIC (Enabling comprehensive effective and efficient protection and restoration measures for a resilient Baltic Sea ecosystem)

Project coordinator: HELCOM (Baltic Marine Environment Protection Commission)

Presented by: Kimmo Koivumäki, PROTECT data manager, HELCOM

Aim of the project: To enable sufficient spatial protection and restoration of the marine environment to secure positive biodiversity outcomes, ensuring that the function of the ecosystem is maintained, thus securing production of ecosystems services and enabling sustainable use both short and long term.

Type of biological data collected: Species observations, species distribution models (benthic species, birds, fish, decapod crustaceans), sedimentation, turbidity

Geographical area: Baltic Sea and its surrounding countries

Data availability

- Majority of the data is open. Restricted data can be used for modelling and/or analysis.

Spatial data mainly published in HELCOM MADS and biodiversity database, metadata in Metadata catalogue

Data management

- Data Management Plan defines used processes.
- Data is stored in Microsoft SharePoint with limited access for project staff. Original datasets are in HELCOM's internal drive which is restricted for secretariat's data personnel.

Data connection with EMODnet

- There are three datasets collected from EMODnet:
    - *Sedimentation accumulation rate*
    - *Sediment maps*
    - Total suspended matter used as a proxy for *turbidity*

Digital Twin-sustained 4D ecological monitoring of restoration in fishery depleted areas (DIGI4ECO)

Project Pitches
EMODNet Biology Workshop
November 20th, 2024

Drs J. Aguzzi (ICM-CSIC) & M. Clavel-Henry (GEOMAR)

Funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting authority can be held responsible for them.

M1 Digital Twin of Ocean (DTO) **simulates marine biological and environmental components** to understand ecosystems' past and present state and make predictions about their future

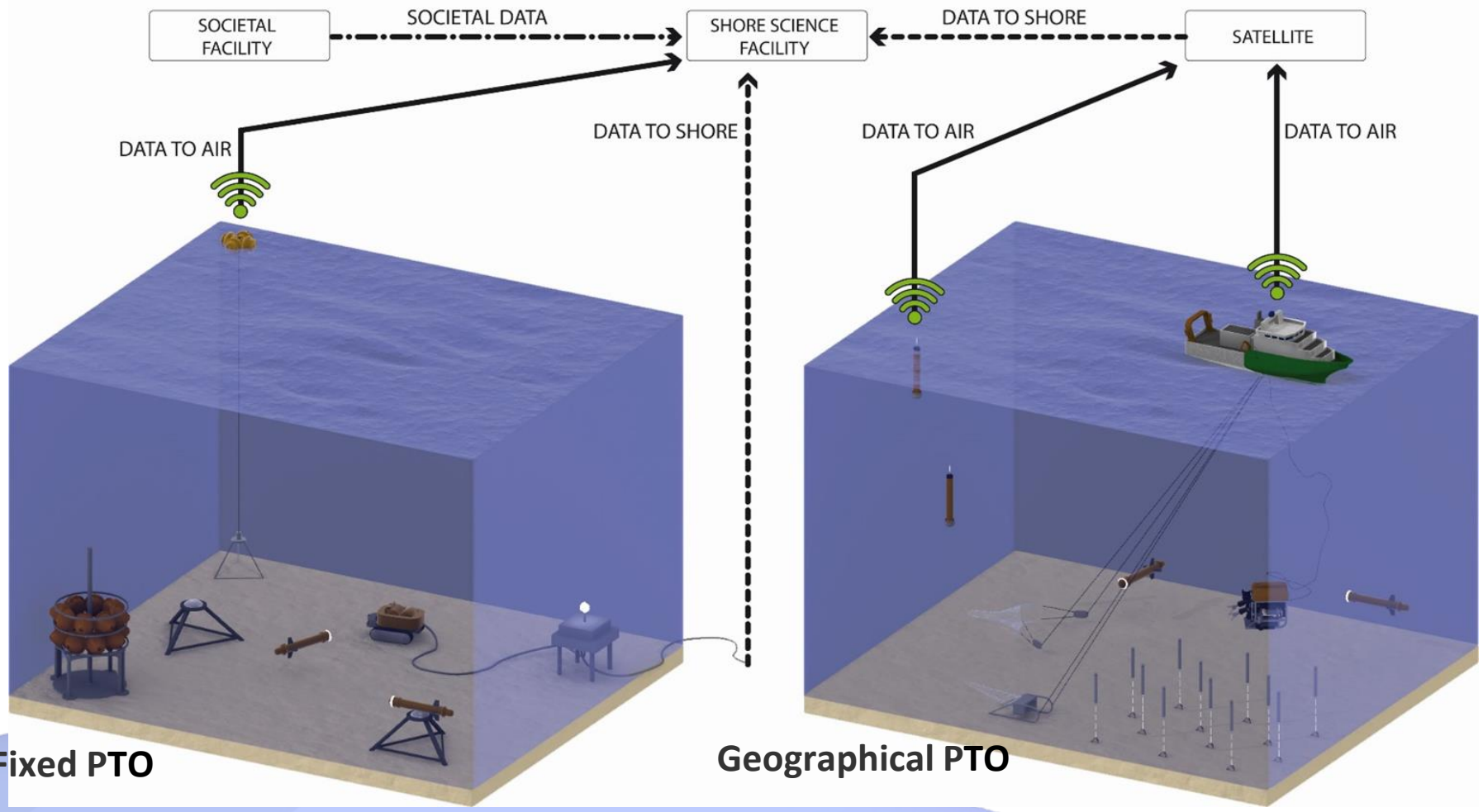
M2 In spite of the vastness of marine ecosystems, any approach to management requires the definition of **strategic areas** where to **repeat** measurements

M3 Therefore, DTO need to identify this space and include in its virtualization, the tools for in-situ data collection

Obj. Implement robotic networks for **demo-missions** delivering real-time data to be **merged with historical and sleeping data**, allowing comparisons for spatial scaling and temporal modelling

Physical Twin of Ocean (PTO)

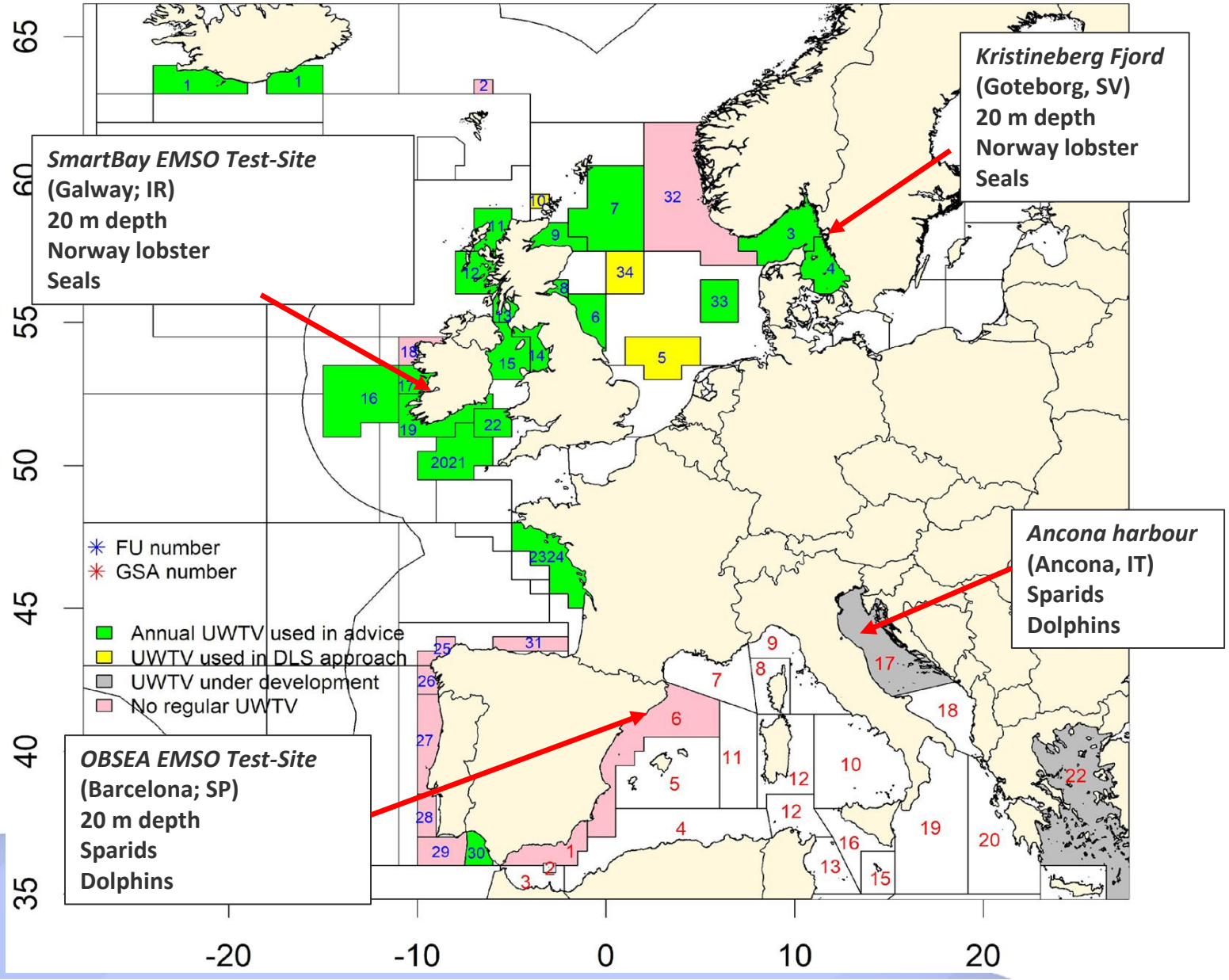
A PTO describes the physical system that allows the multiparametric data collection in the real-world



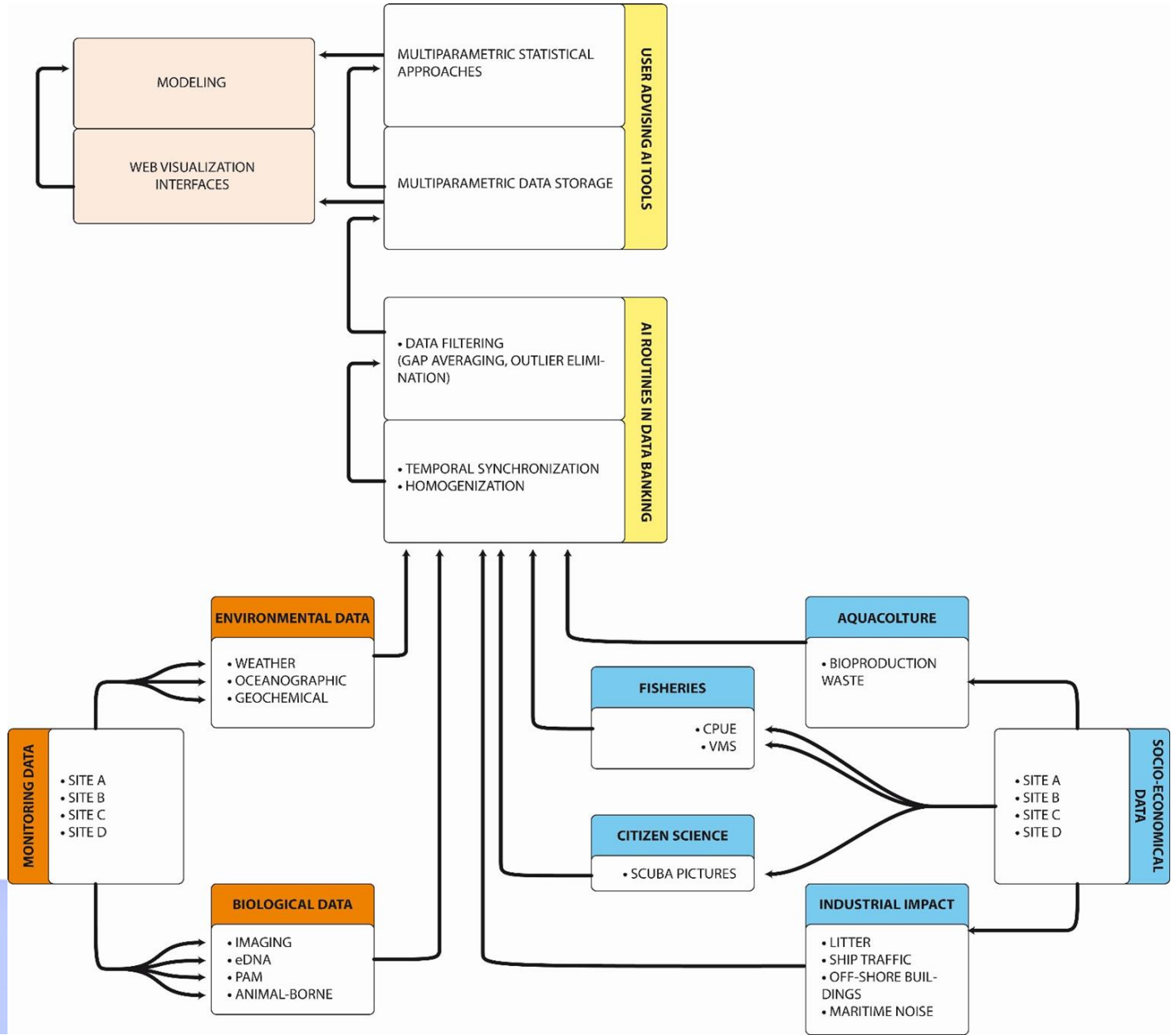
Fixed PTO

Geographical PTO

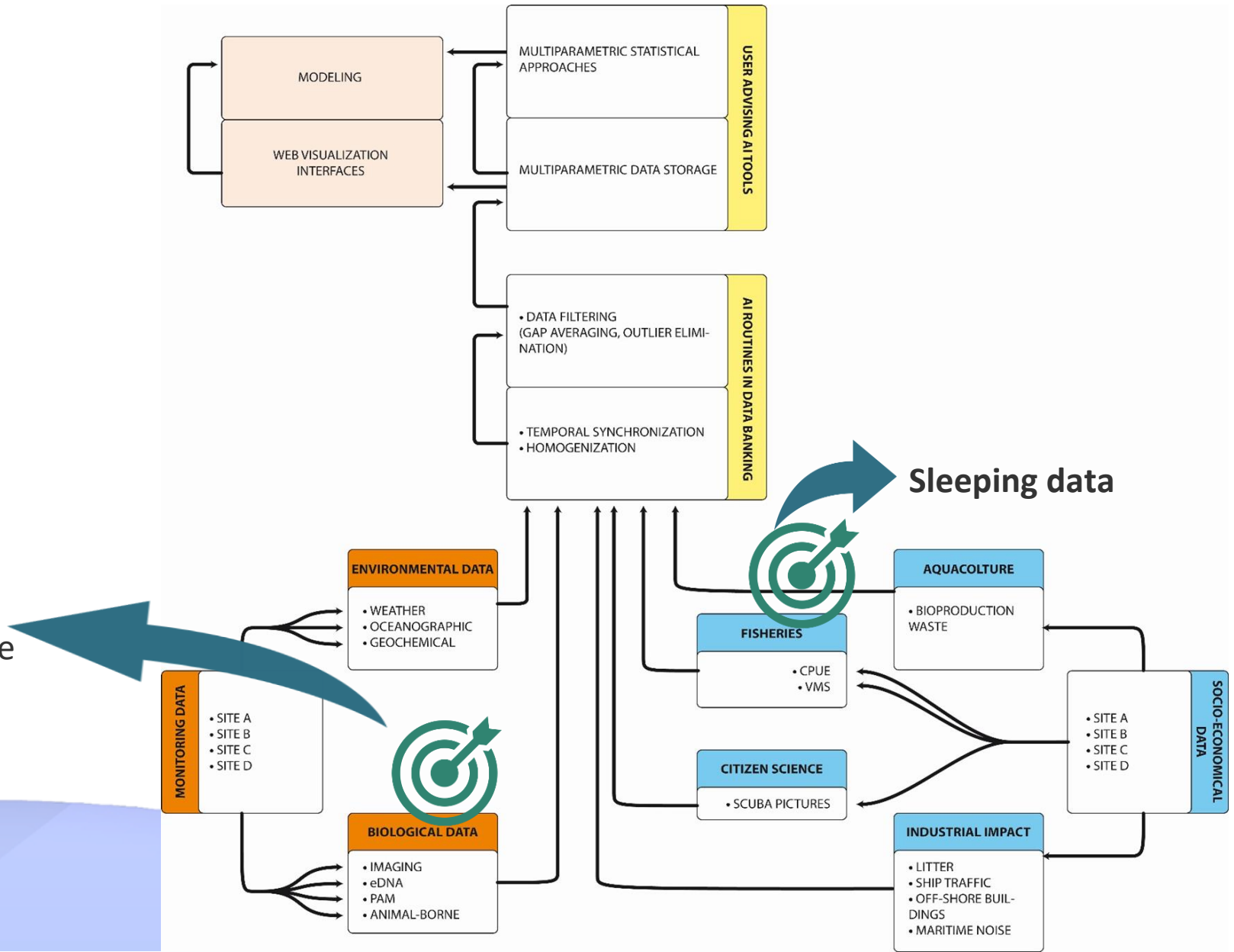
Four strategic areas for fixed PTOs



DTO architecture (Information flow treatment)

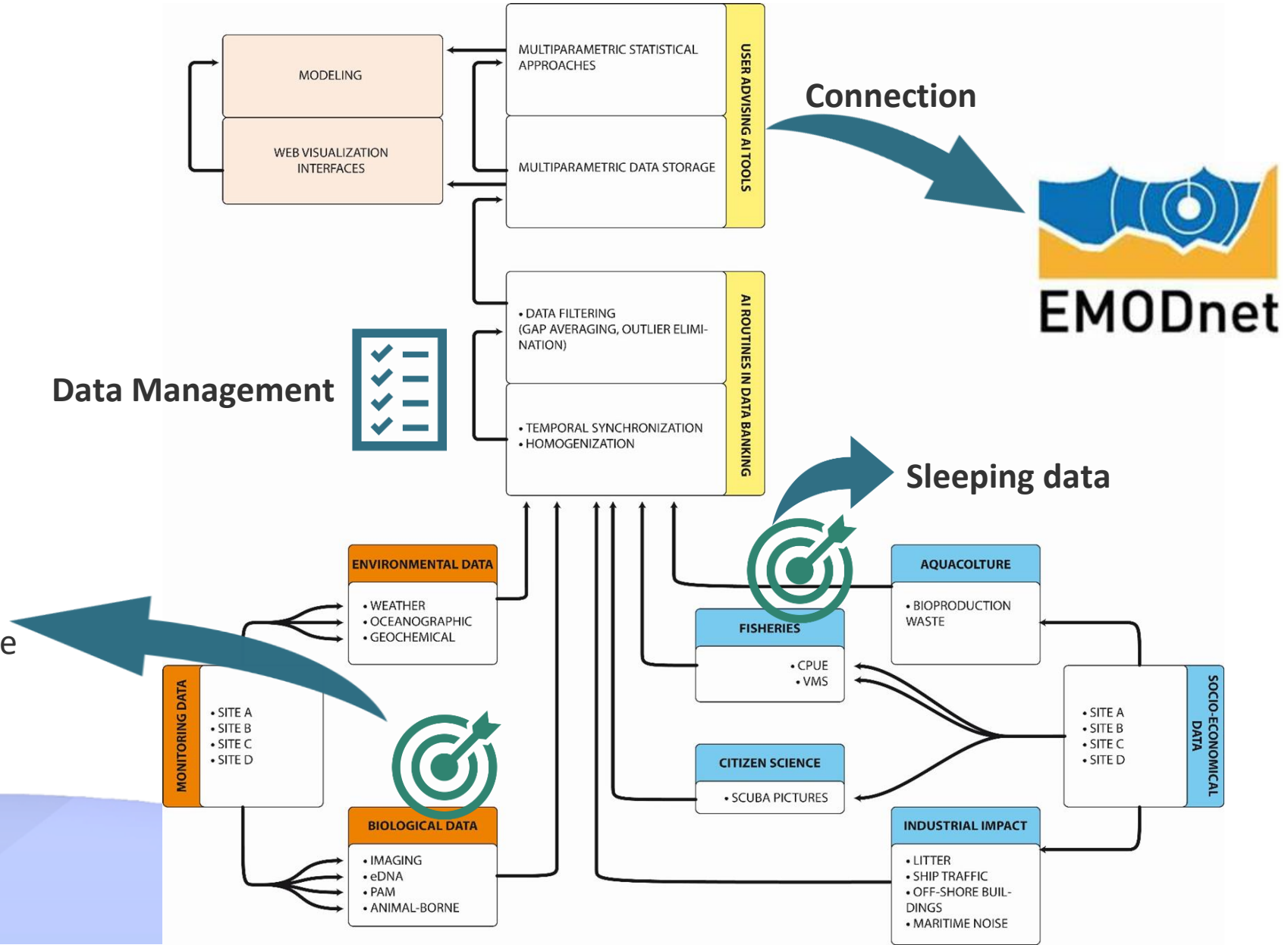


DTO architecture (Information flow treatment)



Detection and time series:
Imaging/eDNA/PAM/animal-borne

DTO architecture (Information flow treatment)



Detection and time series:
Imaging/eDNA/PAM/animal-borne

Pitch summary

Project name: Digital Twin-sustained 4D ecological monitoring of restoration in fishery depleted areas (DIGI4ECO)

Project coordinator: J. Aguzzi

Presented by: J. Aguzzi (ICM-CSIC) and M. Clavel-Henry (GEOMAR)

Aim of the project:
Implement robotic networks for demo-missions delivering real-time data to be merged with historical and sleeping data, allowing comparisons for spatial scaling and temporal modelling

Type of biological data collected:
Imaging, eDNA, PAM, oceanographic and geochemical data, socioeconomical and sleeping entries (e.g., recreational commercial catches)

Geographical area:
Western and Eastern Mediterranean, North Sea and Eastern Atlantic

Pitch summary

Data availability, management, connection with EMODnet

➤ Data Availability:

- I**mage data sets** for time series of species counts (megafauna; manual and AI processing)
- **PAM and eDNA** for augmented species detectability;
- Sleeping and socio-economic data

➤ Management:

- **Raw data storage** at each demo-site, processed for harmonization and stored/**managed** by ICATMAR (Spain) central bank repository;
- **Data quality** based on metadata labelling is provided.

