

# FAIR data on biochemistry in European marine waters

## Current status and way forward

Anouk Blauw, Wouter Schoonveld, Willem Stolte (Deltares)

Peter Thijssse, Tjerk Krijger (MARIS)

July 3rd 2024

# Agenda

- Introduction
- User perspective
  - Questions and answers
- FAIR data
- Data landscape at EU level
  - Questions and answers
- Mentimeter on your experiences on bottlenecks and potential solutions
  - [menti.com](https://www.menti.com/join/71162441): login code: 7116 2441



# Introduction: why we need ocean observation data

The ocean and coastal waters are changing

- Climate change
- More intense use of marine waters, for example for wind farms and aquaculture
- Biodiversity loss

We need to know how these changes impact:

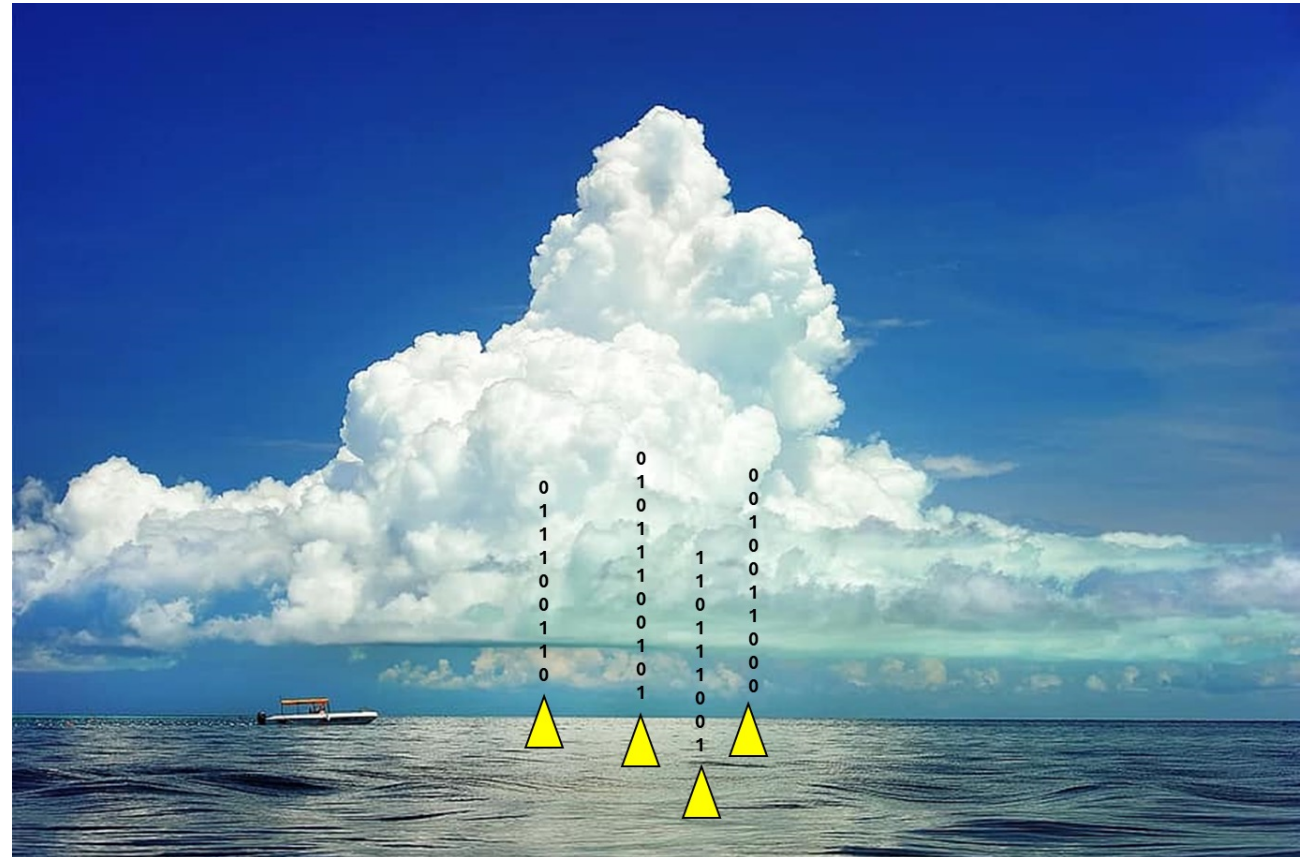
- the natural ecosystem and biodiversity
- conditions for ecosystem services, such as aquaculture and fisheries
- Our safety, through flood risks and coastal erosion

# Europe organizes data and information flows

- COPENICUS programme
  - Satellite data
  - Model data
  - In-situ data
- EMODnet
- Digital twins
- EOSC virtual research environments

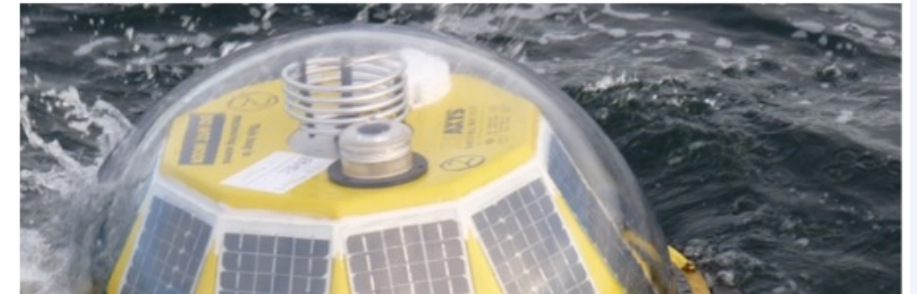
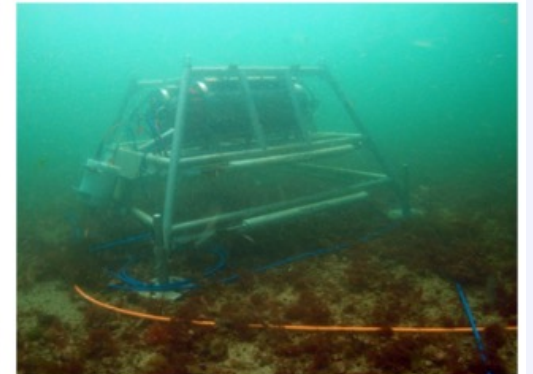
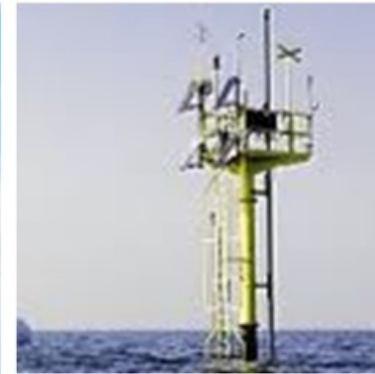
For:

