



Research Data Management

Simona Simoncelli

IR0000032 – ITINERIS, Italian Integrated Environmental Research Infrastructures System
(D.D. n. 130/2022 - CUP B53C22002150006) Funded by EU - Next Generation EU PNRR-
Mission 4 "Education and Research" - Component 2: "From research to business" - Investment
3.1: "Fund for the realisation of an integrated system of research and innovation infrastructures"



Outline

- 🌐 Datascape
- 🌐 the importance of data management
- 🌐 Marine Data Value Chain
- 🌐 Blue Data Infrastructures
- 🌐 Data Life Cycle and Dataflow
- 🌐 Data and Products' Quality
- 🌐 FAIR principles and FAIRness assessment
- 🌐 Data Publication
- 🌐 workflow and Virtual Research Environment



UN Decade of Ocean Science for Sustainable Development 2021-2030

Expected Outcomes

1. **A clean ocean** where sources of pollution are identified, reduced or removed
2. **A healthy and resilient ocean** where marine ecosystems are understood, protected, restored and managed
3. **A productive ocean** supporting sustainable food supply and a sustainable ocean economy
4. **A predicted ocean** where society understands and can respond to changing ocean conditions
5. **A safe ocean** where life and livelihoods are protected from ocean-related hazards
6. **An accessible ocean** with open and equitable access to data, information and technology
7. **An inspiring and engaging ocean** where society understands and values the ocean in relation to human wellbeing and sustainable development

→ **Data and information are key enablers of the Ocean Decade outcomes**

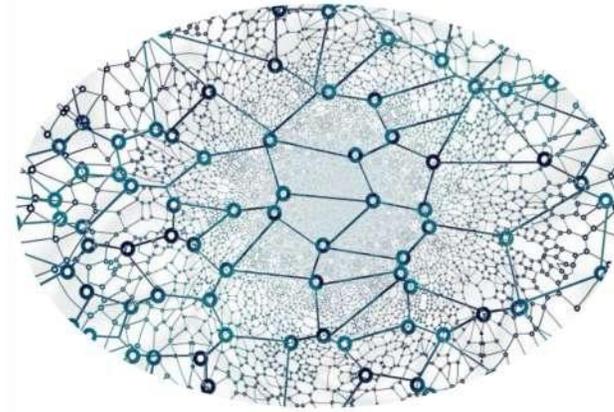
from Implementation Plan (<https://www.oceandecade.org/>)

Ocean Monitoring

An **healthy, resilient and productive ocean** can be restored and preserved through constant and **systematic monitoring of the marine environment** which allows to:

- assess past and present ocean state
 - understand its complex dynamics and evolution
 - understand/mitigate the impact of human activity in a changing environment and climate
 - exploit its natural resources sustainably
- the **management of the deriving marine data** represents a societal priority
- **data sharing approach and open data policies** maximize users' uptake, promoting knowledge and innovation, reducing costs, avoiding duplication of efforts
- adoption of **common vocabulary, formats and standards** enable a rapid and efficient access to the data
- **FAIR principles** (Findable Accessible Interoperable and Reusable - *Wilkinson et al., 2016*) and **linked data approach** are the foundation of this revolution to reduce uncertainties but preserving transparency and traceability

Ocean Decade: digital ecosystem



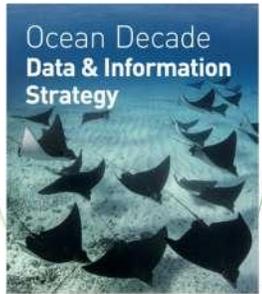
Data Strategy Implementation Group

→ **Data & Information Strategy** describes how data systems can co-create a distributed, robust and collaborative digital **ecosystem** to be actively used for decision making and to support sustainable ocean management

3 main components:

1. observation and data collection
2. data management and sharing
3. data processing (analytics, modeling and prediction)

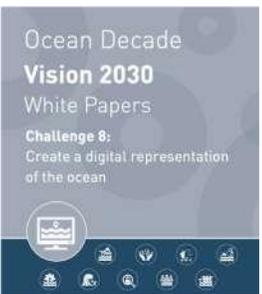
UNESCO-IOC (2023). Ocean Decade Data & Information Strategy. Paris, UNESCO. (The Ocean Decade Series, 45)



The United Nations
Decade of Ocean Science
for Sustainable Development
2021-2030



The United Nations
Decade of Ocean Science
for Sustainable Development
2021-2030



The United Nations
Decade of Ocean Science
for Sustainable Development
2021-2030

Ocean Decade: digital ecosystem

Vision

A trusted, inclusive, and interconnected ocean data and information ecosystem that is actively used for decision making to support sustainable ocean management.