➤ Connection with EMODnet:

- The project can create **a continuous pipeline of biological data** into EMODnet based on the expertise of UPC/MI (which are already data providers).

Biology

EMODnet Biology: Data submission guidance

Ruben Perez Perez, Flanders Marine Institute (VLIZ)
Lynn Delgat, Flanders Marine Institute (VLIZ)

2024-11-20 | EMODnet Biology online workshop

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.

EMODnet Biology data submission guidance

Data flows

Darwin Core

DNA derived data

Publishing data

Quality Control

EMODnet
European Marine Observation and Data Network

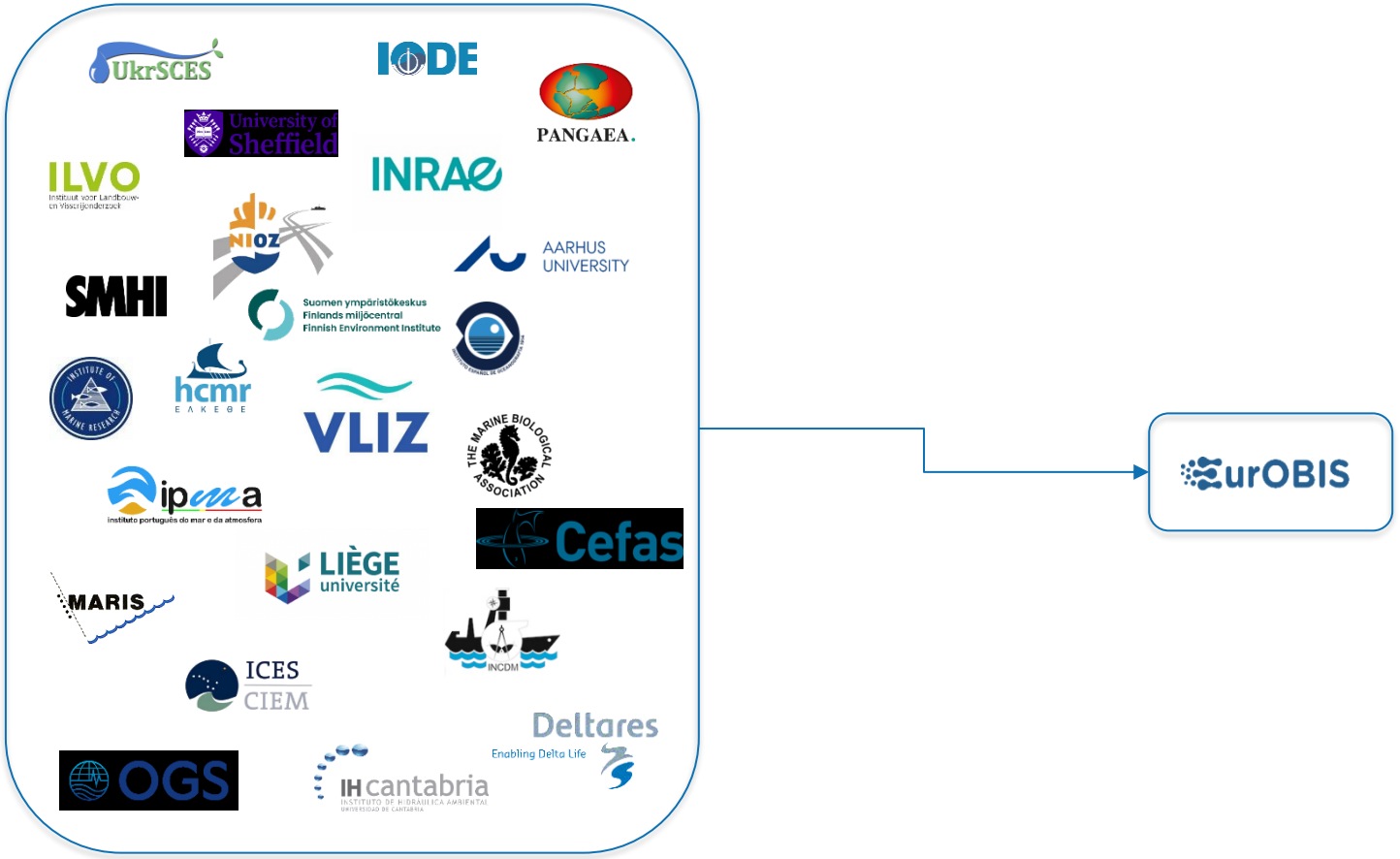
EMODnet Biology. The content

- 1400+ datasets
- 40M+ occurrence records
- 100M+ measurements or facts

Images copyright: from top to bottom: i) VLIZ (Nick Decombel); ii) Lotte Pohl; iii) VLIZ Lifewatch Belgium; iv) John Hill; v) VLIZ (Sven Van Haelst).

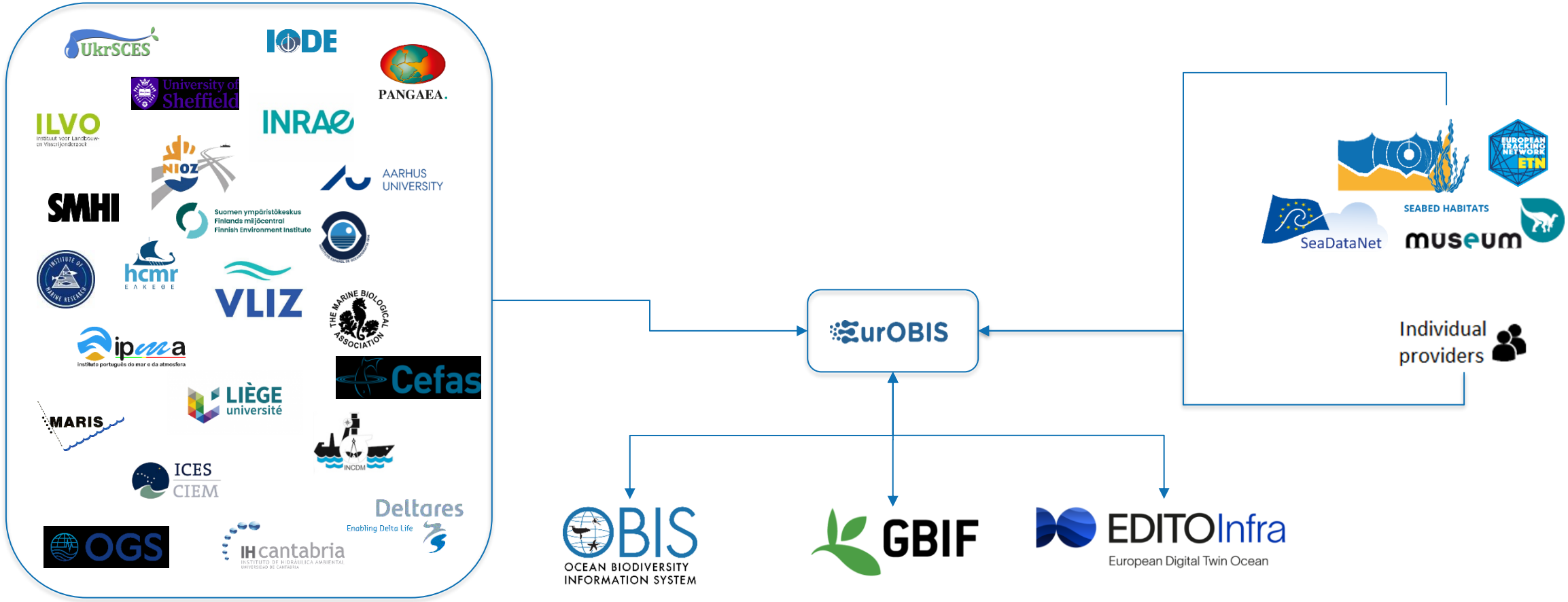
EMODnet Biology. The data mobiliser

- Data flow powered by EMODnet Biology



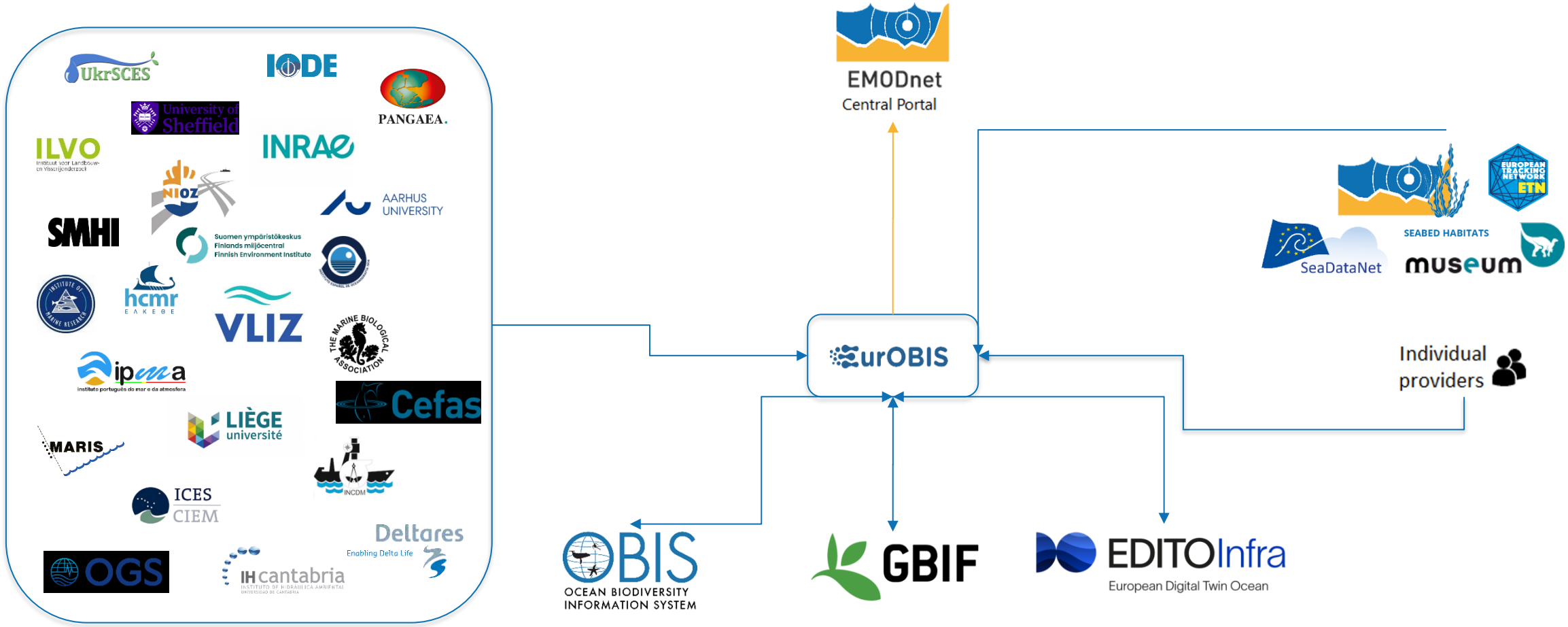
EMODnet Biology. The data mobiliser

- Data flow powered by EMODnet Biology



EMODnet Biology. The data mobiliser

- Data flow powered by EMODnet Biology



EMODnet Biology. The method

(Meta)data standards

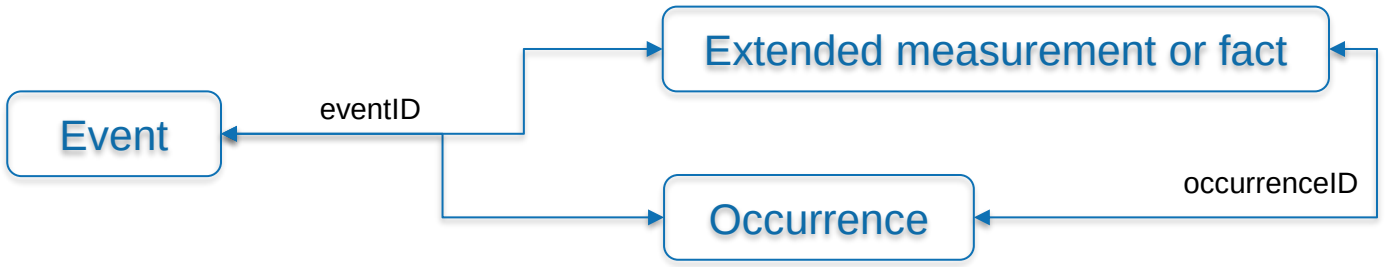
Ecological Metadata Language

Darwin Core

OBIS ENV- Data

Controlled vocabularies

- WoRMS – taxonomy
- Marine Regions – geography
- BODC - parameters



- Organisms quantifications, facts and biometrics
- Sampling methodologies
- Environmental data

EMODnet Biology data submission guidance

Data flows

Darwin Core

DNA derived data

Publishing data

Quality Control

EMODnet
European Marine Observation and Data Network

Darwin Core: The scope

Darwin Core is a standard maintained by the Darwin Core Maintenance Interest Group. It includes a glossary of terms (in other contexts these might be called properties, elements, fields, columns, attributes, or concepts) 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.

Darwin Core: The terms

List of terms

- **Standardized**
- **Maintained**

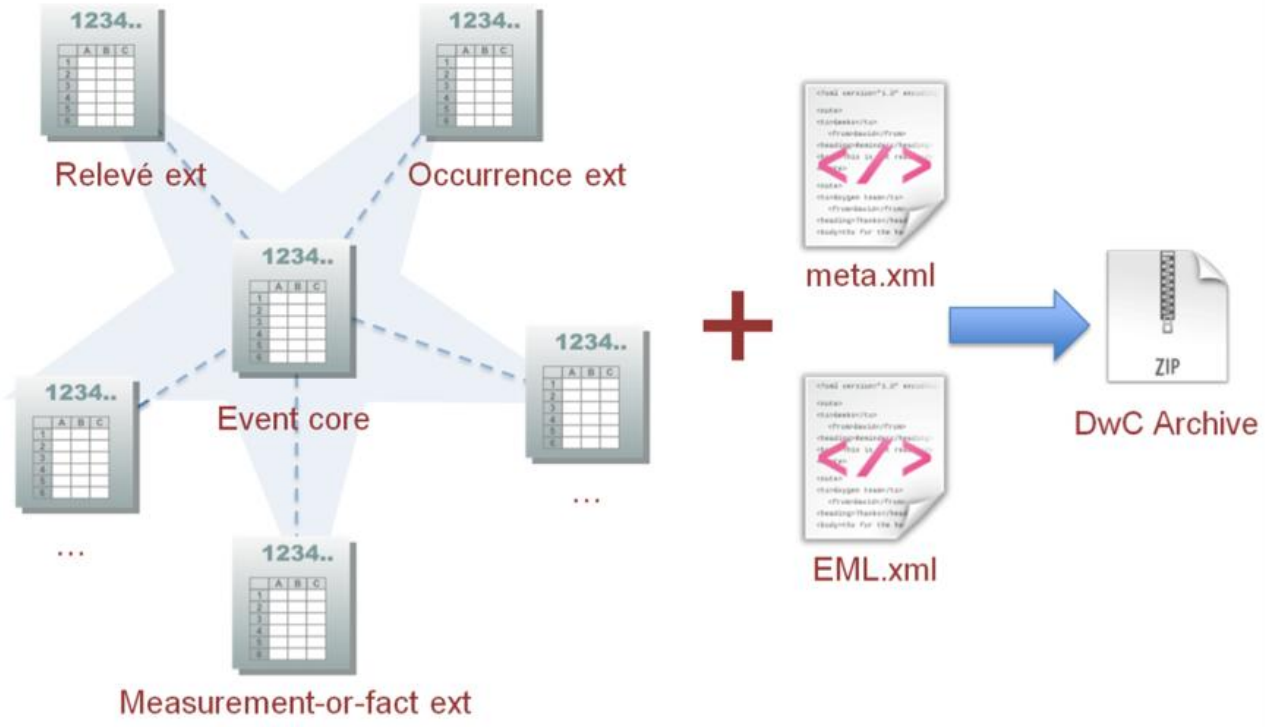
occurrenceID | catalogNumber | recordNumber | recordedBy | recordedByID | individualCount | organismQuantity | organismQuantityType | sex | lifeStage | reproductiveCondition | caste | behavior | vitality | establishmentMeans | degreeOfEstablishment | pathway | georeferenceVerificationStatus | occurrenceStatus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

basisOfRecord	
Identifier	http://rs.tdwg.org/dwc/terms/basisOfRecord
Definition	The specific nature of the data record.
Comments	Recommended best practice is to use a controlled vocabulary such as the set of local names of the identifiers for classes in Darwin Core.