- Research
- Policy
- Actors at sea



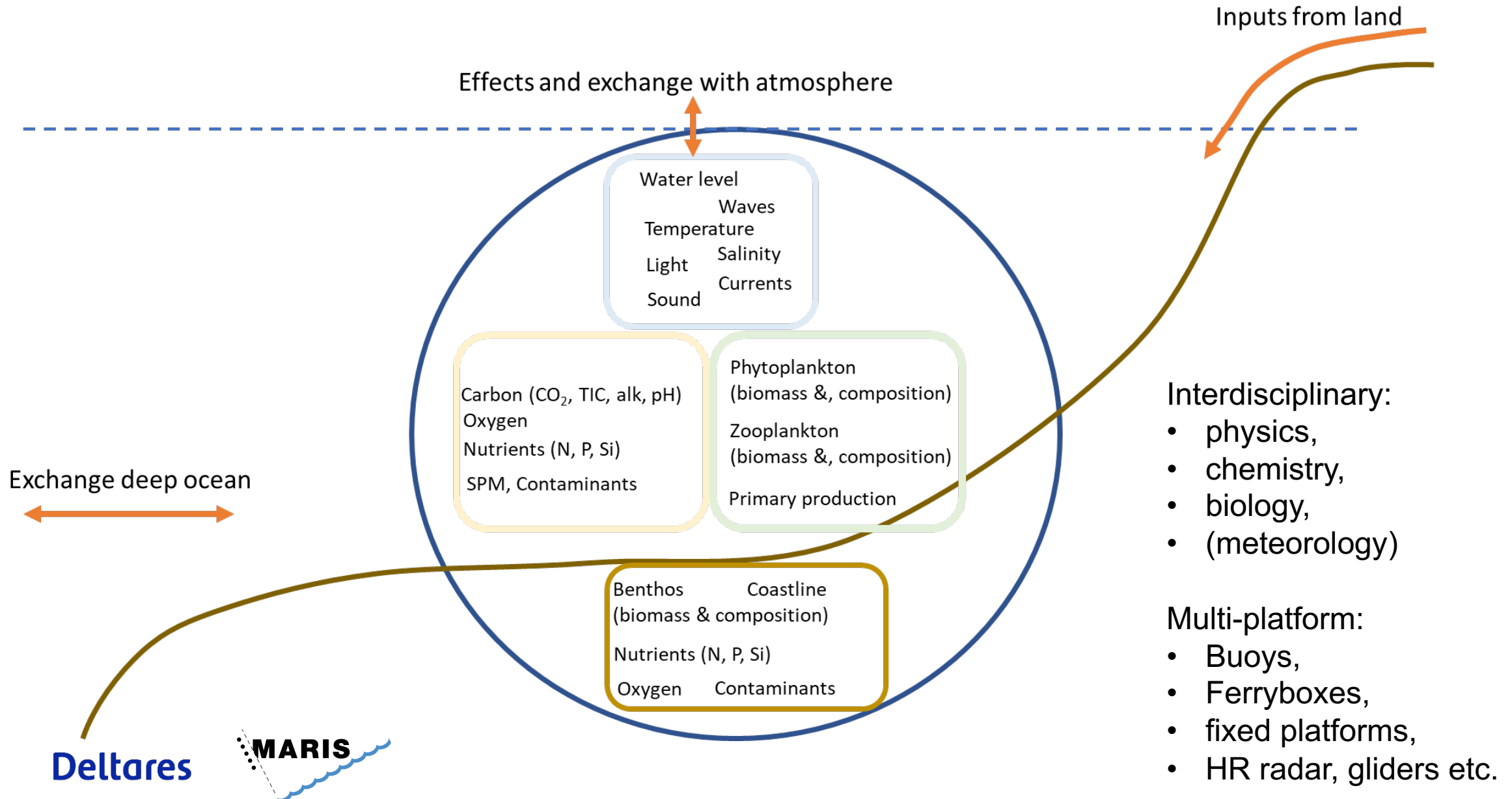
# Where do the in-situ data come from?

- Long term monitoring programmes
  - Samples – delayed mode
  - Operational oceanography with sensors:
    - near-realtime
    - quality controlled delayed mode
- Research projects
- Citizen observations
- Project-based monitoring for industry or infrastructures