Mission

To catalyse a solution-oriented, global digital transformation for the digital ecosystem we need to overcome the Decade Challenges.

Strategic Objectives

1

Develop an ocean digital ecosystem that encourages the sharing and equitable access of multidisciplinary data, information and knowledge by all.

2

Improve data and information discovery and usability across the ocean digital ecosystem.

3

Build trust in data and information shared across the ocean digital ecosystem.

4

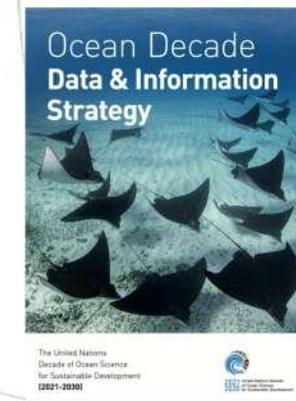
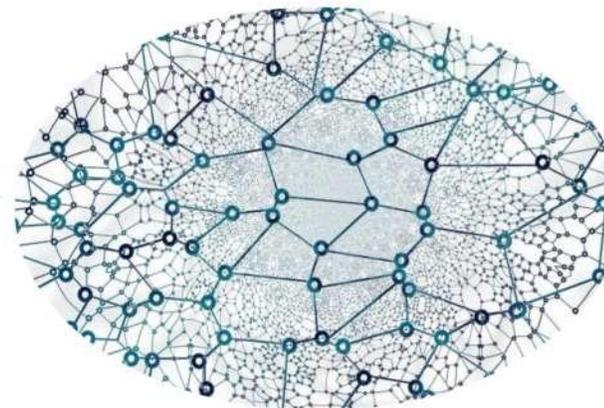
Prioritise digital solutions that support decisions for sustainable ocean management.

5

Expand, empower, and mobilise global communities to advance and maintain the ocean digital ecosystem.

Enablers

Technological Innovation // Partnerships // Durable Resourcing // Policy & Regulatory Frameworks



UNESCO-IOC (2023). Ocean Decade Data & Information Strategy. Paris, UNESCO. (The Ocean Decade Series, 45)

Digital Ecosystem Drivers

- 🌐 **Open science** and **FAIR** (Findable, Accessible, Interoperable and Reusable) principles drive the development of a digital ocean ecosystem capable of supporting the preservation and sharing of a wide array of data, knowledge and information products
- 🌐 **Ocean Best Practices** and a **FAIR digital asset** facilitate technological advances and community approaches for all methods to better understand and sustain our oceans
- 🌐 **Virtual Research Environments** will provide to the new generation of scientists an easy access to big amount of data, analytics and HPC resources to create innovative applications and tackle the emerging societal challenges (DTO - Digital Twin of the Ocean)