Examples	`MaterialEntity`
	`PreservedSpecimen`
	`FossilSpecimen`

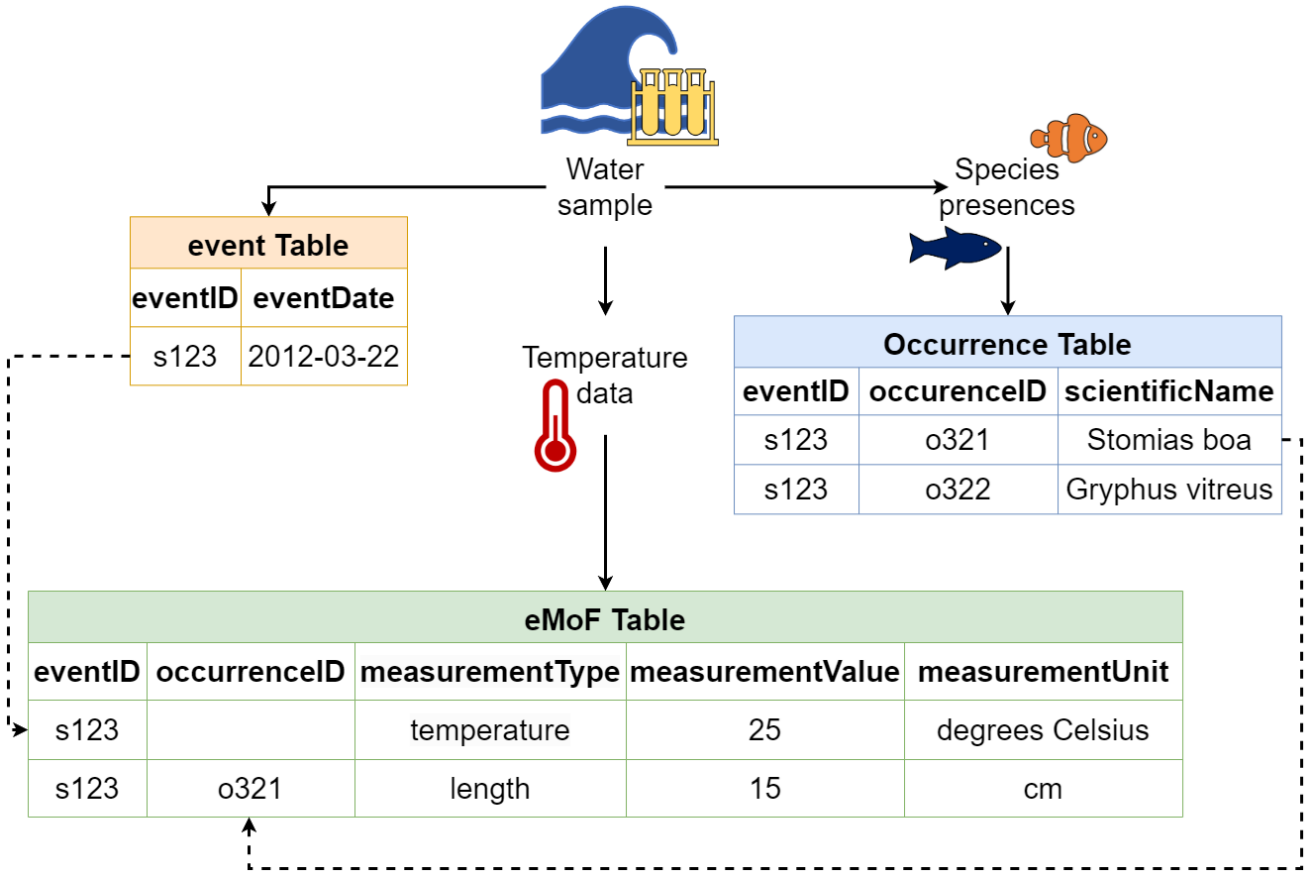
Darwin Core Archives: The data package

- Data tables
  - Core
  - Extensions
- Meta.xml
- Eml.xml



Darwin Core Archives: The data package

- Data tables
  - Core
  - Extensions
- Meta.xml
- Eml.xml



Darwin Core Archives: The data package

- Data tables
  - Core
  - Extensions
- **Meta.xml**
- Eml.xml

```
<archive xmlns="http://rs.tdwg.org/dwc/text/" metadata="eml.xml">
  <core encoding="UTF-8" fieldsTerminatedBy="\t" linesTerminatedBy="\n" fieldsEnclosedBy="">
    <files>
      <location>occurrence.txt</location>
    </files>
    <coreid index="0" />
    <field index="1" term="http://rs.tdwg.org/dwc/terms/basisOfRecord"/>
    <field index="2" term="http://rs.tdwg.org/dwc/terms/occurrenceID"/>
    <field index="3" term="http://rs.tdwg.org/dwc/terms/occurrenceRemarks"/>
    <field index="4" term="http://rs.tdwg.org/dwc/terms/recordedBy"/>
    <field index="5" term="http://rs.tdwg.org/dwc/terms/behavior"/>
    <field index="6" term="http://rs.tdwg.org/dwc/terms/occurrenceStatus"/>
    <field index="7" term="http://rs.tdwg.org/dwc/terms/previousIdentifications"/>
    <field index="8" term="http://rs.tdwg.org/dwc/terms/eventID"/>
    <field index="9" term="http://rs.tdwg.org/dwc/terms/identifiedBy"/>
    <field index="10" term="http://rs.tdwg.org/dwc/terms/identificationQualifier"/>
    <field index="11" term="http://rs.tdwg.org/dwc/terms/scientificNameID"/>
    <field index="12" term="http://rs.tdwg.org/dwc/terms/scientificName"/>
  </core>
  <extension encoding="UTF-8" fieldsTerminatedBy="\t" linesTerminatedBy="\n" fieldsEnclosedl
    <files>
      <location>extendedmeasurementorfact.txt</location>
    </files>
    <coreid index="0" />
    <field index="1" term="http://rs.tdwg.org/dwc/terms/measurementID"/>
    <field index="2" term="http://rs.tdwg.org/dwc/terms/occurrenceID"/>
    <field index="3" term="http://rs.tdwg.org/dwc/terms/measurementType"/>
    <field index="4" term="http://rs.iobis.org/obis/terms/measurementTypeID"/>
    <field index="5" term="http://rs.tdwg.org/dwc/terms/measurementValue"/>
    <field index="6" term="http://rs.iobis.org/obis/terms/measurementValueID"/>
    <field index="7" term="http://rs.tdwg.org/dwc/terms/measurementUnit"/>
    <field index="8" term="http://rs.iobis.org/obis/terms/measurementUnitID"/>
  </extension>
</archive>
```

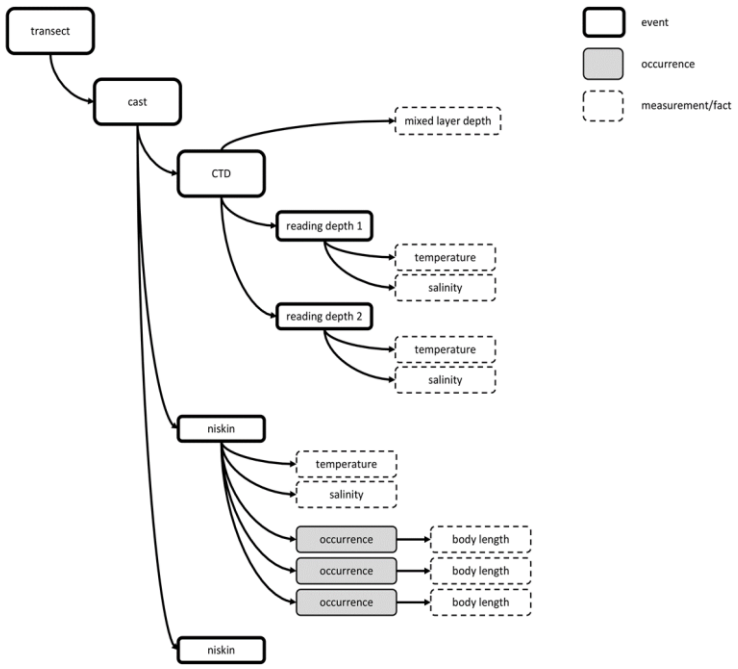
Darwin Core Archives: The data package

- Data tables
  - Core
  - Extensions
- Meta.xml
- **Eml.xml**

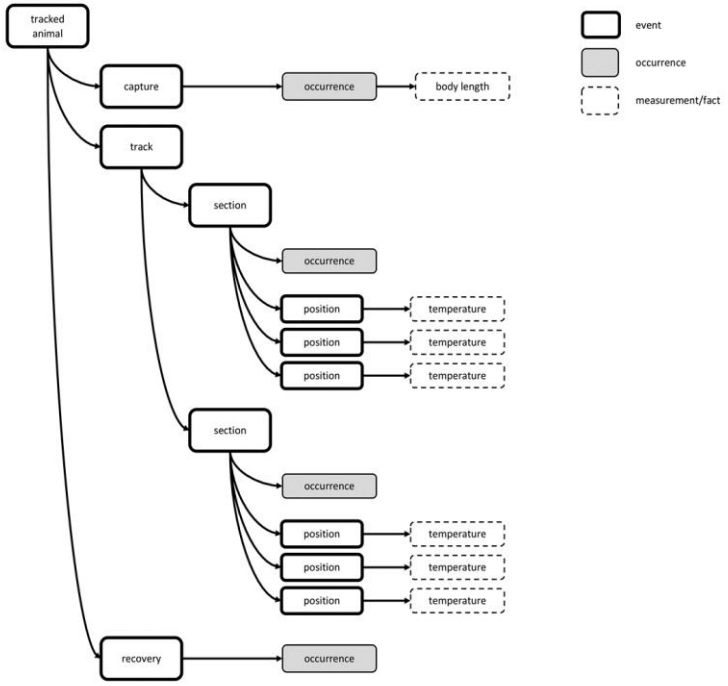
```
<eml:eml xmlns:eml="eml://ecoinformatics.org/eml-2.1.1"
         xmlns:dc="http://purl.org/dc/terms/"
         xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
         xsi:schemaLocation="eml://ecoinformatics.org/eml-2.1.1 http://rs.gbif.org/schema/eml-gbif-profile/1.1
         packageId="cd4f3685-c3dd-4461-894c-b4e94c17585f/v1.3" system="http://gbif.org" scope="system"
         xml:lang="eng">

<dataset>
  <alternateIdentifier>10.17031/tpgjrv</alternateIdentifier>
  <alternateIdentifier>cd4f3685-c3dd-4461-894c-b4e94c17585f</alternateIdentifier>
  <alternateIdentifier>https://www.dassh.ac.uk/ipt/resource?r=dassh-113</alternateIdentifier>
  <title xml:lang="eng">1999-2001 University Marine Biological Station Millport (UMBSM) Clyde Sea Rapid Method
    <creator>
    <organizationName>University Marine Biological Station Millport (UMBSM)</organizationName>
    </creator>
    <metadataProvider>
    <organizationName>The archive for marine species and habitats data (DASSH)</organizationName>
    <address>
      <city>Plymouth</city>
      <country>GB</country>
    </address>
    <electronicMailAddress>dassh.enquiries@mba.ac.uk</electronicMailAddress>
    </metadataProvider>
    <associatedParty>
    <individualName>
      <givenName>Data</givenName>
    <surName>Team</surName>
    </individualName>
    <electronicMailAddress>data@mba.ac.uk</electronicMailAddress>
    <role>user</role>
    </associatedParty>
```

Darwin Core Archives: Choosing the structure

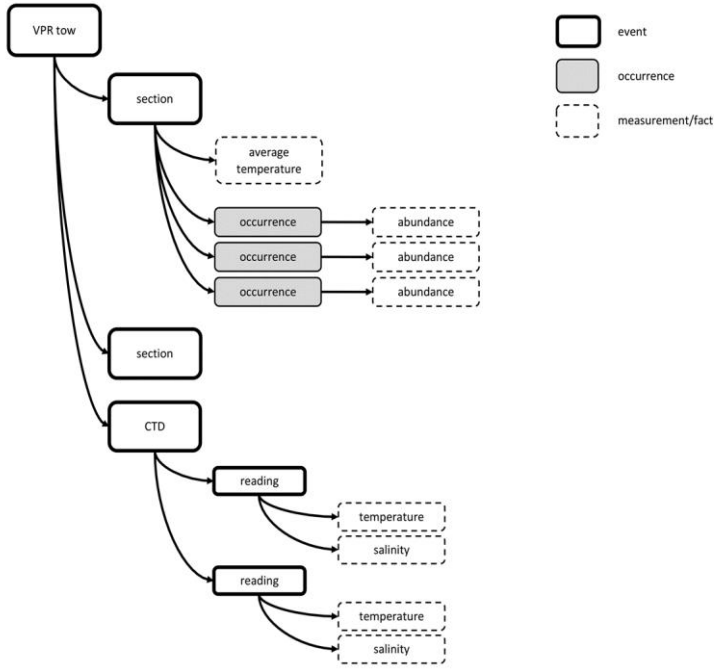
CTD measurements



Tagged animals



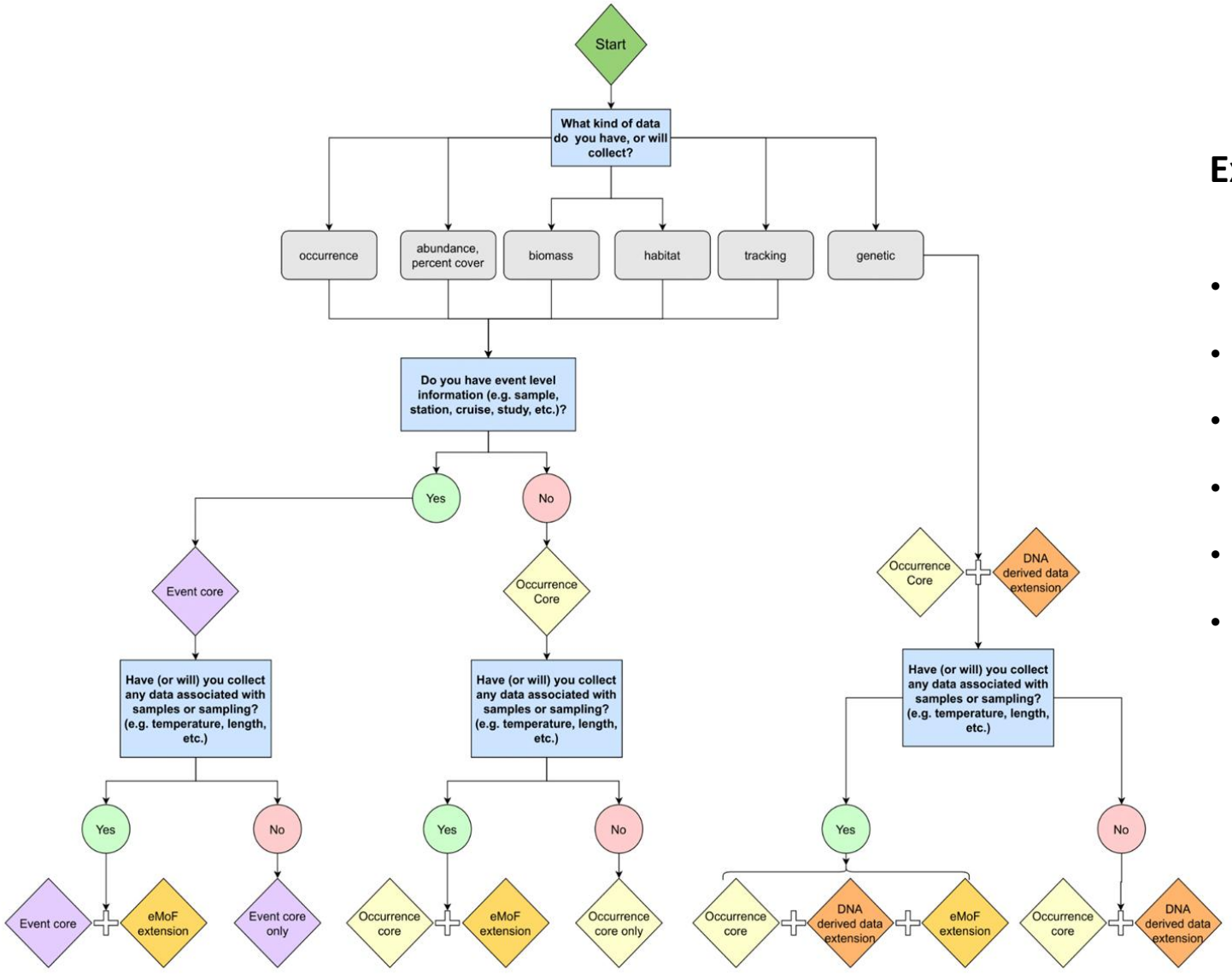
VPR



Darwin Core Archives: Choosing the structure

Cores

- Event
- Occurrence
- Checklist



Extensions

- Occurrence
- Extended Measurement Or Facts
- Resource Relationships
- Chronometric Age
- DNA derived data
- ...

OBIS ENV Data format

Event

eventDate	locality	datasetNa	institution	eventID	decimalLo	decimalLa	coordinat	maximum	minimumDepthInMe
1992-11-07	cadiz bay	Data surve	University	biom_cb_071	-6.26667	36.53333	10000	40	30
1992-11-11	cabo de ga	Data surve	University	biom_cg_111	-2.18333	36.73333	9300	15	5

extended measurements or facts

eventID	occurrenc	measurem	measurem	measurem	measurem	measurem	measurem
biom_cb_071		SamplingP	Agassiz tra	not applic	http://voc	https://vo	https://vo
biom_cg_111		SamplingP	diver	not applic	http://voc	https://vo	https://vo
biom_cb_071	biom_cb_(	Lifestage	adult	not applic	http://voc	https://vo	https://vo
biom_cb_071	biom_cb_(	Lifestage	juvenile	not applic	http://voc	https://vo	https://vo
biom_cb_071	biom_cb_(	Lifestage	juvenile	not applic	http://voc	https://vo	https://vo
biom_cb_071	biom_cb_(	Lifestage	adult	not applic	http://voc	https://vo	https://vo
biom_cg_111	biom_cg_1	Lifestage	adult	not applic	http://voc	https://vo	https://vo
biom_cg_111	biom_cg_1	Lifestage	adult	not applic	http://voc	https://vo	https://vo
biom_cb_071	biom_cb_(	Wet weigh	10.4	kg/m2	http://voc		https://vo
biom_cb_071	biom_cb_(	Wet weigh	3.6	kg/m2	http://voc		https://vo
biom_cb_071	biom_cb_(	Wet weigh	8.43	kg/m2	http://voc		https://vo
biom_cb_071	biom_cb_(	Wet weigh	15.22	kg/m2	http://voc		https://vo

Occurrence

scientificNam	eventID	occurrenc	scientificNameID	occurrenc	basisOfRe
Mustelus aste	biom_cb_(	biom_cb_(	urn:lsid:marinespeci	present	MaterialSa
Mustelus aste	biom_cb_(	biom_cb_(	urn:lsid:marinespeci	present	MaterialSa
Mustelus can	biom_cb_(	biom_cb_(	urn:lsid:marinespeci	present	MaterialSa
Carcharhinus	biom_cb_(	biom_cb_(	urn:lsid:marinespeci	present	MaterialSa
Exaptasia pal	biom_cg_1	biom_cg_1	urn:lsid:marinespeci	present	HumanOb
Breviolum mi	biom_cg_1	biom_cg_1	urn:lsid:marinespeci	present	HumanOb

Event:

An **action** that occurs at a particular **place** and **time**.

Occurrence:

An existence of an **organism**(or **homogeneous** group of organisms) at a particular **place** and **time**

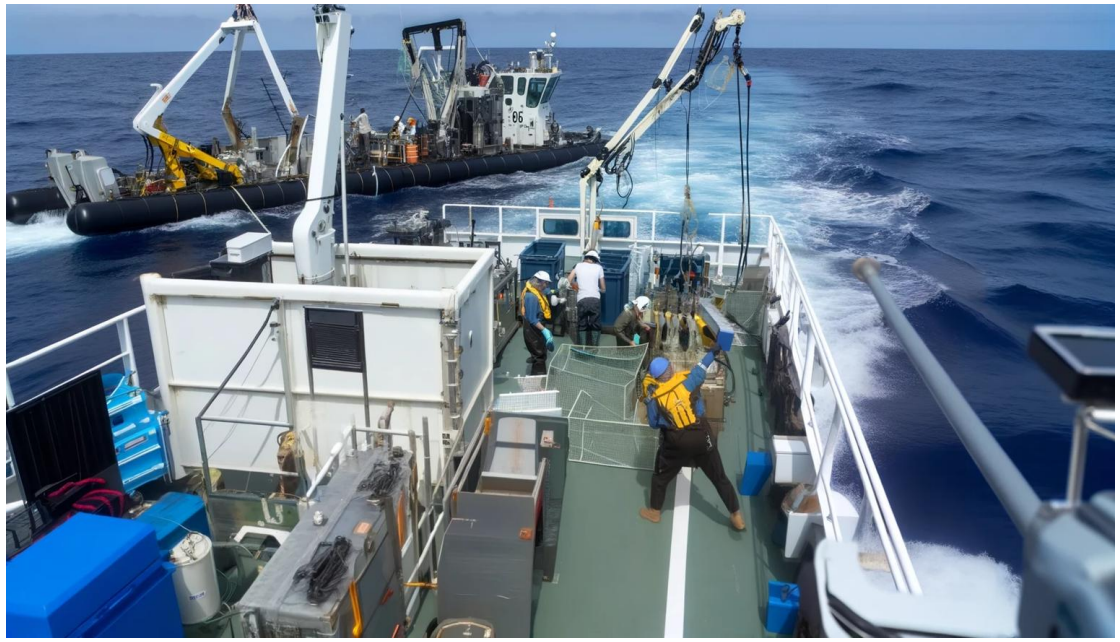
Measurement or Fact:

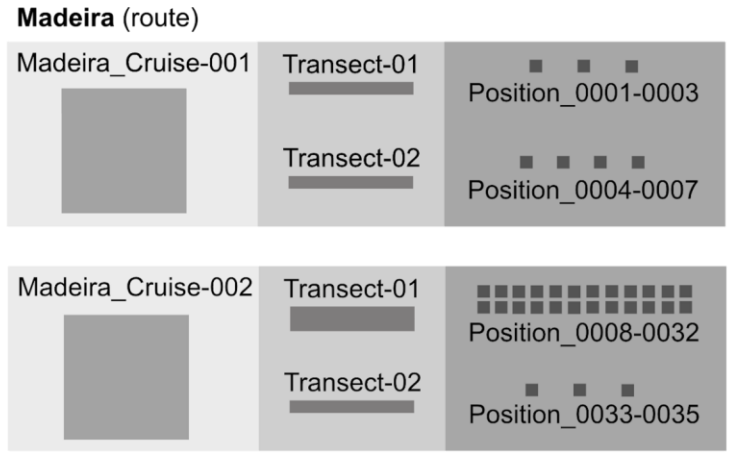
A known **characteristic** of **something**

OBIS ENV Data format

Event

eventDate	locality	datasetNa	institution	eventID	decimalLo	decimalLa	coordinat	maximum	minimumDepthInMe
1992-11-07	cadiz bay	Data surve	University	biom_cb_071	-6.26667	36.53333	10000	40	30
1992-11-11	cabo de ga	Data surve	University	biom_cg_111	-2.18333	36.73333	9300	15	5





Event:

An **action** that occurs at a particular **place** and **time**.