# Long-term sensor observations in Jerico-RI



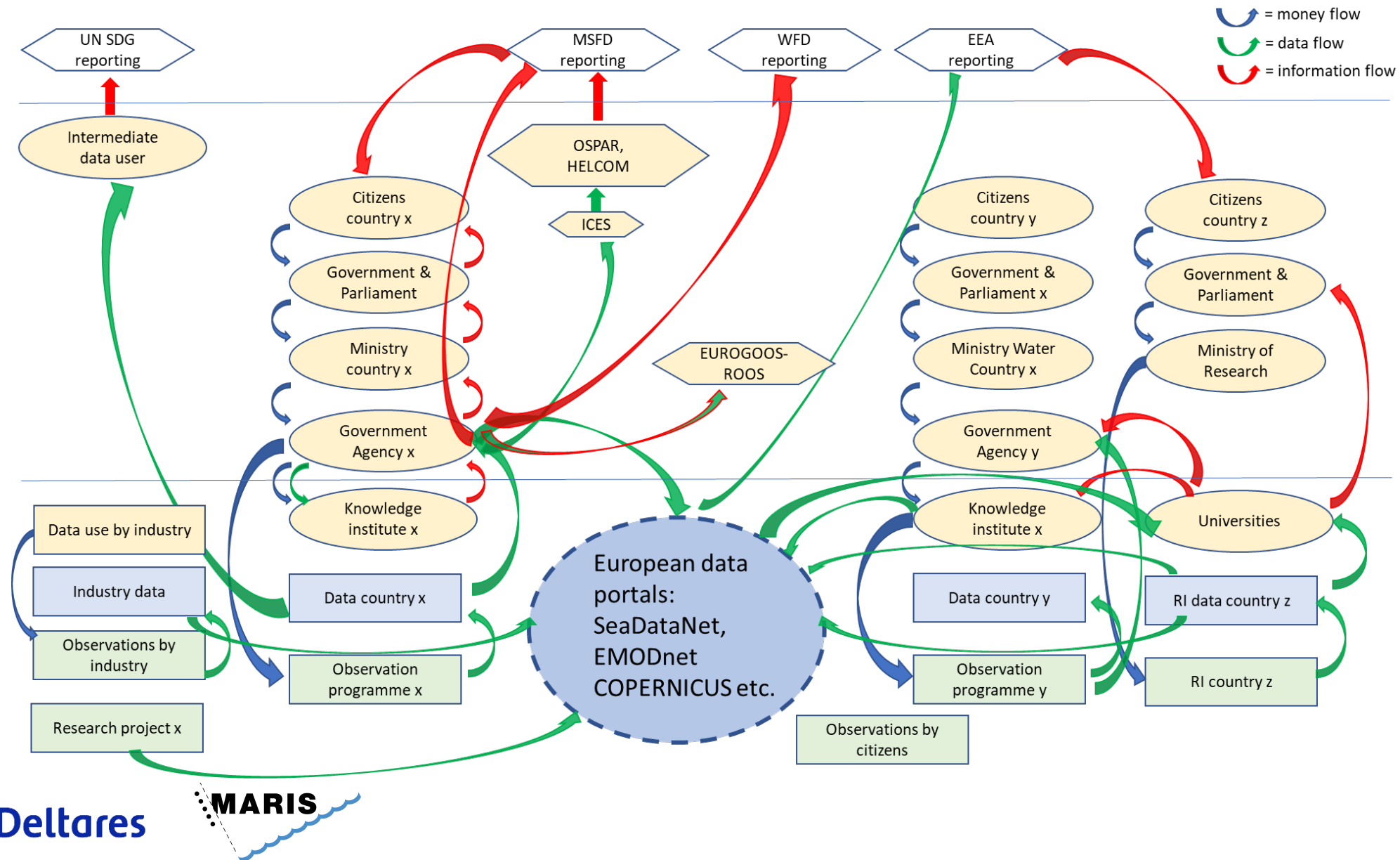
# Objectives of this webinar

- Share lessons learnt from Jerico-S3 project on: data collection, data management and data use
- Give an overview of European FAIR data management landscape
- Evaluate to what extent the current data flow to European information services is adequate
- Identify bottlenecks and potential solutions

# User perspective



# Users and producers of data for policy evaluation



# Data requirements for policy assessments:

- Representative observations per assessment area
- Representative observations of seasonal variability
- Reliability, quality control
- Transparency
- Consistency between assessment periods and areas

This is currently not provided by EMODNET, so OSPAR and HELCOM organize their assessment data through ICES