Need of a Community Effort

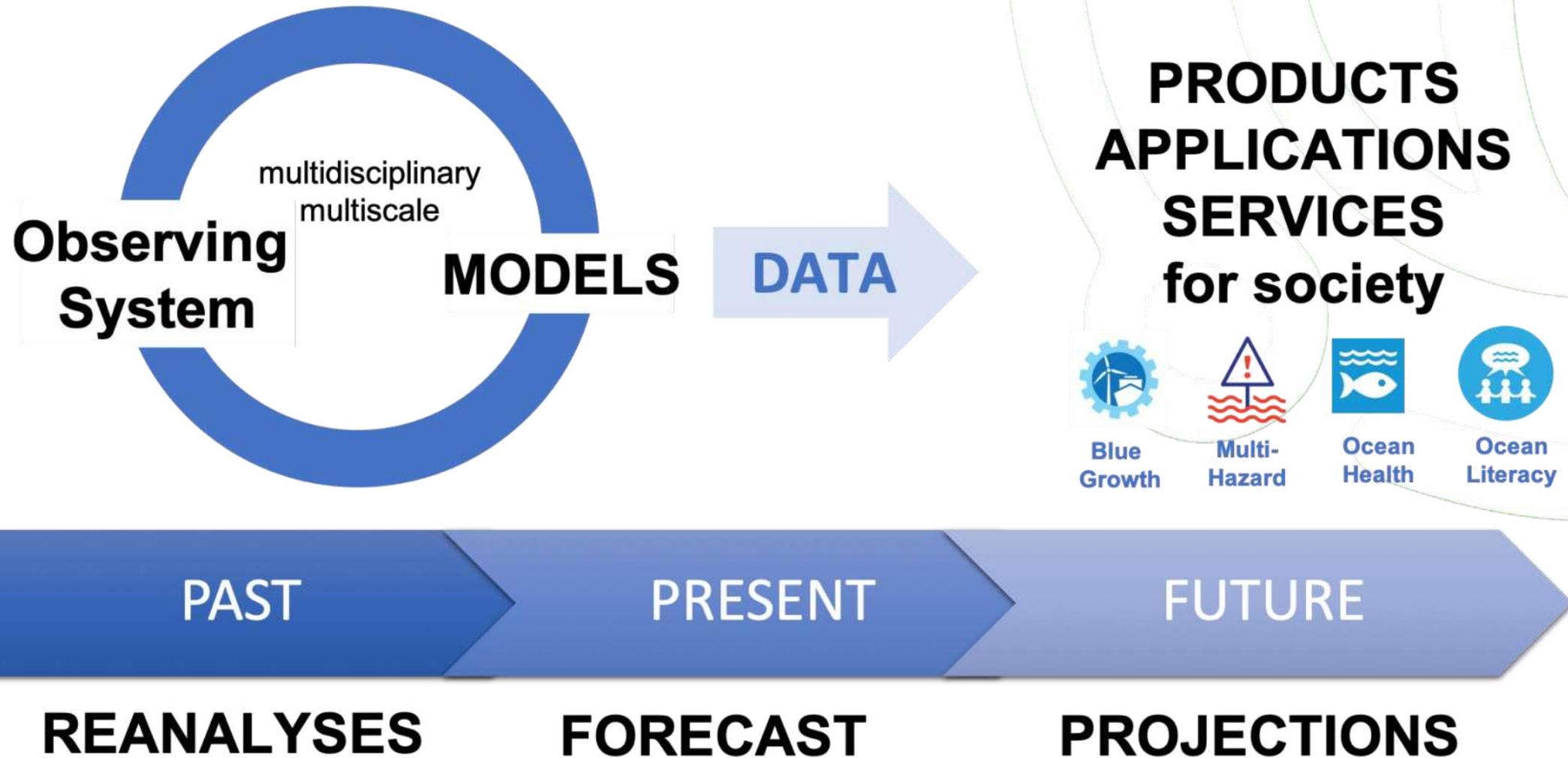
Societal benefit of science can be enhanced through a community effort to collect, manage, and share the data acquired with specific purposes and to enable multiple stakeholders to reuse them many times and for different purposes

Ocean knowledge and information rely on an **integrated multidisciplinary and multiscale ocean observing and monitoring system**

This implies the **engagement of multiple stakeholders along with the data value chain**

- to continuously tailor data products **requirements** to the emerging societal needs
- to identify **gaps**
- to optimize the data flow

Integrated Observing and Monitoring System



Simoncelli et al. 2022 <https://doi.org/10.1016/B978-0-12-823427-3.00001-3>

Global Ocean Observing System (GOOS)