Occurrence:

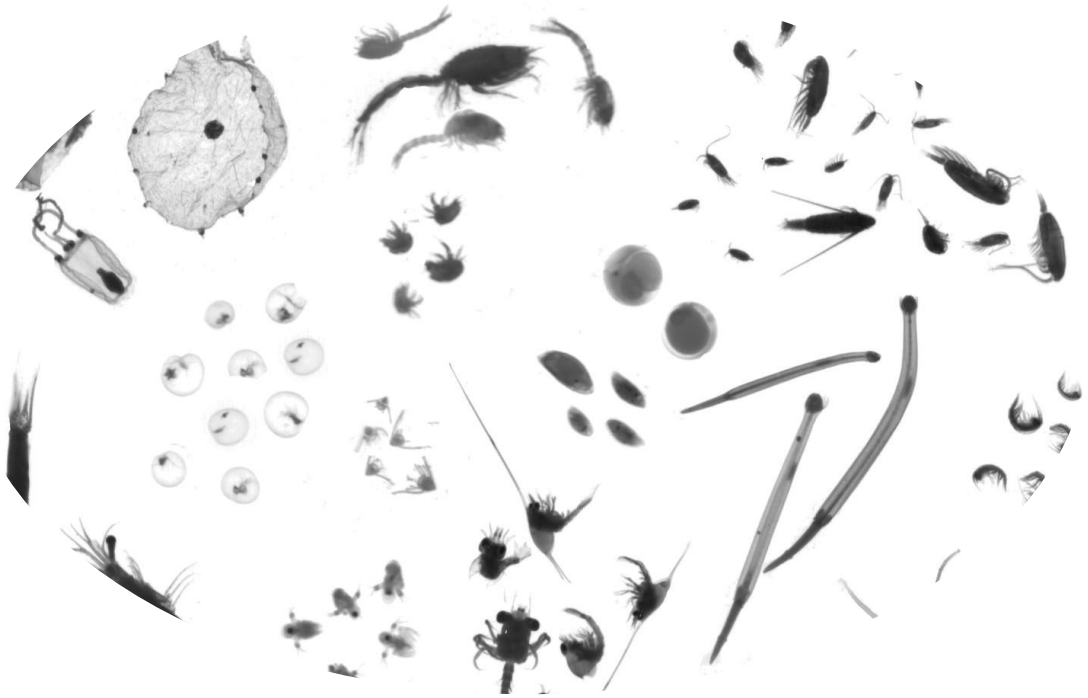
An existence of an **organism**(or **homogeneous** group of organisms) at a particular **place** and **time**

Measurement or Fact:

A known **characteristic** of **something**

Image source: DALL-E 2014-04-11

OBIS ENV Data format



- Taxonomy
- Size class
- Sex
- Life stage
- Behaviour
- …

Occurrence

scientificNam	eventID	occurrenc	scientificNameID	occurrenc	basisOfRe
Mustelus aste	biom_cb_	biom_cb_	urn:lsid:marinespeci	present	MaterialSa
Mustelus aste	biom_cb_	biom_cb_	urn:lsid:marinespeci	present	MaterialSa
Mustelus can	biom_cb_	biom_cb_	urn:lsid:marinespeci	present	MaterialSa
Carcharhinus	biom_cb_	biom_cb_	urn:lsid:marinespeci	present	MaterialSa
Exaptasia pal	biom_cg_	biom_cg_	urn:lsid:marinespeci	present	HumanOb
Breviolum mi	biom_cg_	biom_cg_	urn:lsid:marinespeci	present	HumanOb

Event:

An **action** that occurs at a particular **place** and **time**.

Occurrence:

An existence of an **organism**(or **homogeneous** group of organisms) at a particular **place** and **time**

Measurement or Fact:

A known **characteristic** of **something**

Image source: LifeWatch VLIZ

OBIS ENV Data format

Organism quantifications

Abiotic measurements

Species biometrics

Sampling methodologies

Organism characteristics

Habitat descriptions

extended measurements or facts

eventID	occurrenc	measurem	measurem	measurem	measurem	measurem	measurem
biom_cb_071		SamplingP	Agassiz tra	not applic	http://voc	https://vo	https://vo
biom_cg_111		SamplingP	diver	not applic	http://voc	https://vo	https://vo
biom_cb_071	biom_cb_(	Lifestage	adult	not applic	http://voc	https://vo	https://vo
biom_cb_071	biom_cb_(	Lifestage	juvenile	not applic	http://voc	https://vo	https://vo
biom_cb_071	biom_cb_(	Lifestage	juvenile	not applic	http://voc	https://vo	https://vo
biom_cb_071	biom_cb_(	Lifestage	adult	not applic	http://voc	https://vo	https://vo
biom_cg_111	biom_cg_1	Lifestage	adult	not applic	http://voc	https://vo	https://vo
biom_cg_111	biom_cg_1	Lifestage	adult	not applic	http://voc	https://vo	https://vo
biom_cb_071	biom_cb_(	Wet weigh	10.4	kg/m2	http://voc		https://vo
biom_cb_071	biom_cb_(	Wet weigh	3.6	kg/m2	http://voc		https://vo
biom_cb_071	biom_cb_(	Wet weigh	8.43	kg/m2	http://voc		https://vo
biom_cb_071	biom_cb_(	Wet weigh	15.22	kg/m2	http://voc		https://vo

Event:

An **action** that occurs at a particular **place** and **time**.

Occurrence:

An existence of an **organism**(or **homogeneous** group of organisms) at a particular **place** and **time**

Measurement or Fact:

A known **characteristic** of **something**

Image source: LifeWatch VLIZ

EMODnet
European Marine Observation and Data Network

Practical example

Data transformation Example

spec name	date	station	depth (m)	gear	wwb (kg/m2)	lifestage
M. asterias	7/11/1992	cadiz bay	30-40	Agassiz trawl	10.4	ad
M. asterias	7/11/1992	cadiz bay	30-40	Agassiz trawl	3.6	juv
M. canis	7/11/1992	cadiz bay	30-40	Agassiz trawl	8.43	juv
Carcharhinus albimarginatus	7/11/1992	cadiz bay	30-40	Agassiz trawl	15.22	ad
Exaptasia pallida	11/11/1992	cabo de gata	5 to 15	diver		adult
Breviolum minutum	11/11/1992	cabo de gata	5 to 15	diver		adult

Darwin Core fields

scientificN	eventDate	locality		SamplingP		Lifestage	datasetNa	institution	eventID	occurrenc	scientificN	decimalLo	decimalLa	coordinat	maximum	minimuml	occurrenc	basisOfRe
spec name	date	station	depth (m)	gear	wwb (kg/r	lifestage												
M. asteria	7/11/1992	cadiz bay	30-40	Agassiz tra	10.4	ad												
M. asteria	7/11/1992	cadiz bay	30-40	Agassiz tra	3.6	juv												
M. canis	7/11/1992	cadiz bay	30-40	Agassiz tra	8.43	juv												
Carcharhir	7/11/1992	cadiz bay	30-40	Agassiz tra	15.22	ad												
Exaptasia	11/11/1992	cabo de ga	5 to 15	diver	-	adult												
Breviolum	11/11/1992	cabo de ga	5 to 15	diver	-	adult												

- **Fields mapping** ●
- **Mandatory fields** ●

- occurrenceID
- scientificName
- scientificNameID
- occurrenceStatus
- basisOfRecord
- maximumDepthInMeters
- minimumDepthInMeters

- eventID
- eventDate
- decimalLatitude
- decimalLongitude
- coordinateUncertaintyInMeters
- datasetName
- institutionCode

https://manual.obis.org/formatting.html#data-formatting-tools
https://dwc.tdwg.org/terms/
https://manual.obis.org/checklist.html#darwin-core-term-checklist-for-obis

Standardising the content

scientificN	eventDate	locality		SamplingP		Lifestage	datasetNa	institution	eventID	occurrenc	scientificN	decimalLo	decimalLa	coordinate	maximum	minimumI	occurrenc	basisOfRe
			depth (m)		wwb (kg/r													
M. asteria	1992-11-07	cadiz bay	30-40	Agassiz tra	10.4	adult	Data surve	University	biom_cb_	biom_cb_					40	30	present	MaterialS
M. asteria	1992-11-07	cadiz bay	30-40	Agassiz tra	3.6	juvenile	Data surve	University	biom_cb_	biom_cb_					40	30	present	MaterialS
M. canis	1992-11-07	cadiz bay	30-40	Agassiz tra	8.43	juvenile	Data surve	University	biom_cb_	biom_cb_					40	30	present	MaterialS
Carcharhi	1992-11-07	cadiz bay	30-40	Agassiz tra	15.22	adult	Data surve	University	biom_cb_	biom_cb_					40	30	present	MaterialS
Exaptasia	1992-11-11	cabo de ga	5 to 15	diver		adult	Data surve	University	biom_cg_	biom_cg_					15	5	present	HumanOb
Breviolum	1992-11-11	cabo de ga	5 to 15	diver		adult	Data surve	University	biom_cg_	biom_cg_					15	5	present	HumanOb

- **Standardise**
- **Enhance**

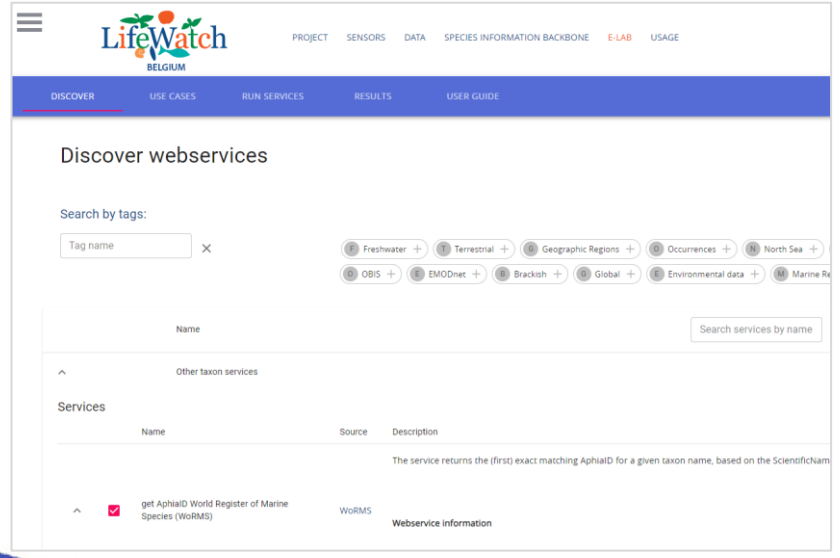
Standardising the content

scientificN	eventDate	locality		SamplingP		Lifestage	datasetNa	institution	eventID	occurrenc	scientificN	decimalLo	decimalLa	coordinate	maximum	minimumI	occurrenc	basisOfRe
			depth (m)		wwb (kg/r													
Mustelus a	1992-11-07	cadiz bay	30-40	Agassiz tra	10.4	adult	Data surve	University	biom_cb_(	biom_cb_	urn:lsid:m	-6.26667	36.53333	10000	40	30	present	MaterialSa
Mustelus a	1992-11-07	cadiz bay	30-40	Agassiz tra	3.6	juvenile	Data surve	University	biom_cb_(	biom_cb_	urn:lsid:m	-6.26667	36.53333	10000	40	30	present	MaterialSa
Mustelus (	1992-11-07	cadiz bay	30-40	Agassiz tra	8.43	juvenile	Data surve	University	biom_cb_(	biom_cb_	urn:lsid:m	-6.26667	36.53333	10000	40	30	present	MaterialSa
Carcharhi	1992-11-07	cadiz bay	30-40	Agassiz tra	15.22	adult	Data surve	University	biom_cb_(	biom_cb_	urn:lsid:m	-6.26667	36.53333	10000	40	30	present	MaterialSa
Exaptasia	1992-11-11	cabo de ga	5 to 15	diver		adult	Data surve	University	biom_cg_1	biom_cg_	urn:lsid:m	-2.18333	36.73333	9300	15	5	present	HumanOb
Breviolum	1992-11-11	cabo de ga	5 to 15	diver		adult	Data surve	University	biom_cg_1	biom_cg_	urn:lsid:m	-2.18333	36.73333	9300	15	5	present	HumanOb

- **Standardise**

WoRMS World Register of Marine Species

Marineregions.org



https://www.lifewatch.be/en/manual

Adopting the schema

Event

eventDate	locality	datasetNa	institution	eventID	decimalLo	decimalLa	coordinat	maximum	minimumDepthInMe
1992-11-07	cadiz bay	Data surve	University	biom_cb_071	-6.26667	36.53333	10000	40	30
1992-11-11	cabo de ga	Data surve	University	biom_cg_111	-2.18333	36.73333	9300	15	5

Occurrence

scientificNam	eventID	occurrenc	scientificNameID	occurrenc	basisOfRe
Mustelus aste	biom_cb_	biom_cb_	urn:lsid:marinespeci	present	MaterialSa
Mustelus aste	biom_cb_	biom_cb_	urn:lsid:marinespeci	present	MaterialSa
Mustelus can	biom_cb_	biom_cb_	urn:lsid:marinespeci	present	MaterialSa
Carcharhinus	biom_cb_	biom_cb_	urn:lsid:marinespeci	present	MaterialSa
Exaptasia pal	biom_cg_	biom_cg_	urn:lsid:marinespeci	present	HumanOb
Breviolum mi	biom_cg_	biom_cg_	urn:lsid:marinespeci	present	HumanOb

- **Separate tables**
- **Standardise**

National Oceanography Centre

extended measurements or facts

eventID	occurrenc	measuren	measuren	measuren	measuren	measuren	measuren
biom_cb_071		SamplingP	Agassiz tra	not applic	http://voc	https://vo	https://vo
biom_cg_111		SamplingP	diver	not applic	http://voc	https://vo	https://vo
biom_cb_071	biom_cb_	Lifestage	adult	not applic	http://voc	https://vo	https://vo
biom_cb_071	biom_cb_	Lifestage	juvenile	not applic	http://voc	https://vo	https://vo
biom_cb_071	biom_cb_	Lifestage	juvenile	not applic	http://voc	https://vo	https://vo
biom_cb_071	biom_cb_	Lifestage	adult	not applic	http://voc	https://vo	https://vo
biom_cg_111	biom_cg_	Lifestage	adult	not applic	http://voc	https://vo	https://vo
biom_cg_111	biom_cg_	Lifestage	adult	not applic	http://voc	https://vo	https://vo
biom_cb_071	biom_cb_	Wet weigh	10.4	kg/m2	http://voc		https://vo
biom_cb_071	biom_cb_	Wet weigh	3.6	kg/m2	http://voc		https://vo
biom_cb_071	biom_cb_	Wet weigh	8.43	kg/m2	http://voc		https://vo
biom_cb_071	biom_cb_	Wet weigh	15.22	kg/m2	http://voc		https://vo

https://vocab.nerc.ac.uk/search_nvs/

EMODnet Biology data submission guidance

EMODnet
European Marine Observation and Data Network

DNA derived data

- Information from molecular studies (metabarcoding, qPCR, metagenomics)
  - Raw information stored in sequence databases
  - Not only useful for phylogenetic research or molecular ecology
  - ‘Occurrences’ can be derived
    - Coordinates + date + taxonomic information
  - Submitting derived occurrences to a biodiversity database makes them useful in a broader context
  - Valuable source of information
    - Undescribed taxa
    - Inconspicuous / unobservable taxa

DNA derived data – Transformation to DwC

- Additional extension needed:
  - DNA derived data extension
    - Terms from:
      - MIxS
      - GGBN
      - MIQE
  - Occurence Core
    - Recent development: Event Core possible

DNA derived data – Transformation to DwC

- 1) Transform data tables

OTU/ASV table

	Sample 1	Sample 2	Sample 3	Sample 4	Sample 5
ASV 1	0	41	0	0	52
ASV 2	5	0	35	0	0
ASV 3	0	0	0	46	0
ASV 4	8	0	0	21	62
ASV 5	0	51	0	0	0
ASV 6	23	0	0	0	0
ASV 7	0	0	64	0	0
ASV 8	80	0	0	25	8
ASV 9	1	0	0	63	0
ASV 10	3	7	0	0	35

Sample table

Sample name	Date	Location name	Latitude	Longitude
Sample 1	13/11/2024	Site 1	51.25195	2.924496
Sample 2	13/11/2024	Site 2	51.24787	2.883469
Sample 3	13/11/2024	Site 3	51.23336	2.894284
Sample 4	14/11/2024	Site 1	51.25195	2.924496
Sample 5	15/11/2024	Site 2	51.24787	2.883469

Taxonomy table

	Species
ASV 1	Taxon A
ASV 2	Taxon B
ASV 3	Taxon B
ASV 4	Taxon C
ASV 5	Taxon D
ASV 6	Taxon E
ASV 7	Taxon F
ASV 8	Taxon G
ASV 9	Taxon G
ASV 10	Taxon H

FASTA/FASTQ file

- ASV/OTU and sample identifiers link the files
- Combining the tables → Derived occurrences

DNA derived data – Transformation to DwC

- 1) Transform data tables

OTU/ASV table

	Sample 1	Sample 2	Sample 3	Sample 4	Sample 5
ASV 1	0	41	0	0	52
ASV 2	5	0	35	0	0
ASV 3	0	0	0	46	0
ASV 4	8	0	0	21	62
ASV 5	0	51	0	0	0
ASV 6	23	0	0	0	0
ASV 7	0	0	64	0	0
ASV 8	80	0	0	25	8
ASV 9	1	0	0	63	0
ASV 10	3	7	0	0	35

Sample table

Sample name	Date	Location name	Latitude	Longitude
Sample 1	13/11/2024	Site 1	51.25195	2.924496
Sample 2	13/11/2024	Site 2	51.24787	2.883469
Sample 3	13/11/2024	Site 3	51.23336	2.894284
Sample 4	14/11/2024	Site 1	51.25195	2.924496
Sample 5	15/11/2024	Site 2	51.24787	2.883469

Taxonomy

	Species
ASV 1	Taxon A
ASV 2	Taxon B
ASV 3	Taxon B
ASV 4	Taxon C
ASV 5	Taxon D
ASV 6	Taxon E
ASV 7	Taxon F
ASV 8	Taxon G
ASV 9	Taxon G
ASV 10	Taxon H