# Example for policy: do we get sufficient information?

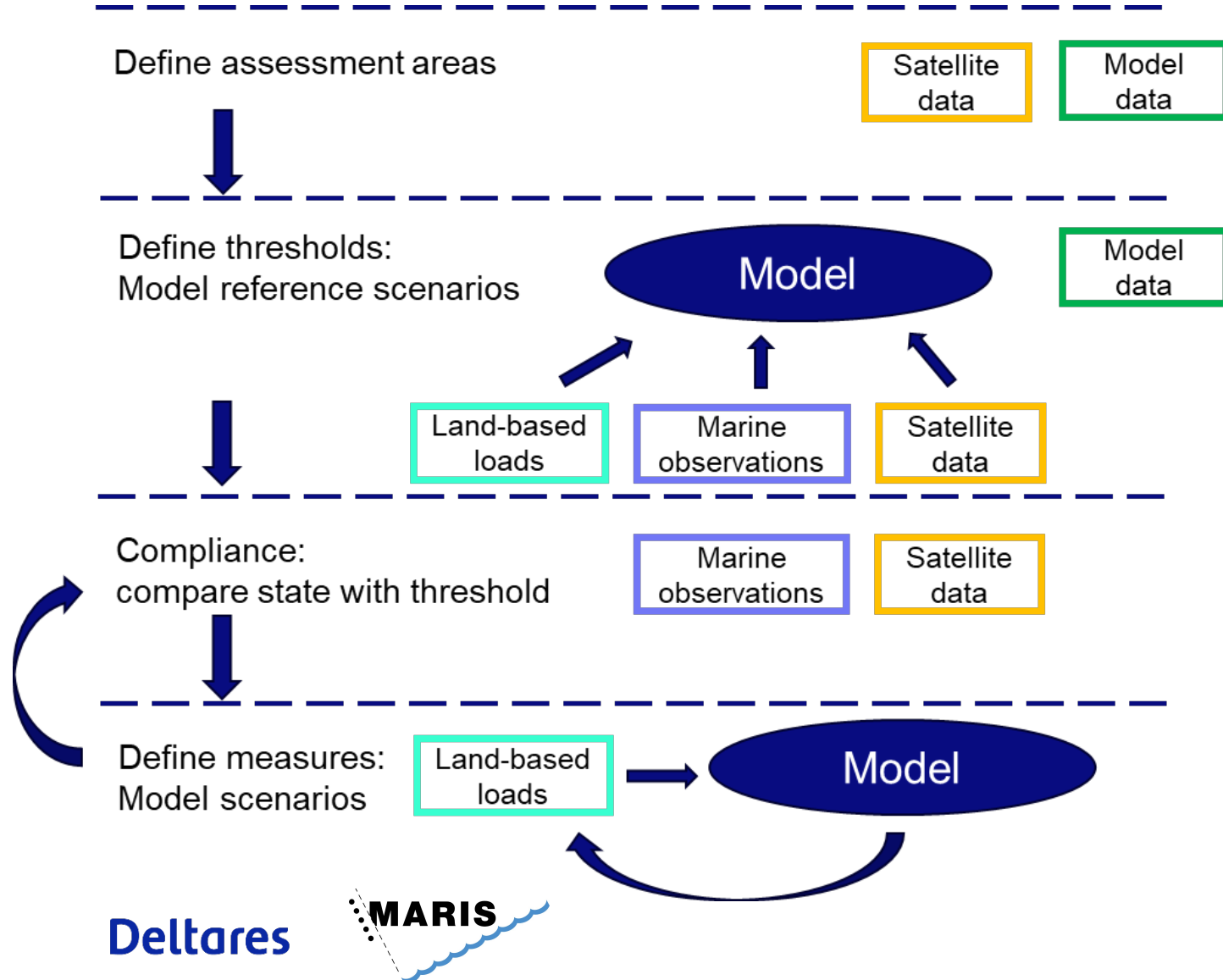
Spatial confidence oxygen assessment



Temporal confidence oxygen assessment



# Example: data use for OSPAR eutrophication assessment



- Models used for:
  - threshold setting
  - Definition of assessment areas
  - Required reductions of land-based loads
- In situ observations used for:
  - Model inputs on land-based loads
  - Model validation
  - Compliance checking
- Satellite data on chlorophyll-a used for:
  - Definition of assessment areas
  - Model validation
  - Compliance checking

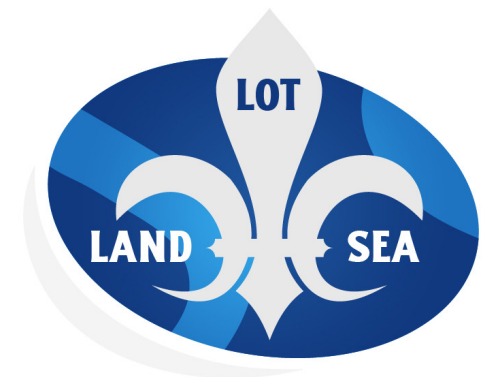
# Conclusions

Policy assessments are hampered by limited observation data

- partly because not enough observations are done
- partly because existing observations cannot be used:
  - Unknown/ hidden
  - Not accessible
  - Unknown quality

European portal for land-based inputs of nutrients and pollutants is lacking

- Requires observations at river mouths of:
  - Discharges
  - Pollutant concentrations
- Work in progress by OSPAR & HELCOM and LandSeaLot project



# Questions?

# FAIR data

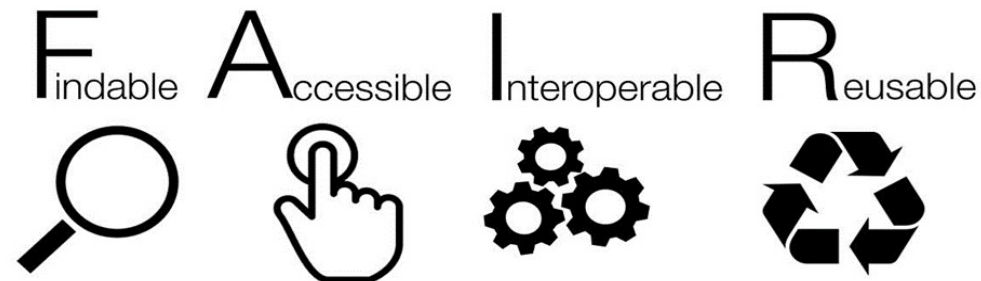


# Growing need for data FAIRness

Research and observation data should be more and more **FAIR**

“**Findable Accessible Interoperable Reusable**” for machines

- **Findable:** Available via the web, in catalogues, with good discovery metadata
- **Accessible:** From download to laptop, to machine readable data development
- **Interoperable:** Use **combined** data from different datasets in same format, with metadata
- **Reusable:** Measure once: use many times. Means reliable and traceable data.



# Growing need for data FAIRness

This is driven by:

- Marine Strategy Framework Directive Reporting
  - Reliable, trustworthy data
  - FAIR data does not guarantee this, but allows for quality assessment
- Data gaps
- Digital Twin developments
- Internet and cloud processing possibilities
  - VRE's/Jupyter notebooks, machine-2-machine access
- Research requirements for reproducibility and traceability
- Growing data availability and need for better findability of data
- And on the other side the cost of observation data and benefit of multiple use.

# The FAIR metrics F&A

Source: GO-FAIR

- **Findable**
  - F1. (Meta)data are assigned a globally unique and persistent identifier
  - F2. Data are described with rich metadata (defined by R1 below)
  - F3. Metadata clearly and explicitly include the identifier of the data they describe
  - F4. (Meta)data are registered or indexed in a searchable resource
- **Accessible**
  - A1. (Meta)data are retrievable by their identifier using a standardised communications protocol
    - A1.1 The protocol is open, free, and universally implementable
    - A1.2 The protocol allows for an authentication and authorisation procedure, where necessary
  - A2. Metadata are accessible, even when the data are no longer available

# The FAIR metrics I&R

- **Interoperable => think here: domain specific metadata and vocabulary solutions**
- I1. (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
- I2. (Meta)data use vocabularies that follow FAIR principles
- I3. (Meta)data include qualified references to other (meta)data
- **Reusable => the biggest challenge for achieving reliability assessment**
- R1. (Meta)data are richly described with a plurality of accurate and relevant attributes
- **Context, purpose of data collection, lab conditions, parameter settings, etc.**
- R1.1. (Meta)data are released with a clear and accessible data usage license
- R1.2. (Meta)data are associated with detailed provenance
- **Where does the data come from, who collected it, how was it processed, which original data was included.**
- R1.3. (Meta)data meet domain-relevant community standards
- **Formal data standard, combine various datasets**

# How to achieve FAIRness?

## Main steps

- Make the data Findable and Accessible with persistent identifiers in well accessible catalogues
- Create extensive metadata following **EU community standards, including community endorsed vocabularies and quality information**
- Create community supported data exchange formats
- Add a usage license

## Validation

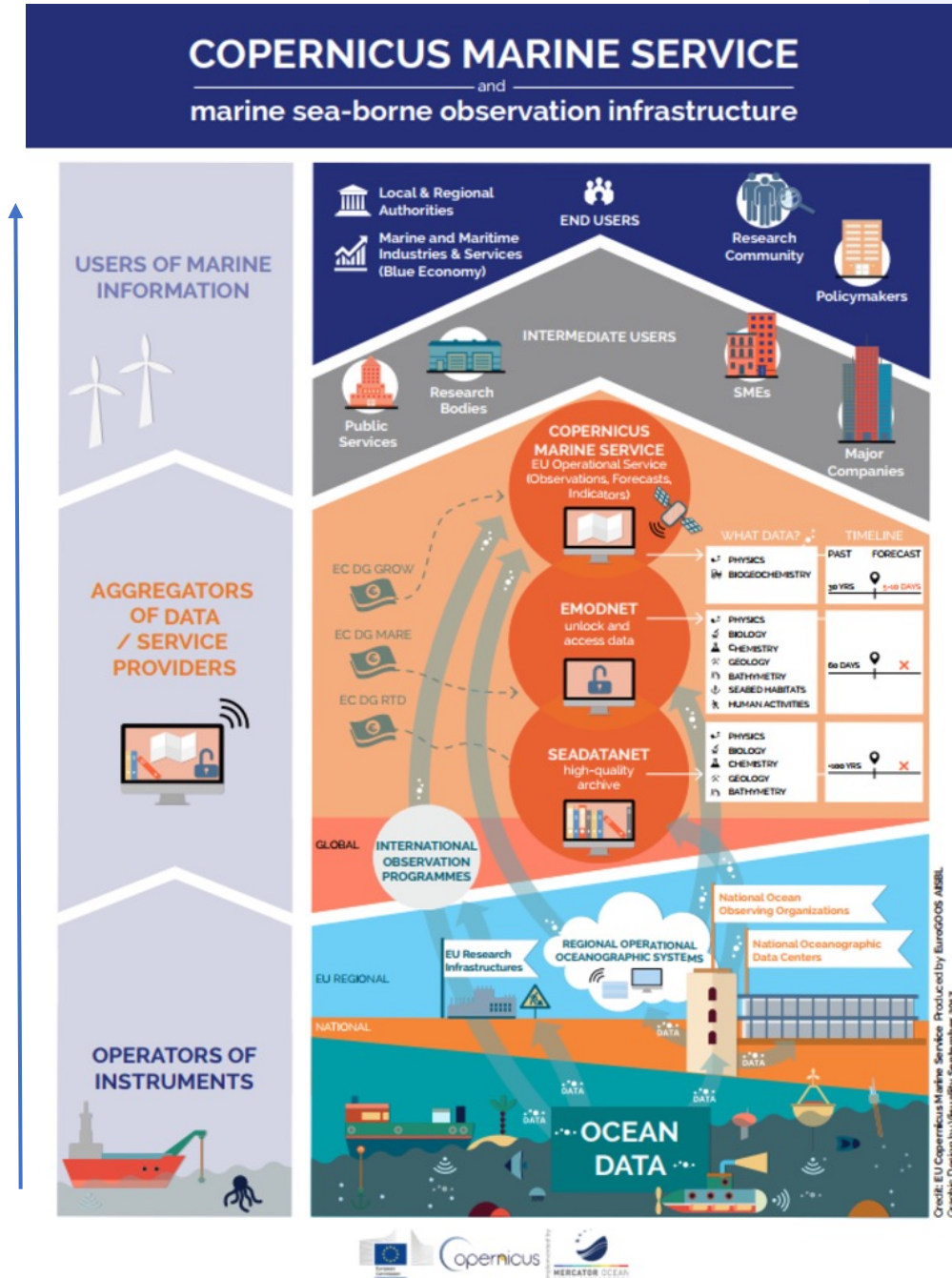
- FAIRness process steps guidance => [FIP wizard](#)