FASTA/FASTQ file

```
>ASV 1
gtgggtctgaggcaccgccaagtcctttgggttttaagctggcgctcgtagtgcccaggcggatagctg
>ASV 2
gtgggtttgaagcaccgccaagtcctttgggtttcaagctaatgctcgtagtaccctggcggataagtt
>ASV 3
gtgggtttgaagcaccgccaagtcctttgggttttaagctaatgctcgtagtaccctggcggataaggg
>ASV 4
gtgggtctaaagcaccgccaagtcctttgggttttaagctggtgctcgtagttcccaggcggacagtgg
>ASV 5
gtgggtctaaagcaccgccaagtcctttgggttttaagctaatgcttgtagtacccgggcggataaaat
>ASV 6
gggggtctaaagcaccgccaagtcctttgggttttaggcaggttgcacgtaattcccaggcggatagcg
>ASV 7
ggggatctaaagtaccgccaagtcctttgggttttaagctagtgctcgtagtacccgggcggatagatg
>ASV 8
gggggtctaaagtaccgccaggtcctttgggttttaagctaatgctcgtagtacccgggcggataggtt
>ASV 9
ggggatctaaagtaccgccaagtcctttgggttttaagctaatgctcgtagtccccaggcggatagaga
>ASV 10
ggggatctaaagtaccgccaagtcctttgggtttcaagctgatgctcgtagtacccaggcggatagagg
```

- Combine tables
- Remove non-detections
- Map
- Split in Core and extension(s)
- Standardize

DwC Occurrence

occurrenceID	scientificName	organismQuantity	organismQuantityType	eventID	eventDate	decimalLatitude	decimalLongitude	locality
StudyID_Occ1	Taxon B	5	DNA sequence reads	Sample 1	2024-11-13	51.25195	2.924496	Site 1
StudyID_Occ2	Taxon C	8	DNA sequence reads	Sample 1	2024-11-13	51.25195	2.924496	Site 1
StudyID_Occ3	Taxon E	23	DNA sequence reads	Sample 1	2024-11-13	51.25195	2.924496	Site 1
StudyID_Occ4	Taxon G	80	DNA sequence reads	Sample 1	2024-11-13	51.25195	2.924496	Site 1
StudyID_Occ5	Taxon G	1	DNA sequence reads	Sample 1	2024-11-13	51.25195	2.924496	Site 1
StudyID_Occ6	Taxon H	3	DNA sequence reads	Sample 1	2024-11-13	51.25195	2.924496	Site 1
StudyID_Occ7	Taxon A	41	DNA sequence reads	Sample 2	2024-11-13	51.247867	2.883469	Site 2
StudyID_Occ8	Taxon D	51	DNA sequence reads	Sample 2	2024-11-13	51.247867	2.883469	Site 2
StudyID_Occ9	Taxon H	7	DNA sequence reads	Sample 2	2024-11-13	51.247867	2.883469	Site 2

DwC DNA extension

occurrenceID	DNA_sequence
StudyID_Occ1	gtgggtttgaagcaccgcca
StudyID_Occ2	gtgggtctaaagcaccgcc
StudyID_Occ3	gggggtctaaagcaccgcc
StudyID_Occ4	gggggtctaaagtaccgcc
StudyID_Occ5	ggggatctaaagtaccgcc
StudyID_Occ6	ggggatctaaagtaccgcc
StudyID_Occ7	gtgggtctgaggcaccgcc
StudyID_Occ8	gtgggtctaaagcaccgcc
StudyID_Occ9	ggggatctaaagtaccgcc

DNA derived data – Transformation to DwC

- 1) Transform data tables
- 2) Document additional (meta)data (enhance)
  - Same as other datasets (e.g. basisOfRecord, sampling protocol, etc.)
  - But also DNA specific information:
    - Link to raw sequences
    - Primers
    - Target locus
    - Total number of reads in sample
    - Bio-informatic methods (otu/asv generation / taxonomic annotation)
      - Software + version + parameters
      - Reference database
    - Taxonomic ID from other taxonomic database (eg. NCBI)
    - …
- Which info to which DwC term follow this guide: Publishing DNA-derived data through biodiversity data platforms

DNA derived data – Resources

- GBIF Guidelines: Publishing DNA-derived data through biodiversity data platforms
- OBIS Manual: DNA derived data
- EMODnet Biology DNA derived data guidelines: coming May 2025
- Ocean Teacher Biological Data Management Course: chapter on DNA derived data
- GBIF Metabarcoding Data Toolkit: Tool to convert metabarcoding datasets to DwC (pilot phase)

- Recent developments feedback on guidelines welcome

EMODnet Biology data submission guidance

Data flows

Darwin Core

DNA derived data

Publishing data

Quality Control

EMODnet
European Marine Observation and Data Network

EMODnet
European Marine Observation and Data Network

The Integrated Publishing Toolkit (IPT)

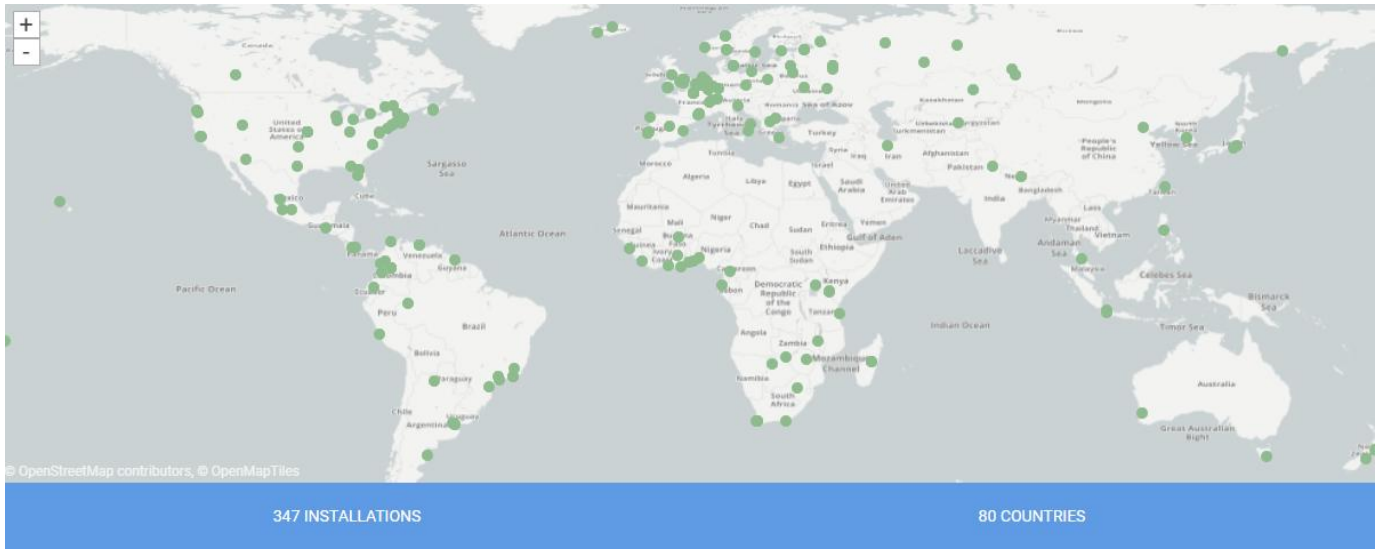
Concept

“The Integrated Publishing Toolkit (IPT) is a free open-source software used to create and manage distributed data repositories that share biodiversity data into the GBIF network”

Concept

- Data repository
- Creation of Darwin Core archive files
- Creation of data paper templates
- Automated data flow with OBIS/GBIF

- Limited to GBIF formats and scope



Publishing data

- Uploading data
- Mapping data to Darwin Core
- Filling in Metadata
- Publishing

- .txt, .csv, .xlsx, .zip, DwC-A files
- DB connection
- URL

- Manual to Automated
- Allows data modifications

- Manual or via EML.xml file

- Quality Control checks

EMODnet Biology data submission guidance

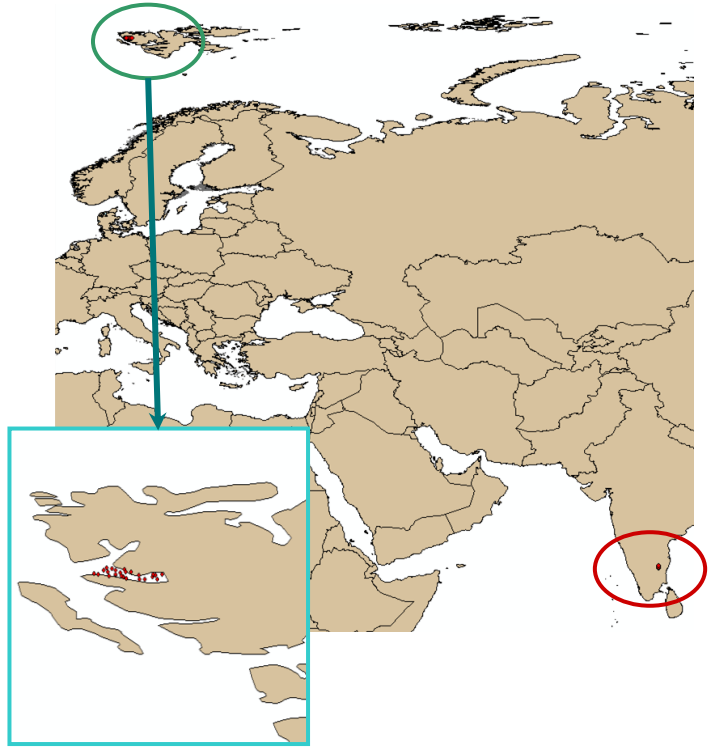
EMODnet
European Marine Observation and Data Network

Definition

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

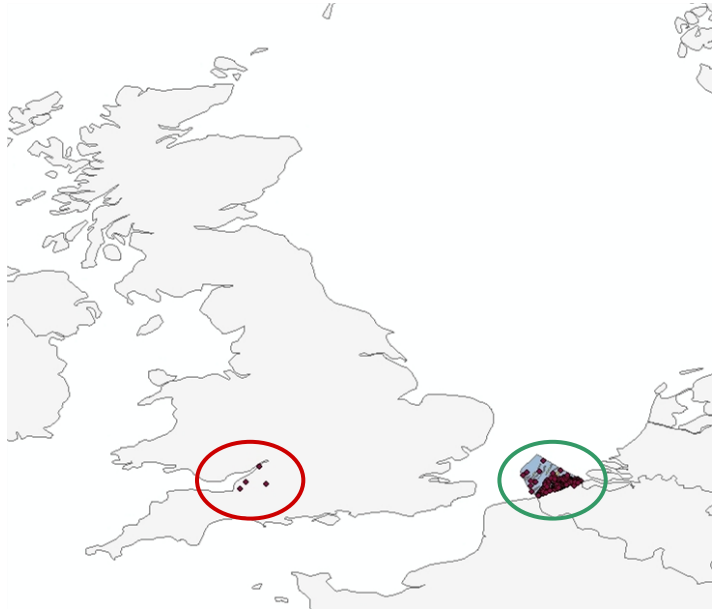
Rationale

“Monitoring in Kongsfjorden area”



Latitude & longitude switched

“Monitoring in Belgian part of the North Sea”



“+” & “-” signs switched

Rationale

	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
Mediteranean	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

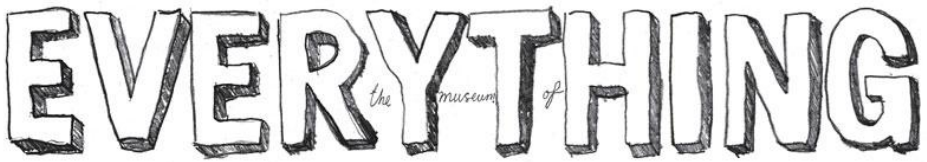
"From 6,172 unique taxon names […] to 4,525, mostly due to spelling variations and synonymy."

" … Such [taxonomic] quality control is highly needed, since a misspelled or obsolete name could be compared to the introduction of a rare species, with adverse effects on further (biodiversity) calculations…"

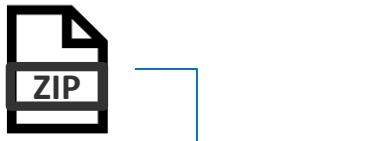
Source: Vandepitte *et al.* (2010)

What needs to be QCed

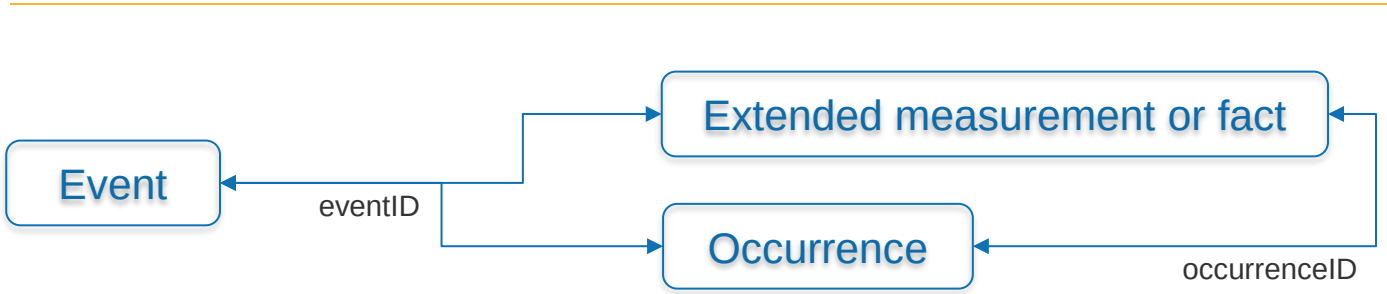
- Adherence to chosen standard
- Content format (dates, coordinates, etc.)
- Unique value fields (IDs)
- Duplicated records and redundant information
- Impossible values and outliers
- Mandatory data
- Completeness (units, geodetic datums, etc.)
- ...