# Data landscape at EU level

# European landscape of marine data management

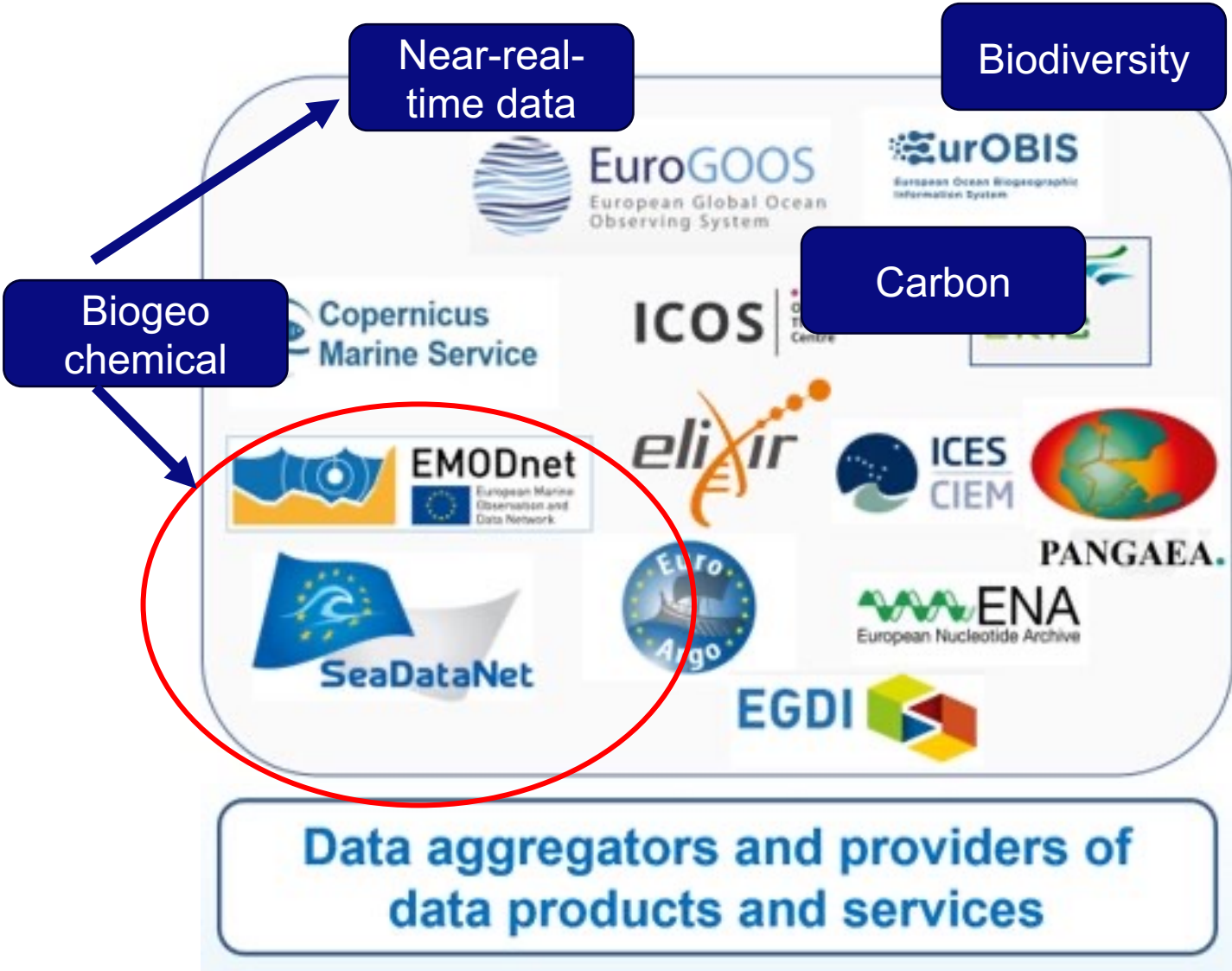
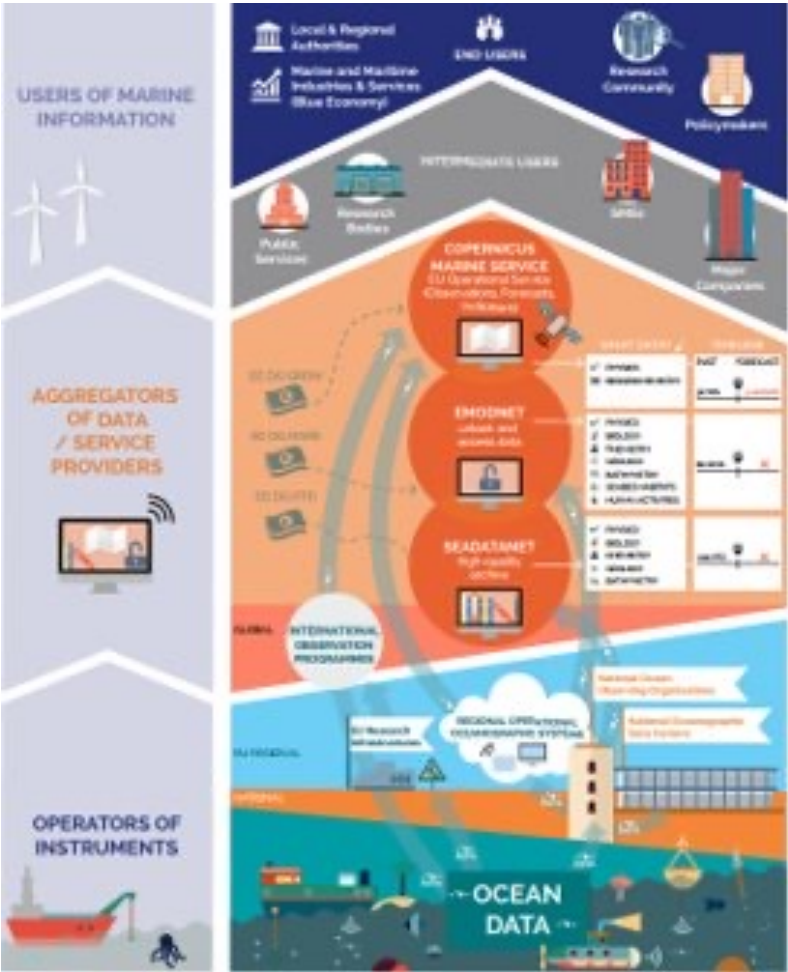
Flow of data:

- Observation platforms
- National/thematic data centers (First use)
- Aggregators at EU level (secondary "re-use")
- Intermediate users, consultants
- End-users: Policy makers, students, researchers





# Thematic organisation in aggregators



# How to achieve FAIRness for Biogeochemical data?

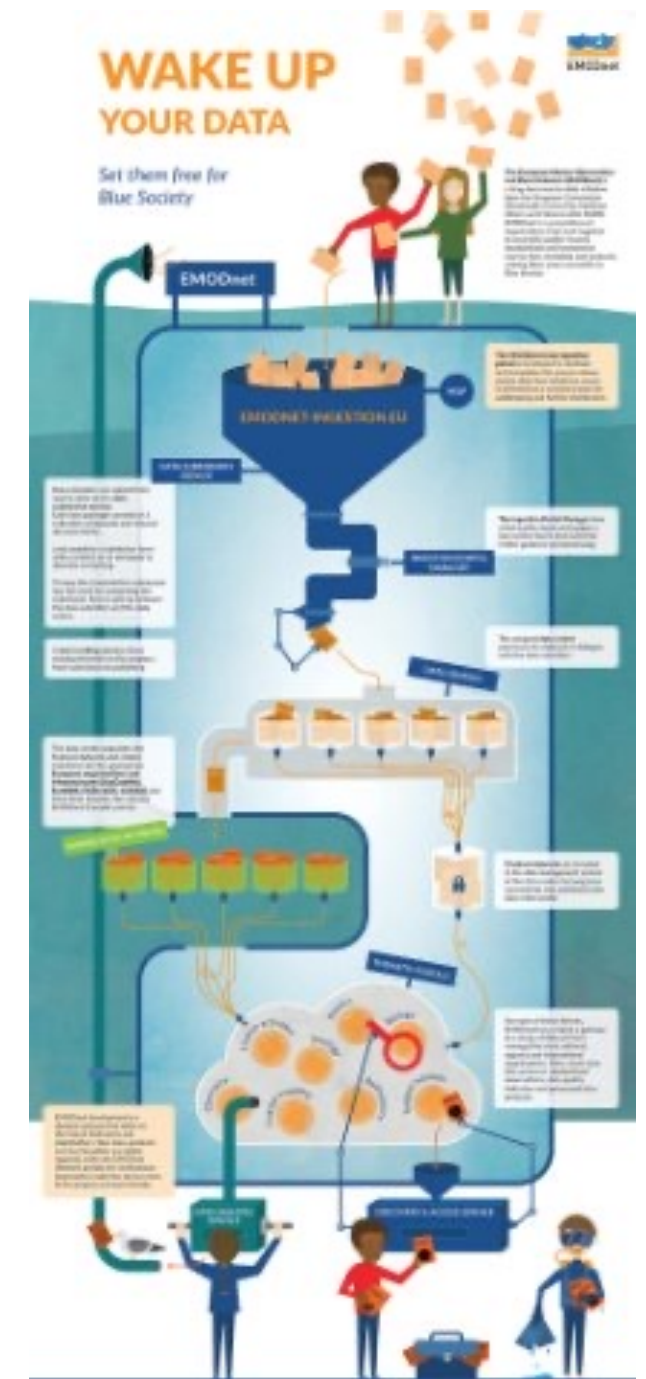
- Provide data to SeaDataNet:
  - Become a SeaDataNet node, or provide data to existing nodes in your country (e.g. National marine data centers)
- Publish metadata:
  - following SeaDataNet community metadata template, including community endorsed vocabularies
  - Vocabularies: "codelists" to describe e.g. parameters, instrument, organisation, etc
- Publish the dataset:
  - Following data exchange formats
  - ODV ASCII or NetCDF CF standard (using again vocabularies for e.g. parameters, units)
- Add a usage license: CC-BY 4.0 preferred, but more restricted is possible

# Streamlined workflow to European portals

**What if you are not connected to National Oceanographic Data centres but still want to publish on the European portals?**

Possible solution: Submit datasets to [EMODNet Data ingestion](#)

- Creates a DOI on the dataset
- Allows for submitting any type of data via a standardized template
- Publishes your datasets 'as-is' with metadata in a central catalogue
- If the datasets are useful for certain EMODnet products the data will be reformatted and checked and possibly flow into any of the EMODnet lots





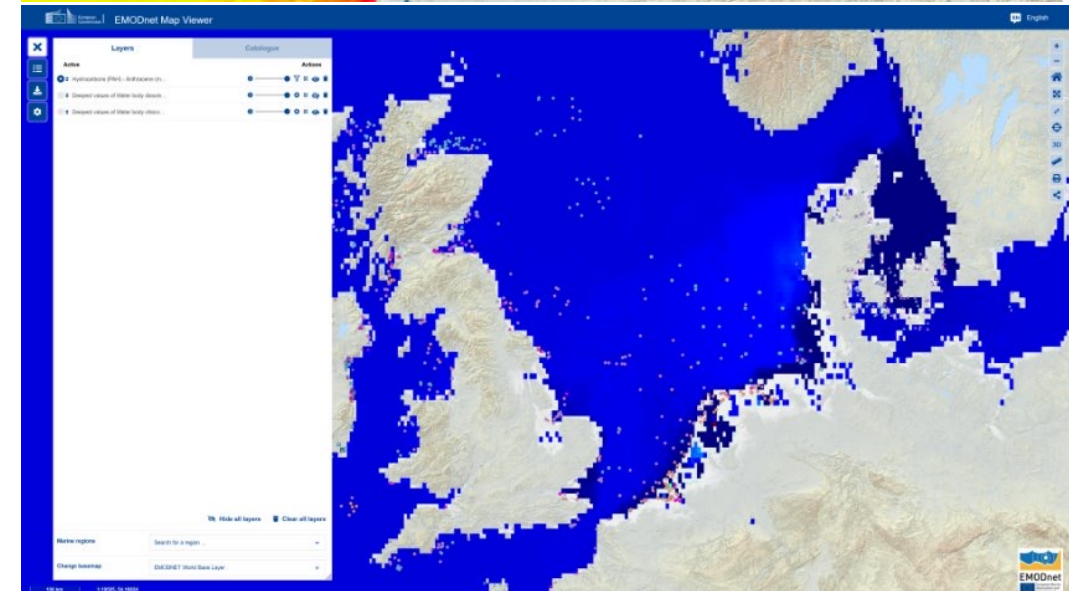
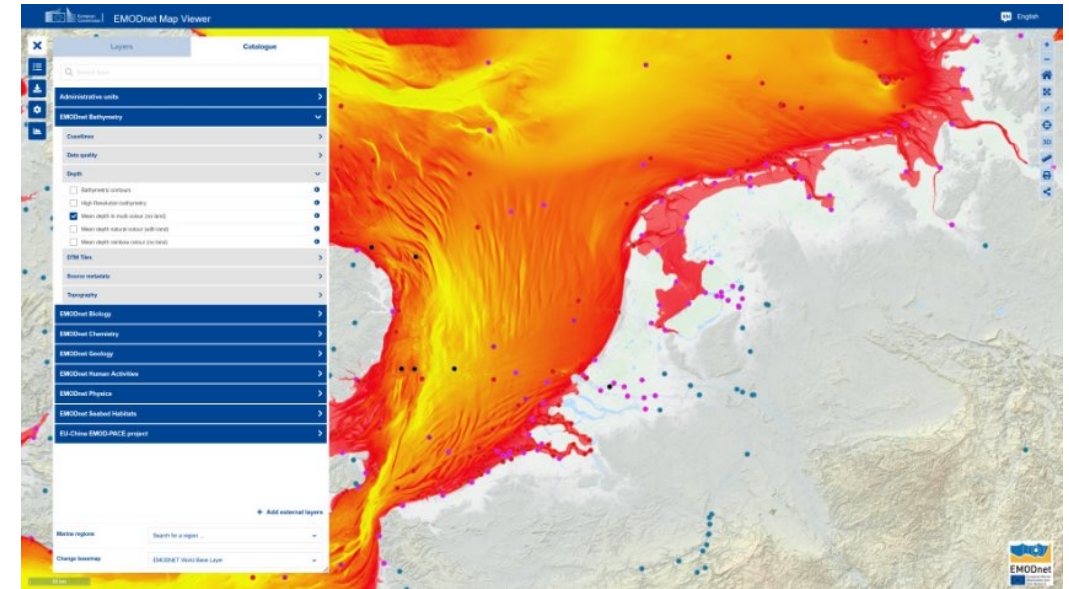
# End goal: Data uptake