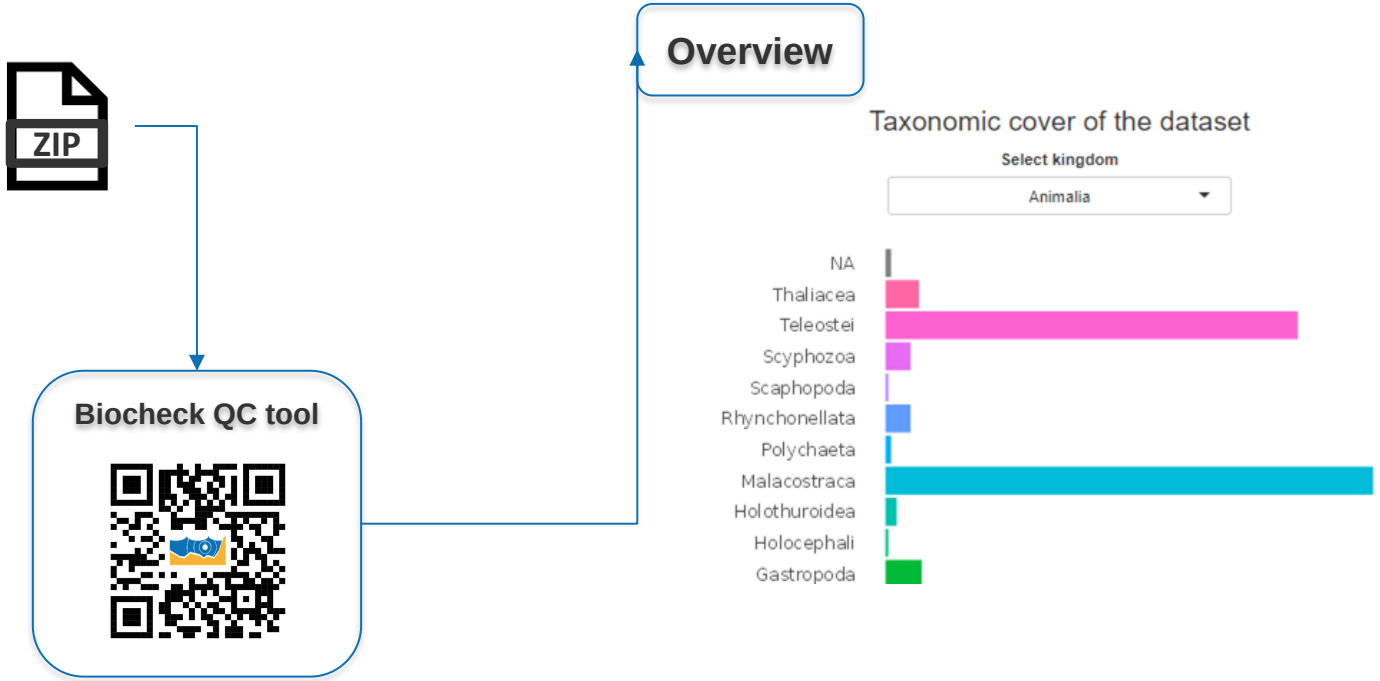
The Biocheck tool

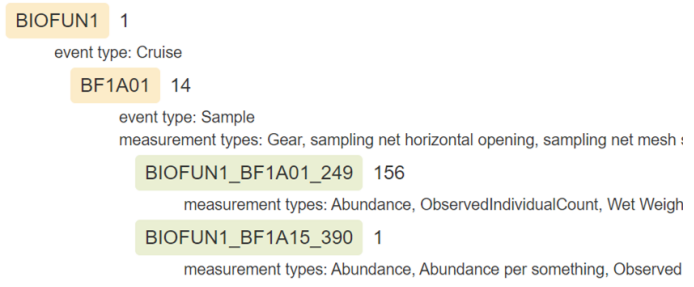


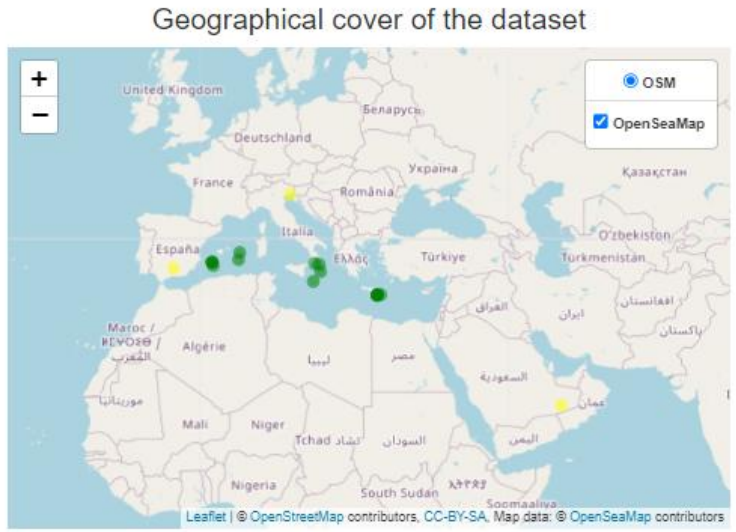
Biocheck QC tool

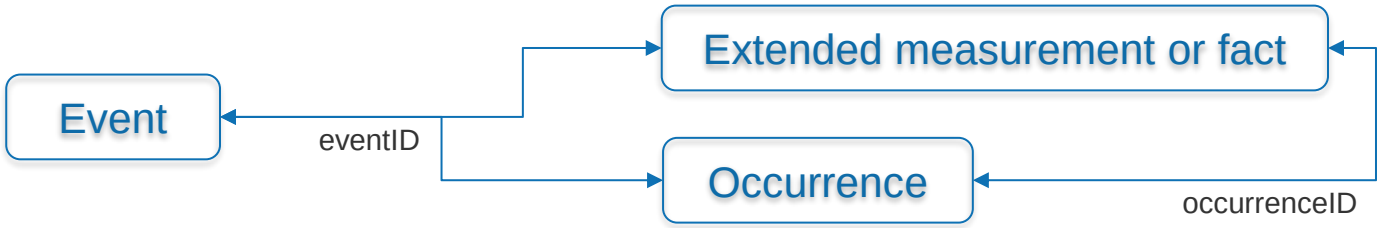


The Biocheck tool





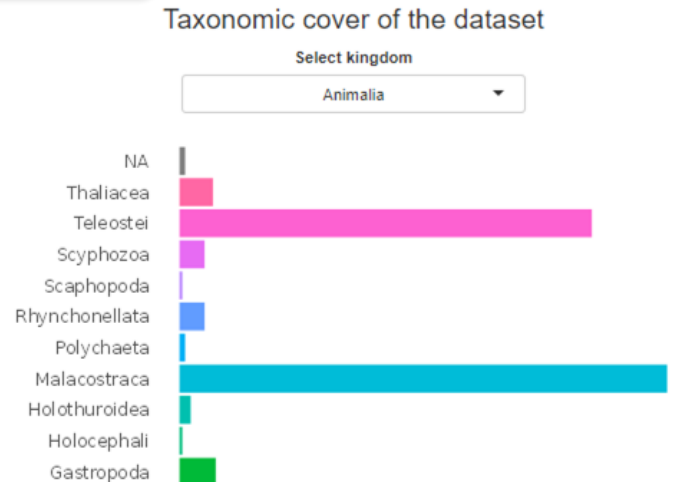


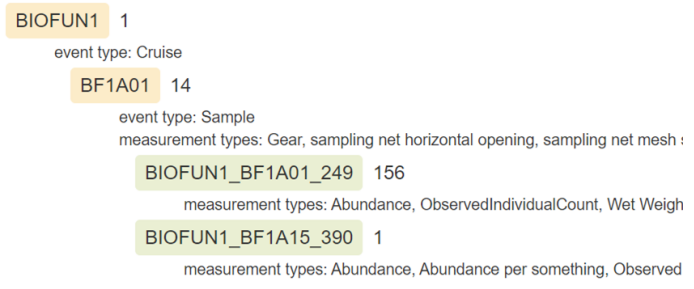


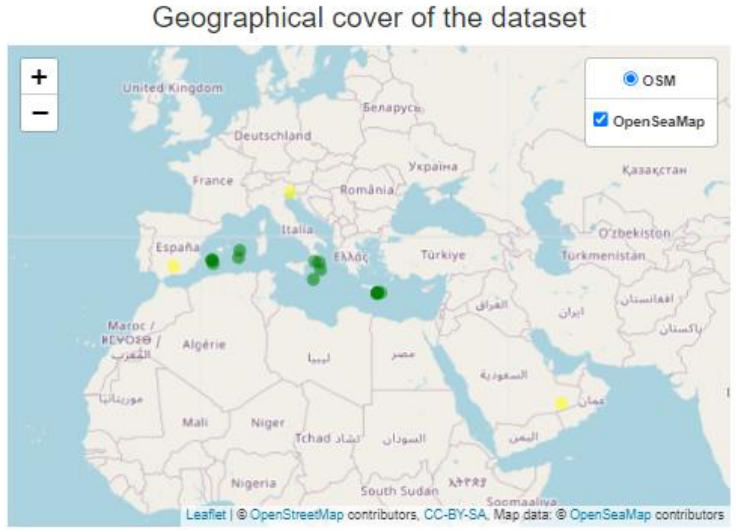
The Biocheck tool

Biocheck QC tool

Overview

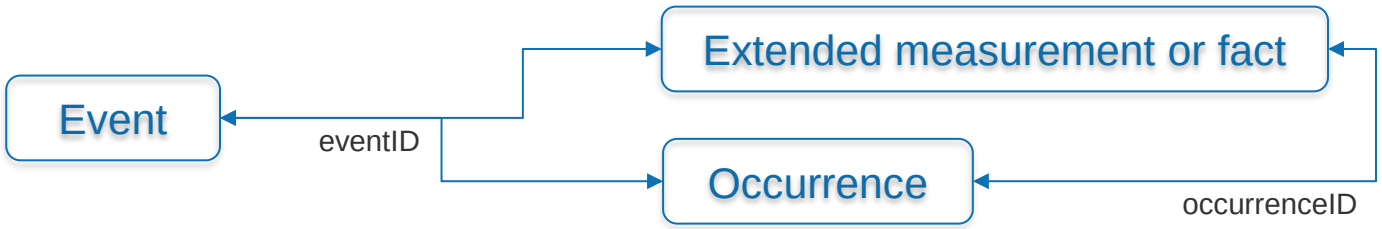






Issues

- Format and integrity
- Taxonomy
- Biogeography
- Parameters



Demo

Biocheck QC tool

https://rshiny.lifewatch.be/BioCheck/

Biology

EMODnet Biology product guidance

Workshop for EU funded projects

Salvador Jesús Fernández Bejarano VLIZ / EMODnet Biology WP3

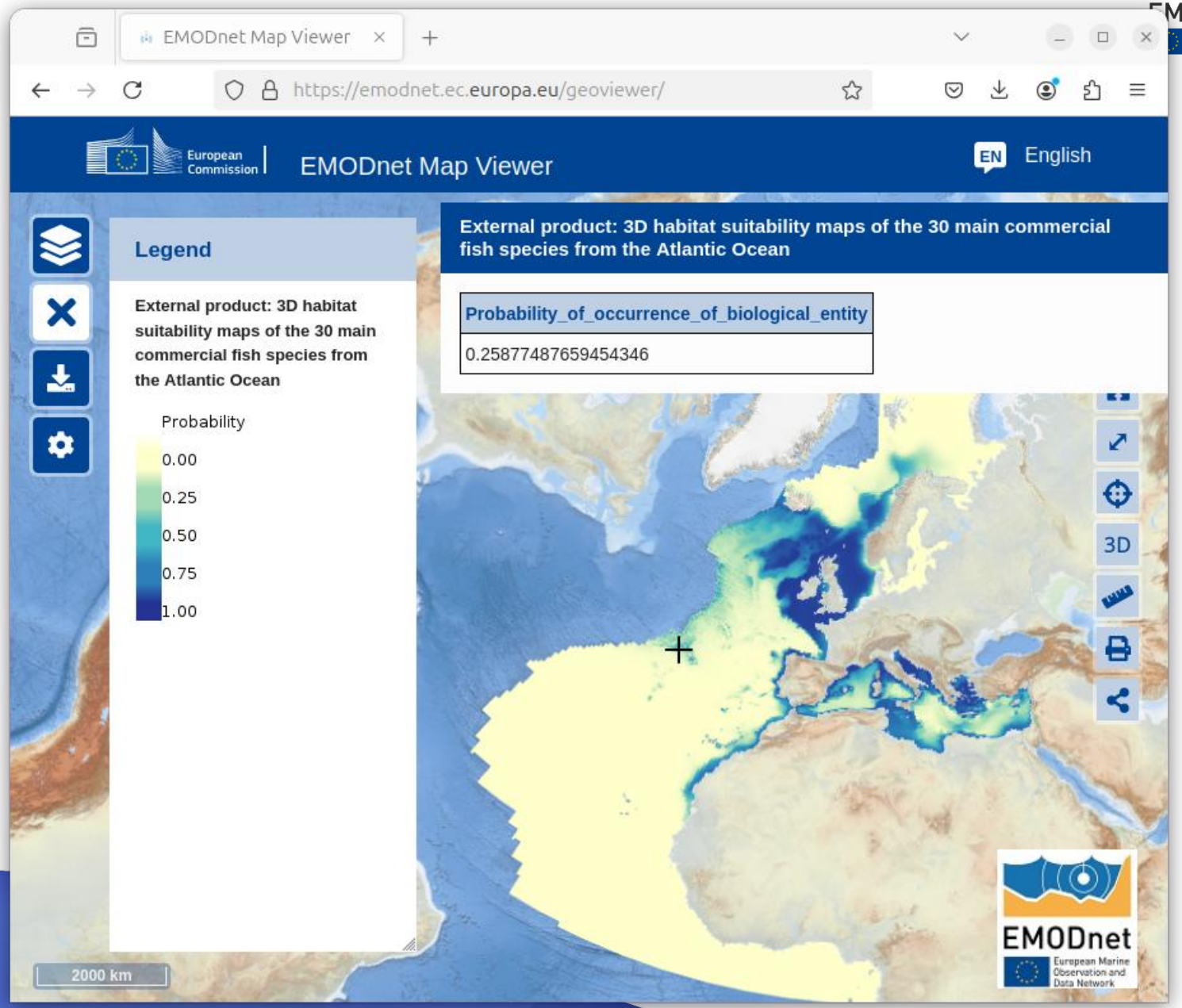
19th and 20th November 2024, Online

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.

What is a data product?

A geospatial file. Typically raster. In certain cases, the data might be of vector type.

If Raster -> **NetCDF**
If Vector -> GPKG, SHP, GeoJSON...



What is a data product?

Free Open Source Software

emodnet.wfs 2.0.2.9000

emodnet.wfs: Access EMODnet Web Feature Service data through R

The goal of emodnet.wfs is to allow interrogation of and access to EMODnet's, European Marine Observation and Data Network, geographic vector data in R through the EMODnet Web Feature Services. Web Feature services (WFS) represent a change in the way geographic information is created, modified and exchanged on the Internet and offer direct fine-grained access to geographic information at the feature and feature property level. emodnet.wfs aims at offering an user-friendly interface to this rich data.

Installation and setup

You can install the development version of emodnet.wfs from GitHub with:

```
# install.packages("pak")
pak::pak("EMODnet/emodnet.wfs")
```

If you want to avoid reading messages from emodnet.wfs such as "WFS client created successfully", set the `"emodnet.wfs.quiet"` option to `TRUE`.

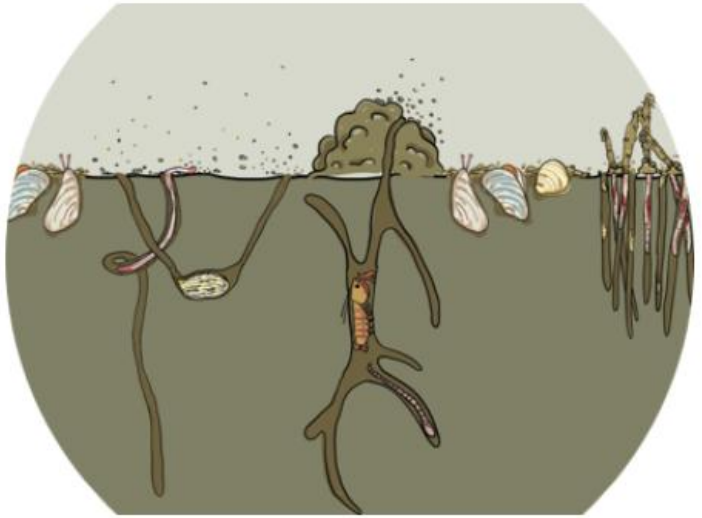
```
options("emodnet.wfs.quiet" = TRUE)
```

Btrait 0.0.0

Btrait: working with biological data and trait information

Karline Soetaert and Olivier Beauchard

Netherlands Institute for Sea Research



Middle Adriatic plankton time series

Phytoplankton

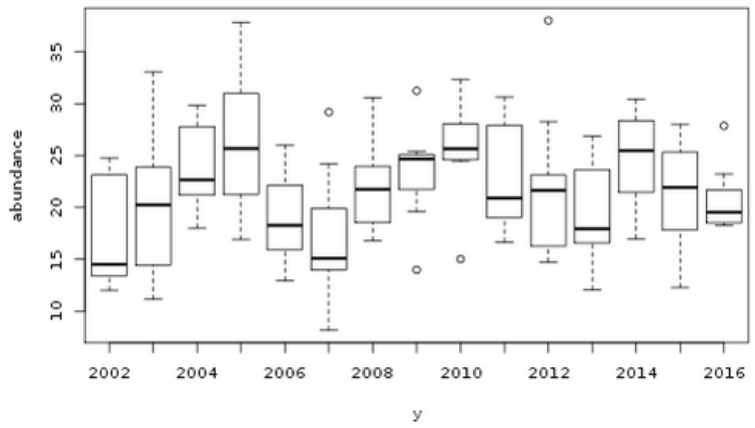
Choose a species (group)

Dinophyceae

value double sqrt transformed

Observations | Multiv 1 | about

Dinophyceae

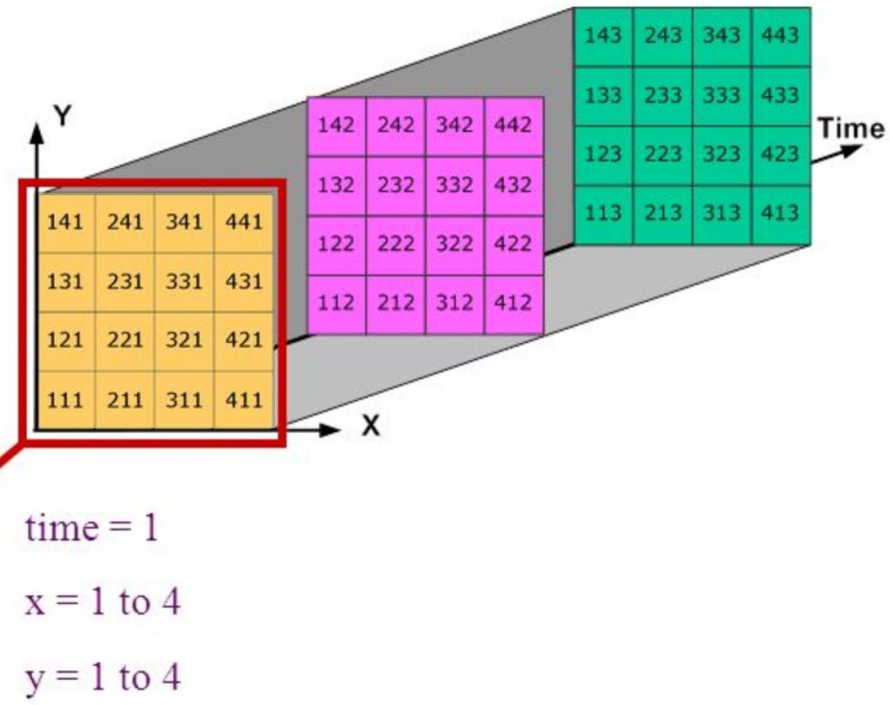


NetCDF

A netCDF file is a format for storing multi-dimensional data. It contains **dimensions** and **variables:**

- **Dimensions:** Define the shape of data, with measurements like location, time, or climate change scenarios.
- **Variables:** Store main data arrays with defined names, types, and shapes, like temperature or abundance.
  - **Coordinate Variables:** special case of variables representing the dimensions in 1D arrays.

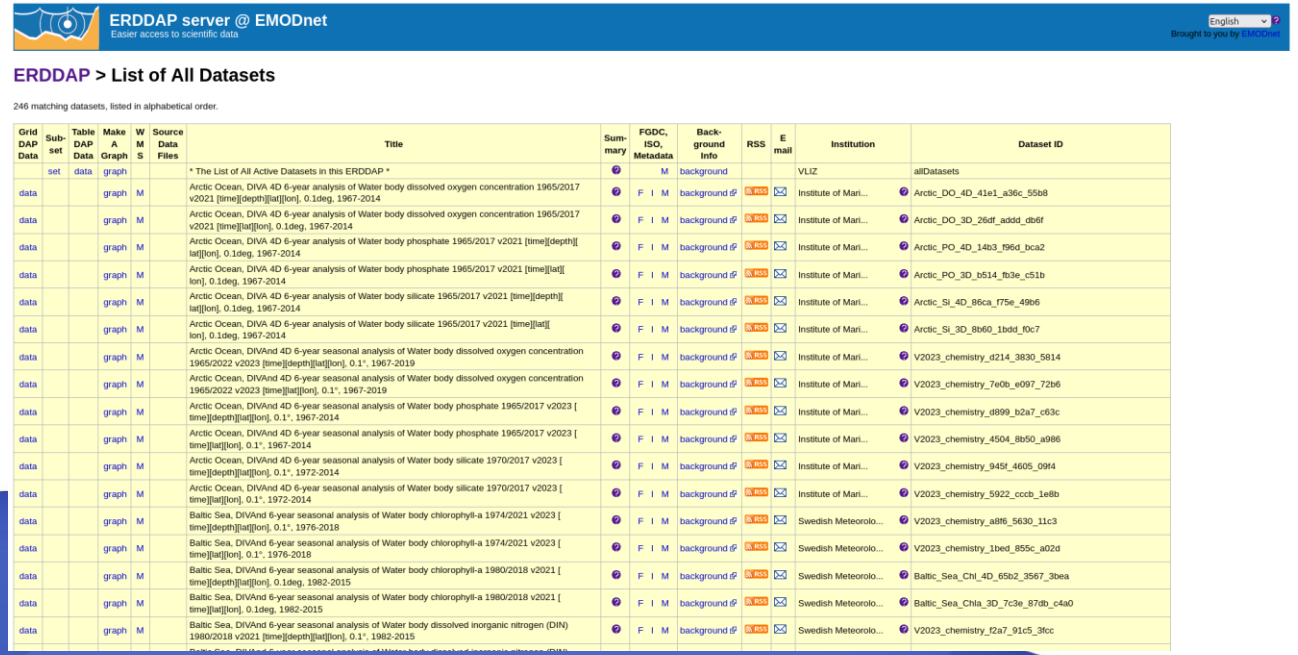
```
netcdf mynetcdf {
dimesions:
      x=4;
      y=4;
      time=UNLIMITED;
variables:
      float x(x);
      float y(y);
      int time(time);
      float temperature(time,x,y);
data:
      x = 10, 20, 30, 40;
      y = 110, 120, 130, 140;
      time = 31, 59, 90;

Temperature = 111, 211, 311, 411, 121, 221, 321,
      421, 131, 231, 331, 431, 141, 241, 341, 441;
}
```



Extracted from John Gosset, April Wright (eds): "Data Carpentry Python Ecology lesson." Version 2017.04.0, April 2017, http://www.datacarpentry.org/python-ecology-lesson/

Why NetCDF

- A centralized system was chosen for data sub-setting and downloads, requiring all thematic lots to maintain OGC web services (e.g., WMS, WFS, WCS).
- ERDDAP was selected as it supports gridded and tabular datasets, with an EMODnet ERDDAP instance set up for products from Chemistry, Physics, and other thematic lots.
- Biology's data delivery via a map viewer was limited to selected taxa, so creating NetCDFs for ERDDAP upload was identified as the optimal solution.
- This approach meets Central Portal requirements without needing custom solutions.