[www.seadatanet.org](http://www.seadatanet.org)

<https://emodnet.ec.europa.eu>

All data in products, FAIR,  
and traceable back to original files, platforms etc.



# More detail: How to achieve FAIRness for Biogeochemical data?

- Make the data Findable and Accessible with persistent identifiers in well accessible catalogues
- Become a [SeaDataNet node](#), or provide data to existing node (e.g. National marine data centers)
- Publish data in SeaDataNet ([CDI system](#)), EMODnet, and other relevant catalogues
- Create metadata following community standards, including community endorsed vocabularies
- SeaDataNet community standard: [CDI metadata format](#) (ISO19115 based, plus profile)
- [SeaDataNet vocabularies](#), and [metadata directories](#), supporting the metadata/data format
- For: Parameters, instruments, organisations, projects, platforms, vessels, etc.
- Create community supported [data exchange formats](#)
- ODV ASCII or NetCDF CF standard (using again SDN vocabularies for e.g. parameters, units)
- Add a usage license: CC-BY 4.0 preferred, but more restricted is possible

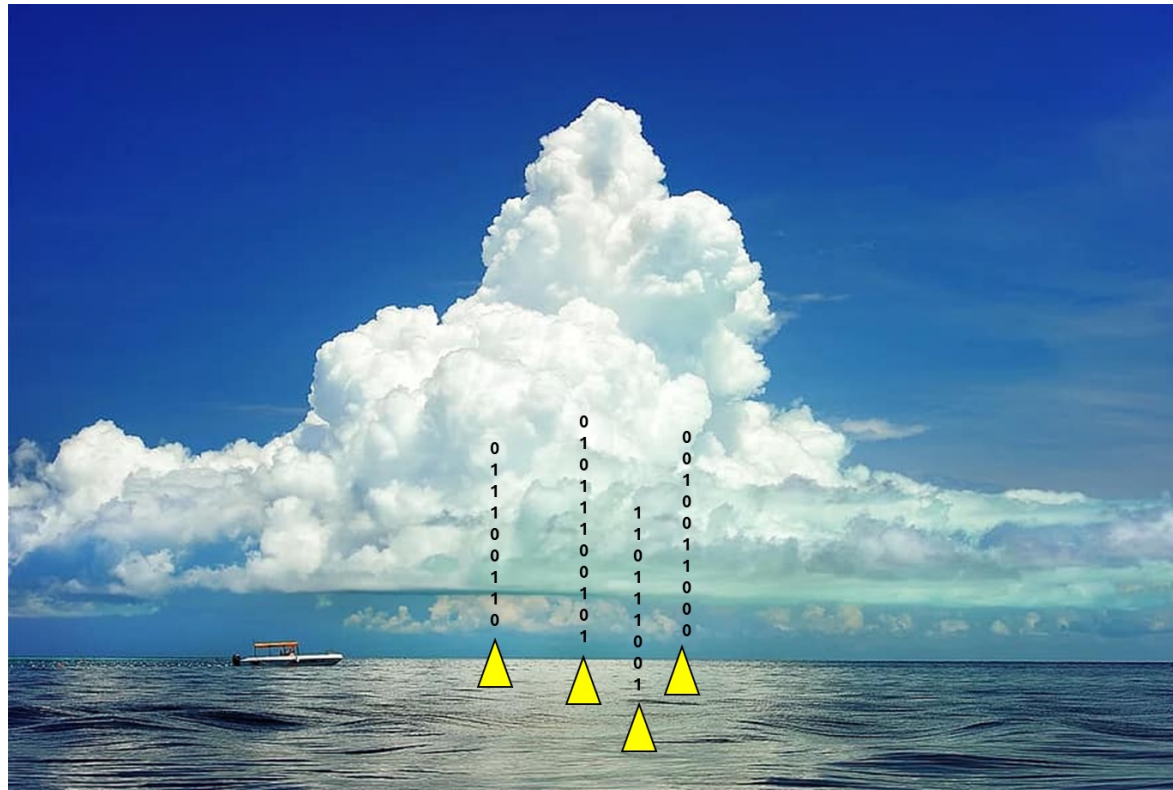
# Questions?



Let us know your experiences

Mentimeter code: 7116 2441





# Closure and follow-up