Grid DAP Data	Sub-set	Table DAP Data	Make A Graph	WMS	Source Data Files	Title	Sum-mary	FGDC, ISO, Metadata	Back-ground Info	RSS	E mail	Institution	Dataset ID
	set	data	graph			* The List of All Active Datasets in this ERDDAP *		M	background			VLIZ	allDatasets
data			graph	M		Arctic Ocean, DIVA 4D 6-year analysis of Water body dissolved oxygen concentration 1965/2017 v2021 [time][depth][lat][lon], 0.1deg, 1967-2014		F I M	background	RSS		Institute of Mari...	Arctic_DO_4D_41e1_a36c_55b8
data			graph	M		Arctic Ocean, DIVA 4D 6-year analysis of Water body dissolved oxygen concentration 1965/2017 v2021 [time][lat][lon], 0.1deg, 1967-2014		F I M	background	RSS		Institute of Mari...	Arctic_DO_3D_26df_addd_db6f
data			graph	M		Arctic Ocean, DIVA 4D 6-year analysis of Water body phosphate 1965/2017 v2021 [time][depth][lat][lon], 0.1deg, 1967-2014		F I M	background	RSS		Institute of Mari...	Arctic_PO_4D_14b3_f96d_bca2
data			graph	M		Arctic Ocean, DIVA 4D 6-year analysis of Water body phosphate 1965/2017 v2021 [time][lat][lon], 0.1deg, 1967-2014		F I M	background	RSS		Institute of Mari...	Arctic_PO_3D_b514_fb3e_c51b
data			graph	M		Arctic Ocean, DIVA 4D 6-year analysis of Water body silicate 1965/2017 v2021 [time][depth][lat][lon], 0.1deg, 1967-2014		F I M	background	RSS		Institute of Mari...	Arctic_Si_4D_86ca_f75e_49b6
data			graph	M		Arctic Ocean, DIVA 4D 6-year analysis of Water body silicate 1965/2017 v2021 [time][lat][lon], 0.1deg, 1967-2014		F I M	background	RSS		Institute of Mari...	Arctic_Si_3D_8b60_1bdd_f0c7
data			graph	M		Arctic Ocean, DIVAnd 4D 6-year seasonal analysis of Water body dissolved oxygen concentration 1965/2022 v2023 [time][depth][lat][lon], 0.1°, 1967-2019		F I M	background	RSS		Institute of Mari...	V2023_chemistry_d214_3830_5814
data			graph	M		Arctic Ocean, DIVAnd 4D 6-year seasonal analysis of Water body dissolved oxygen concentration 1965/2022 v2023 [time][lat][lon], 0.1°, 1967-2019		F I M	background	RSS		Institute of Mari...	V2023_chemistry_7e0b_e097_72b6
data			graph	M		Arctic Ocean, DIVAnd 4D 6-year seasonal analysis of Water body phosphate 1965/2017 v2023 [time][depth][lat][lon], 0.1°, 1967-2014		F I M	background	RSS		Institute of Mari...	V2023_chemistry_d899_b2a7_c63c
data			graph	M		Arctic Ocean, DIVAnd 4D 6-year seasonal analysis of Water body phosphate 1965/2017 v2023 [time][lat][lon], 0.1°, 1967-2014		F I M	background	RSS		Institute of Mari...	V2023_chemistry_4504_8b50_a986
data			graph	M		Arctic Ocean, DIVAnd 4D 6-year seasonal analysis of Water body silicate 1970/2017 v2023 [time][depth][lat][lon], 0.1°, 1972-2014		F I M	background	RSS		Institute of Mari...	V2023_chemistry_945f_4605_09f4
data			graph	M		Arctic Ocean, DIVAnd 4D 6-year seasonal analysis of Water body silicate 1970/2017 v2023 [time][lat][lon], 0.1°, 1972-2014		F I M	background	RSS		Institute of Mari...	V2023_chemistry_5922_cccb_1e8b
data			graph	M		Baltic Sea, DIVAnd 6-year seasonal analysis of Water body chlorophyll-a 1974/2021 v2023 [time][depth][lat][lon], 0.1°, 1976-2018		F I M	background	RSS		Swedish Meteorolo...	V2023_chemistry_a8f6_5630_11c3
data			graph	M		Baltic Sea, DIVAnd 6-year seasonal analysis of Water body chlorophyll-a 1974/2021 v2023 [time][lat][lon], 0.1°, 1976-2018		F I M	background	RSS		Swedish Meteorolo...	V2023_chemistry_1bed_855c_a02d
data			graph	M		Baltic Sea, DIVAnd 6-year seasonal analysis of Water body chlorophyll-a 1980/2018 v2021 [time][depth][lat][lon], 0.1deg, 1982-2015		F I M	background	RSS		Swedish Meteorolo...	Baltic_Sea_Chl_4D_65b2_3567_3bea
data			graph	M		Baltic Sea, DIVAnd 6-year seasonal analysis of Water body chlorophyll-a 1980/2018 v2021 [time][lat][lon], 0.1deg, 1982-2015		F I M	background	RSS		Swedish Meteorolo...	Baltic_Sea_Chla_3D_7c3e_87db_c4a0
data			graph	M		Baltic Sea, DIVAnd 6-year seasonal analysis of Water body dissolved inorganic nitrogen (DIN) 1980/2018 v2021 [time][depth][lat][lon], 0.1°, 1982-2015		F I M	background	RSS		Swedish Meteorolo...	V2023_chemistry_f2a7_91c5_3fcc

Why NetCDF

```
$ tree
.
├── Amblyraja radiata
│   ├── Amblyraja_radiata_A1B_pred2020.asc
│   ├── Amblyraja_radiata_A1B_pred2030.asc
│   ├── Amblyraja_radiata_A1B_pred2040.asc
│   ├── Amblyraja_radiata_A1B_pred2050.asc
│   ├── Amblyraja_radiata_A1B_pred2060.asc
│   ├── Amblyraja_radiata_A1B_Training.asc
│   ├── Amblyraja_radiata_RCP4.5_pred2020.asc
│   ├── Amblyraja_radiata_RCP4.5_pred2030.asc
│   ├── Amblyraja_radiata_RCP4.5_pred2040.asc
│   ├── Amblyraja_radiata_RCP4.5_pred2050.asc
│   ├── Amblyraja_radiata_RCP4.5_pred2060.asc
│   ├── Amblyraja_radiata_RCP4.5_Training.asc
│   ├── Amblyraja_radiata_RCP8.5_pred2020.asc
│   ├── Amblyraja_radiata_RCP8.5_pred2030.asc
│   ├── Amblyraja_radiata_RCP8.5_pred2040.asc
│   ├── Amblyraja_radiata_RCP8.5_pred2050.asc
│   ├── Amblyraja_radiata_RCP8.5_pred2060.asc
│   └── Amblyraja_radiata_RCP8.5_Training.asc
├── Anarhichas lupus
│   ├── Anarhichas_lupus_A1B_pred2020.asc
│   ├── Anarhichas_lupus_A1B_pred2030.asc
# ...
```

Data management hell

- No standard way of organizing files
- Dimensions are contained in the file names (or folder names!!!)
- No metadata

Why NetCDF

```
$ ncdump -h file.nc
netcdf \file {
dimensions:
    aphiaid = 49 ;
    lat = 63 ;
    lon = 79 ;
    time = 6 ;
    emmision_scenario = 3 ;
variables:
    int aphiaid(aphiaid) ;
        aphiaid:long_name = "Life Science Identifier - World Register of Marine Species" ;
        aphiaid:units = "level" ;
    double lat(lat) ;
        lat:units = "degrees_north" ;
        lat:standard_name = "latitude" ;
        lat:long_name = "Latitude" ;
    double lon(lon) ;
        lon:units = "degrees_east" ;
        lon:standard_name = "longitude" ;
        lon:long_name = "Longitude" ;
    double time(time) ;
        time:standard_name = "time" ;
        time:long_name = "Time" ;
        time:units = "days since 1970-01-01 00:00:00" ;
        time:calendar = "gregorian" ;
    int emmision_scenario(emmision_scenario) ;
        emmision_scenario:long_name = "Climate Change Emission Scenarios" ;
        emmision_scenario:units = "level"
    double probability_of_occurrence(time, emmision_scenario, aphiaid, lat, lon) ;
        probability_of_occurrence:_FillValue = -99999. ;
        probability_of_occurrence:long_name = "Probability of occurrence of biological entity"
;
```

- All data in one single file
- Data organized in a 5 dimensions array, including species (aphiaid), time and emission scenario.
- The file includes metadata and CF compliant standard names names.

COARDS and CF compliance

COARDS

Set of rules to organize **geospatial data in a NetCDF file:**

- Basic data structure with **coordinate variables**: lon, lat, depth and time. These must be **numeric** and **monotonic**.
- Longitude and latitude are structured in **regular grids**.
- Time must be stored as **numbers** (e.g., `days since 1970-01-01 00:00:00`).
- All coordinate variables require units, even if they don't have.

CF (Climate Forecast)

More flexible than COARDS, aimed to i**ncrease interoperability**:

- Coordinate variables have **standard_name**, **units**, and **spatial referencing** attributes. **Coordinate variables can be characters.**
- Support **multi-dimensional variables** and consistent handling of **missing values**.
- Global Attributes: Provide self-explanatory and extensive metadata for **increased interoperability.**
- All coordinate variables require units, even if they don't have.

The taxon dimension

The CF way

https://cfconventions.org/cf-conventions/cf-conventions.html#taxon-names-and-identifiers

```
dimension:
  time = 100 ;
  string80 = 80 ;
  taxon = 2 ;
variables:
  float time(time);
    time:standard_name = "time" ;
    time:units = "days since 2019-01-01" ;
  float abundance(time,taxon) ;
    abundance:standard_name = "number_concentration_of_biological_taxon_in_sea_water" ;
    abundance:coordinates = "taxon_lsid taxon_name" ;
  char taxon_name(taxon,string80) ;
    taxon_name:standard_name = "biological_taxon_name" ;
  char taxon_lsid(taxon,string80) ;
    taxon_lsid:standard_name = "biological_taxon_lsid" ;
data:
  time = // 100 values ;
  abundance = // 200 values ;
  taxon_name = "Calanus finmarchicus", "Calanus helgolandicus" ;
  taxon_lsid = "urn:lsid:marinespecies.org:taxname:104464", "urn:lsid:marinespecies.org:taxname:104466"
;
```

The taxon dimension

The EMODnet Biology way

```
dimension:
  time = 100 ;
  string80 = 80 ;
  aphiaid = 2 ;
variables:
  float time(time);
    time:standard_name = "time" ;
    time:units = "days since 2019-01-01" ;
  float abundance(time,aphiaid) ;
    abundance:standard_name = "number_concentration_of_biological_taxon_in_sea_water" ;
  float aphiaid(aphiaid);
    aphiaid:long_name = "Life Science Identifier - World Register of Marine Species";
    aphiaid:units = "level";
  char taxon_name(aphiaid,string80) ;
    taxon_name:standard_name = "biological_taxon_name" ;
  char taxon_lsid(aphiaid,string80) ;
    taxon_lsid:standard_name = "biological_taxon_lsid" ;
data:
  time = // 100 values ;
  abundance = // 200 values ;
  aphiaid = 104464, 104466;
  taxon_name = "Calanus finmarchicus", "Calanus helgolandicus" ;
  taxon_lsid = "urn:lsid:marinespecies.org:taxname:104464", "urn:lsid:marinespecies.org:taxname:104466"
;
```

Climatological statistics

Example: Species distribution models of habitat suitability for future decades:

- CF allows climatological statistics over specific time ranges
- A special time axis represents each time range with a single point (e.g., mid-decade).
- The "climatology" attribute defines the start and end intervals for each range.
- "Cell methods" describe how statistics are calculated (e.g., mean or maximum)

```
dimensions:
  time = 3;
  nv = 2;
variables:
  float habitat_suitability(time);
    habitat_suitability:long_name = "habitat suitability for species";
    habitat_suitability:cell_methods = "time: mean within years time: mean over years";
    habitat_suitability:units = "probability";
    habitat_suitability:_FillValue = -999.9;
  double time(time);
    time:climatology = "climatology_bounds";
    time:units = "days since 2000-1-1";
    time:calendar = "gregorian";
  double climatology_bounds(time, nv);
data:
  time = "2045-1-1", "2055-1-1", "2065-1-1";
  climatology_bounds = "2040-1-1", "2049-12-31",
                       "2050-1-1", "2059-12-31",
                       "2060-1-1", "2069-12-31";
```

Climate change scenarios and other dimensions

In some studies, there are data modelled for different climate change scenarios. In other cases, there are specific dimensions for which data have been modelled.

There is no standard yet for these cases.

Our solution: create a dimension with coordinate variable as numeric series (e.g. 1, 2, 3...). And one more variable of type character that labels the coordinate variable

```
dimensions:
  time = 3;
  string80 = 80;
  emission_scenario = 2;
variables:
  double time(time);
    time:standard_name = "time";
    time:long_name = "Time";
    time:units = "days since 1970-01-01 00:00:00";
    time:calendar = "gregorian";
  int emission_scenario(emission_scenario);
    emission_scenario:long_name = "Climate Change Emission Scenarios";
    emission_scenario:units = "level";
    emission_scenario:description = "1 = RCP 4.5; 2 = RCP 8.5";
  char emission_scenario_char(emission_scenario, string80);
    emission_scenario_char:long_name = "Climate Change Emission Scenarios CHAR";
  double probability_of_occurrence(time, emission_scenario);
    probability_of_occurrence:_FillValue = -99999.0;
    probability_of_occurrence:long_name = "Probability of occurrence of biological entity";
data:
  time = 18263, 18628, 18993;
  emission_scenario = 1, 2;
  emission_scenario_char = "RCP 4.5", "RCP 8.5";
```

Standardizing variables

No CF standard names exist for common metrics in biodiversity

- Habitat suitability
- Probability of occurrence
- Shannon index
- Species loss
- …

We are working in **aligning** the terms we use and submit **controlled vocabularies**.

More info

EMODnet Biology products NetCDF guide
https://github.com/EMODnet/EMODnet-Biology-NetCDF-Guide

EMODnet Biology products demo
https://github.com/EMODnet/EMODnet-Biology-products-erddap-demo

Products template
https://github.com/EMODnet/EMODnet-Biology-Project-Template

EMODnet
European Marine Observation and Data Network

Biology

EMODnet Biology tools and services

Workshop for EU funded projects

Marina Lipizer OGS/EMODnet Biology WP4 “Uptake and Outreach” leader

19th and 20th November 2024, Online

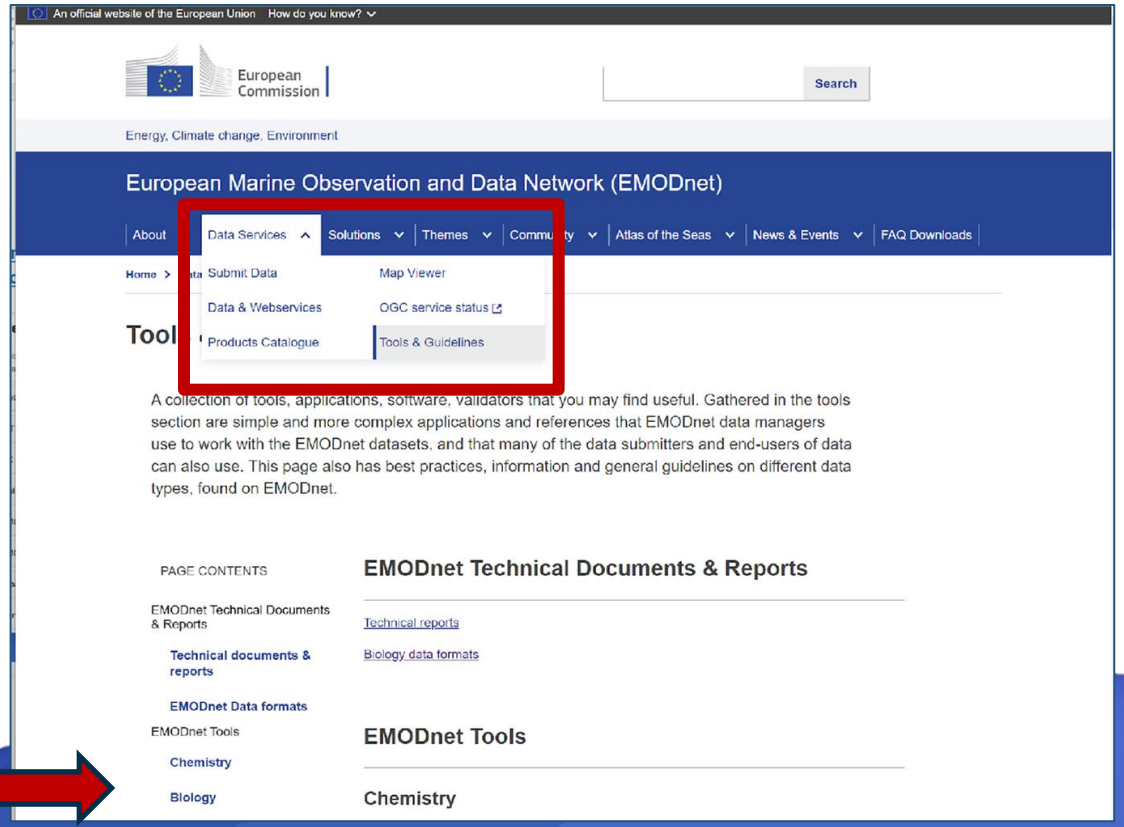
The European Marine Observation and Data Network (EMODnet) is financed by the European Union under Regulation (EU) 2021/1139 of the European Parliament and of the Council of 7 July 2021 establishing the European Maritime, Fisheries and Aquaculture Fund.

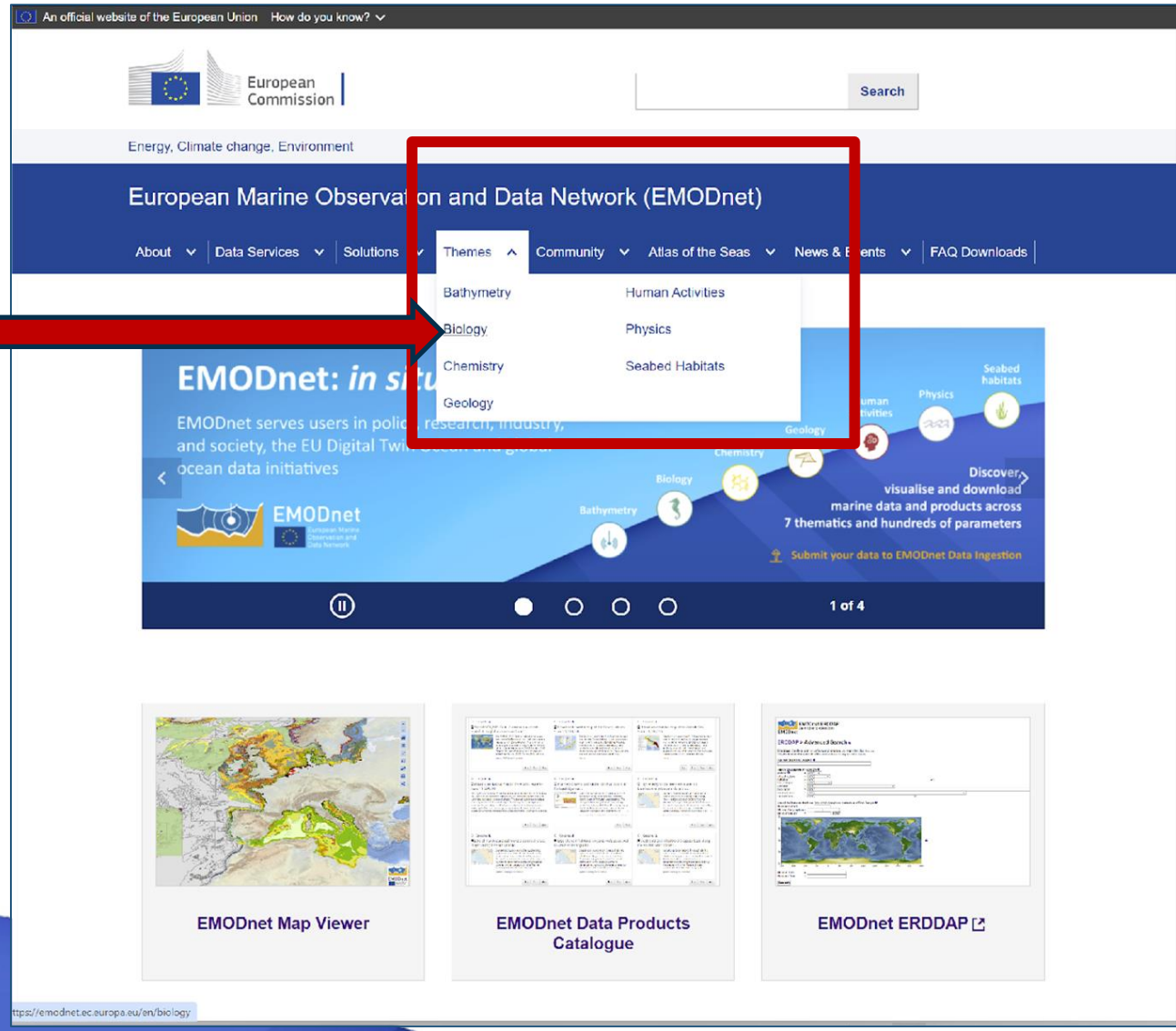
Finding what kind of support EMODnet can provide:

https://emodnet.ec.europa.eu/en

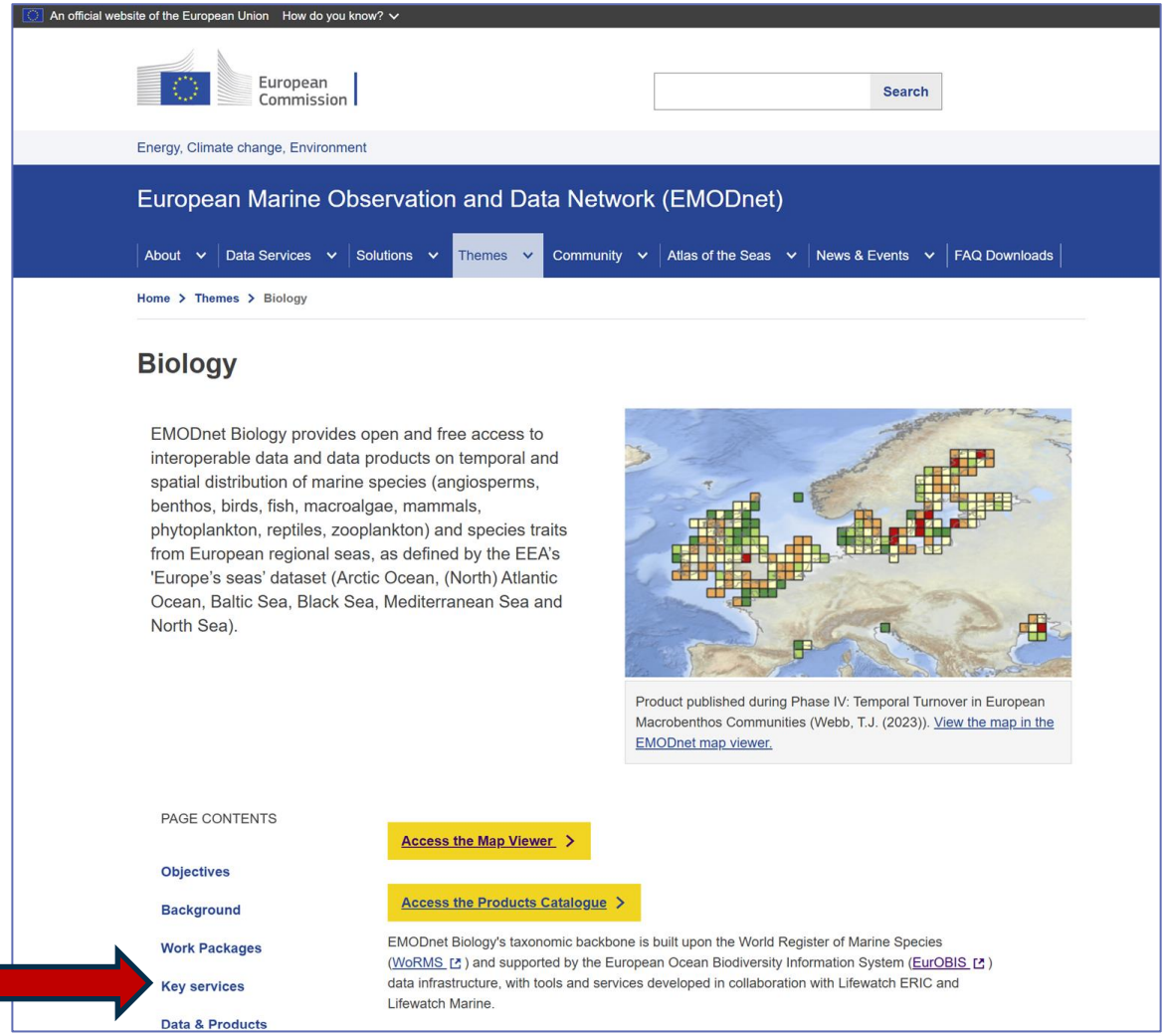
Several services to support:

- ❖Dataset preparation & formatting
- ❖Data Quality Control
- ❖Data access and visualization
- ❖…





https://emodnet.ec.europa.eu/en/biology



EMODnet
European Marine Observation and Data Network

https://emodnet.ec.europa.eu/en/biology#biology-key-services

Key services

EMODnet Biology provides key services and products which allow users to search and visualise data and related data products:

- + Webservices
- + IPT
- + QC Tool
- + Online course (MOOC): "Contributing datasets to EMODnet Biology"
- + EMODnetWCS: Access EMODnet Web Coverage Service data through R
- + EMODnetWFS: Access EMODnet Web Feature Service data through R
- + R package: BTrait - working with biological data and trait information
- + Darwin Core data file template

Webservices

- Occ
  mea
- Retr
- Usin
- Usin
- Exa

EMODnetWCS: Access EMODnet Web Coverage Service data through R

The R package, partially funded by
EMODnet thematic lots' raster data lay
users to query information on and dow
Coverage Service (WCS) endpoints di

Darwin Core data file template

The following file includes information about the mandatory and recommended fields for data submission to EMODnet Biology. Please contact us via bio@emodnet.eu if you have any questions.

iology#biology-

I

The In
using t
inform

EMODnetWFS: Access EMODnet Web Feature Service data through R

The R package, created by EMODnet Biology partners, is designed to make all EMODnet thematic lots' vector data layers easily accessible in R. The package allows users to query information on and download data from all available EMODnet Web Feature Service (WFS) endpoints directly into their R working environment.

R package: BTrait - working with biological data and trait information

BTrait was developed in the scope of EMODnet Biology to facilitates working with species density data, combined with species traits in R. It allows users to query the linked datasets (species density and trait data) and visualise it with an interactive shiny application.

ervices

logy provides key services and products which allow users to search and visualise
ed data products:

services

+ QC Tool
+ Online course (MOOC): "Contributing datasets to EMODnet Biology"
+ EMODnetWCS: Access EMODnet Web Coverage Service data through R
+ EMODnetWFS: Access EMODnet Web Feature Service data through R
+ R package: BTrait - working with biological data and trait information
+ Darwin Core data file template

Training Courses

- Online self paced course (MOOC)
  - ”Contributing datasets to EMODnet Biology”
  - https://classroom.oceanteacher.org/enrol/index.php?id=958
  - Platform managed by OTGA (**Ocean Teacher Global Academy**)

- EMBRC FAIR training
  - 2nd edition in April 2025
  - https://vliz.be/en/embrc-fair-training-course

- OBIS course
  - ”Contributing and publishing datasets to OBIS”
  - https://classroom.oceanteacher.org/course/view.php?id=907

Templates, guidelines

- **Darwin Core (**DwC) file **template**: includes information about the mandatory and recommended fields for data submission to EMODnet Biology.

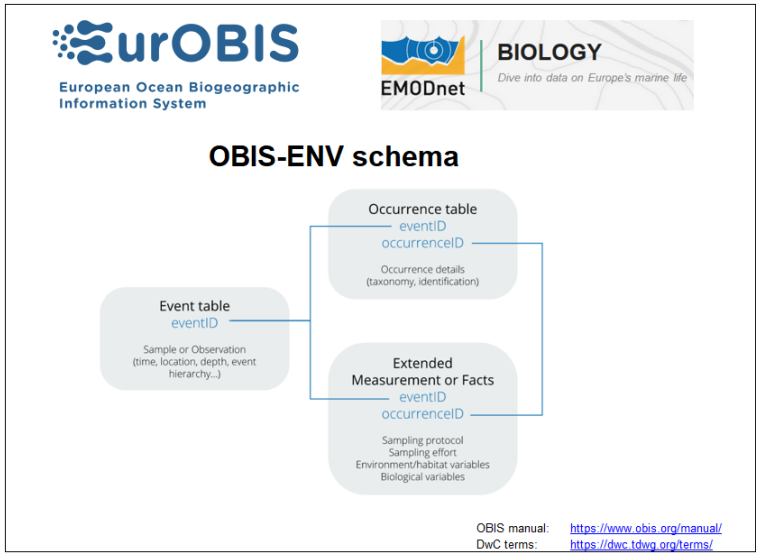
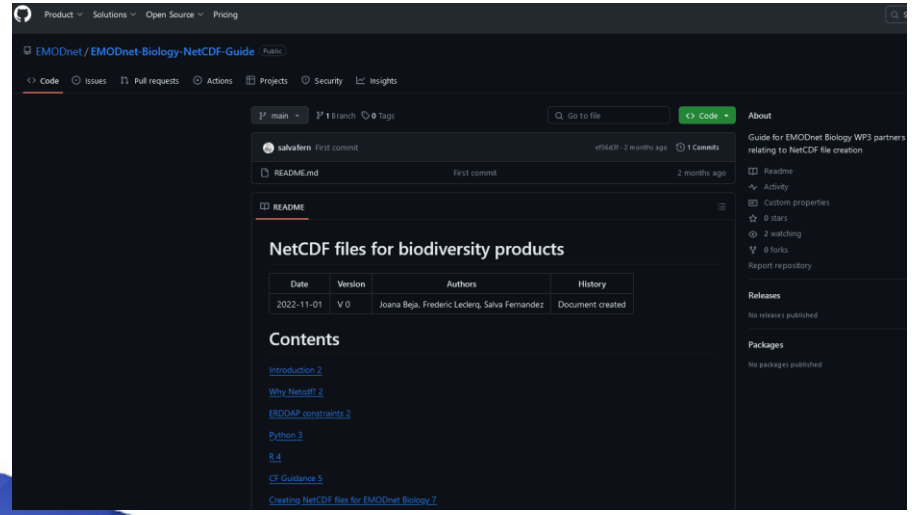


table	rank	field name	definition
Event	Mandatory	eventID	identifier or an identifier specific to the data set
	Mandatory	eventDate	suitable for a time in a geological context.
	Mandatory	decimalLatitude	Location. Positive values are north of the Equator, negative values are south of it. Legal values lie between -90 and 90, inclusive.
	Mandatory	decimalLongitude	Location. Positive values are east of the Greenwich Meridian, negative values are west of it. Legal values lie between -180 and 180,
	Mandatory	institutionCode	The name (or acronym) in use by the institution having custody of the object(s) or information referred to in the record
	Mandatory	datasetName	The name identifying the data set from which the record was derived
	Highly Recommended	maximumDepthInMeters	The greater depth of a range of depth below the local surface, in meters
	Highly Recommended	minimumDepthInMeters	The lesser depth of a range of depth below the local surface, in meters
	Highly Recommended	coordinateUncertaintyInMeters	whole of the Location. Leave the value empty if the uncertainty is unknown, cannot be estimated, or is not applicable (because there are
	Recommendable	footprintWKT	radius representation (see decimalLatitude) and a footprint representation, and they may differ from each other.
	parenEventID exists	type	need a field to explain what the event (http://obis.org/manual/darwincore/#event) refers to (sample, subsample, stationVisit,...) | 2. this
	Mandatory if exists	parenEventID	An identifier for the broader Event that groups this and potentially other Events.
	Other	dataGeneralizations	may be available on request
	Other	eventRemarks	Comments or notes about the Event
	Other	samplingProtocol	The name of, reference to, or description of the method or protocol used during an Event
	Other	locationID	specific to the data set
	Other	locality	(higherGeography, continent, country, stateProvince, county, municipality, waterBody, island, islandGroup). This term may contain
	Other	locationRemarks	Comments or notes about the Location

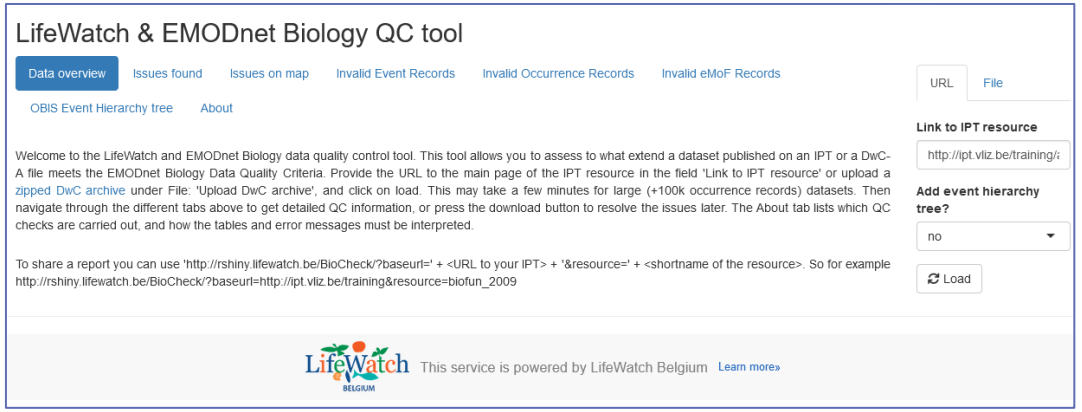
DwC OBIS-ENV schema | **Event** | Occurrence | ExtendedMeasurementOrFact | DwC Term definitions

- Biodiversity data product creation guidance
  - https://github.com/EMODnet/EMODnet-Biology-NetCDF-Guide



Tools

- In collaboration with LifeWatch Belgium we developed a **data QC tool**
  - Available as an rshiny application (https://rshiny.lifewatch.be/BioCheck/) or
  - R Package (https://github.com/EMODnet/EMODnetBiocheck)



- VLIZ hosts and provides technical support for > 30 IPT instances
  - IPT for data sharing (https://ipt.vliz.be/eurobis)

- We have developed R Packages for EMODnet webservices
  - For WCS services (https://github.com/EMODnet/EMODnetWCS) and
  - For WFS services (https://github.com/EMODnet/EMODnetWFS)

- Webservices
  - Data and products available via VLIZ or EMODnet CP Geoserver instances and via EMODnet CP ERDDAP instance
  - Guidance available via https://github.com/EMODnet/EMODnet-Biology-Guidance

Tutorials

External to EMODnet Biology – but the systems are aligned and interoperable

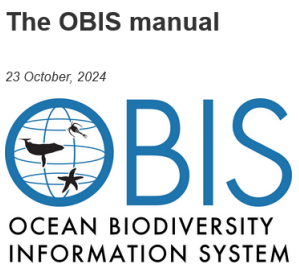
- OBIS manual
  - https://manual.obis.org/

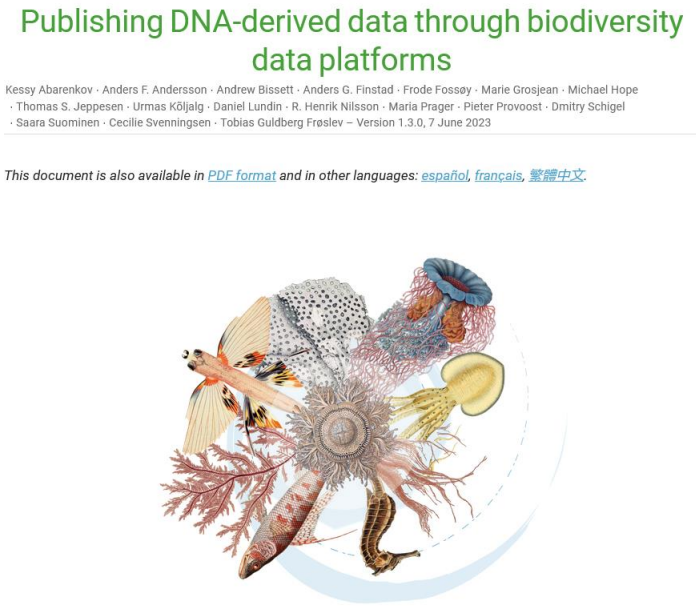
- Introduction to OBIS-ENV data format
  - https://youtu.be/K1vNssRTmyg?feature=shared

- OBIS vocabulary series
- https://www.youtube.com/playlist?list=PLlgUwSvpCFS4hADB7Slf44V1KJauEU6Ul

- OBIS and environmental (eDNA)
  - https://youtu.be/FO9sPtB47r0?feature=shared

- GBIF Publishing DNA-derived data through biodiversity data platforms
  - https://docs.gbif-uat.org/publishing-dna-derived-data/1.0/en/





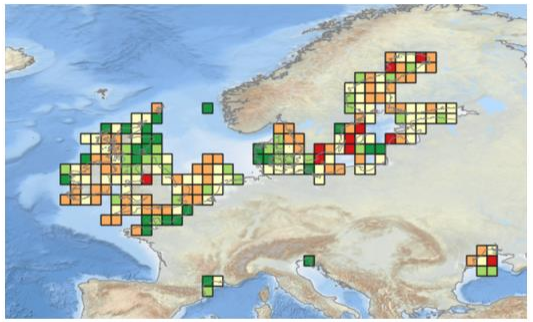
Informative material

- Objectives
- Background
- Work Packages
- Key services
- Data & Products
  - Data sources
  - Data product development
  - Data infrastructure
  - Data format
- Reports
- Communication
- Use Cases
- News

- Biology information
  - https://emodnet.ec.europa.eu/en/biology#biology-key-services
- Reports/Deliverables
  - https://emodnet.ec.europa.eu/en/reports
    - EMODnet componente= Biology
    - Publication type= Deliverable
- EMODnet catalogue
  - https://emodnet.ec.europa.eu/geonetwork/srv/eng/catalog.search#/search
  - Provided by= EMODnet Biology
- EMODnet news
  - https://emodnet.ec.europa.eu/en/news
  - https://emodnet.ec.europa.eu/en/newsletter

Biology

EMODnet Biology provides open and free access to interoperable data and data products on temporal and spatial distribution of marine species (angiosperms, benthos, birds, fish, macroalgae, mammals, phytoplankton, reptiles, zooplankton) and species traits from European regional seas, as defined by the EEA's 'Europe's seas' dataset (Arctic Ocean, (North) Atlantic Ocean, Baltic Sea, Black Sea, Mediterranean Sea and North Sea).



Product published during Phase IV: Temporal Turnover in European Macrobenthos Communities (Webb, T.J. (2023)). View the map in the EMODnet map viewer.

NEWS ARTICLE | 14 Oct 2024

New data product published by EMODnet Biology

The external product, published in EMODnet Biology and developed by the Naval Undersea Warfare Center (NUWC) and presents information on the abundance of Loggerhead turtles (Caretta Caretta) in the Mediterranean basin.

Continue reading

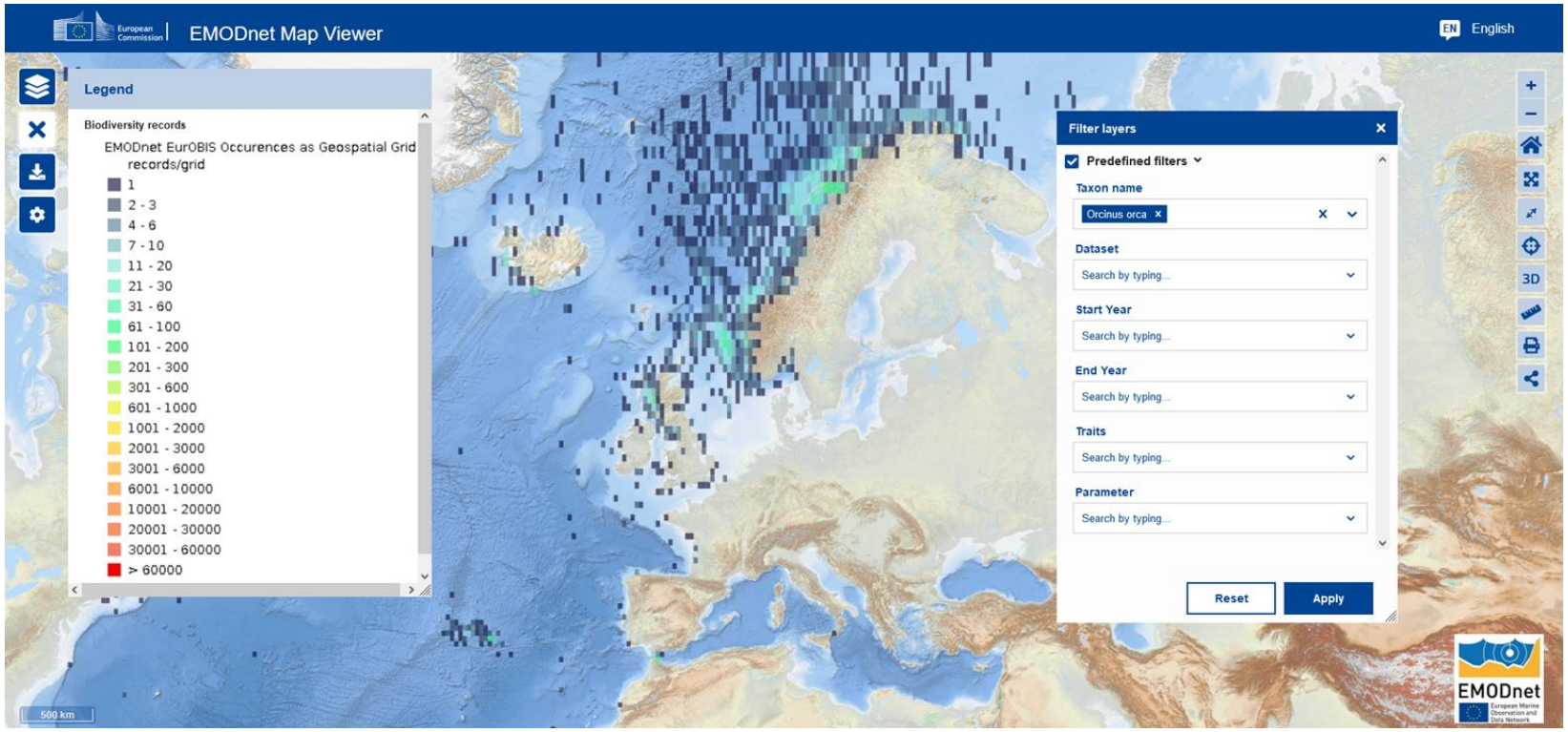
How to access EMODnet Biology data

- EMODnet map viewer

- EMODnet Biology webservices
  - GitHub
  - Guidelines
-

R Packages
  - EMODnet WCS
  - EMODnet WFS
-

Access from other initiatives
  - EU DTO Data Lake
  - EurOBIS IPT
  - OBIS



EMODnet
European Marine Observation and Data Network

Any further information & questions:

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