<https://goosocean.org>

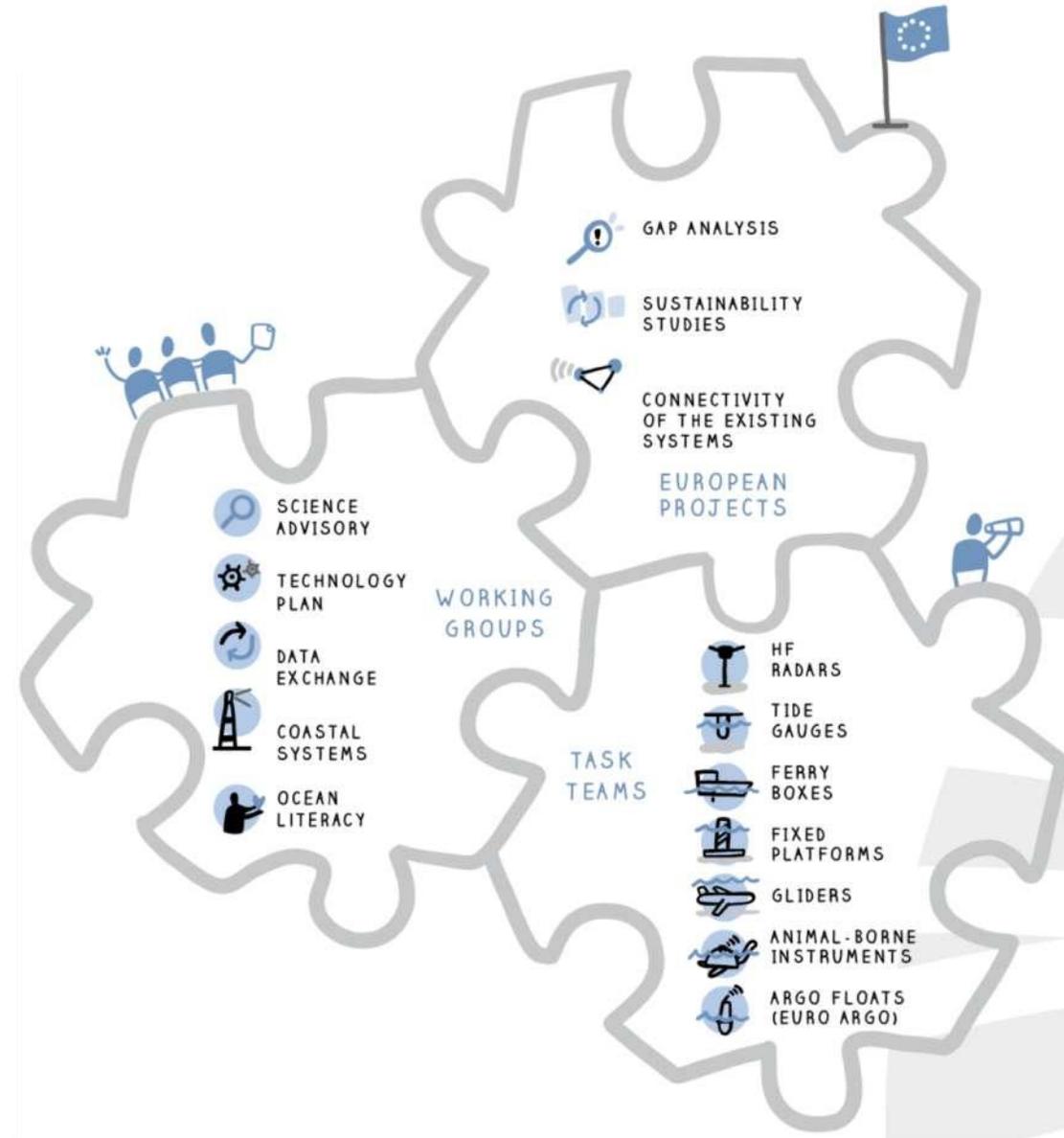
- 🌐 led by the Intergovernmental Oceanographic Commission (IOC) of UNESCO, and co-sponsored by the World Meteorological Organization (WMO), the United Nations Environment Programme (UNEP) and the International Science Council (ISC) to support a community of international, regional and national ocean observing programmes, governments, UN agencies, research organizations and individual scientists
- 🌐 **goal: implementation of an integrated and sustained ocean observing system to deliver maximum impact for our user base and society**
- 🌐 with **systems approach** designed to be flexible and adapt to evolving scientific, technological and societal demands
- 🌐 fit-for-purpose observing system: data flow on Essential Ocean Variables (EOVs), information services and applications that benefit the end users

- 🌐 There are currently ~ 8000 in situ ocean observing platforms monitored by the GOOS operational centre OceanOPS
- 🌐 systems of buoys, moorings, profiling floats, ships (merchant, academic, fishing), gliders and other uncrewed systems (e.g., sail drones), drones, planes, land-based stations (e.g., global sea level), satellites, HF radars, deep sea cables, and animal borne sensors measuring mostly physical data and some biogeochemical data
- 🌐 14 global ocean observing networks that vary in stage of implementation
- 🌐 GOOS is organised in thirteen GOOS Regional Alliances (GRAs)



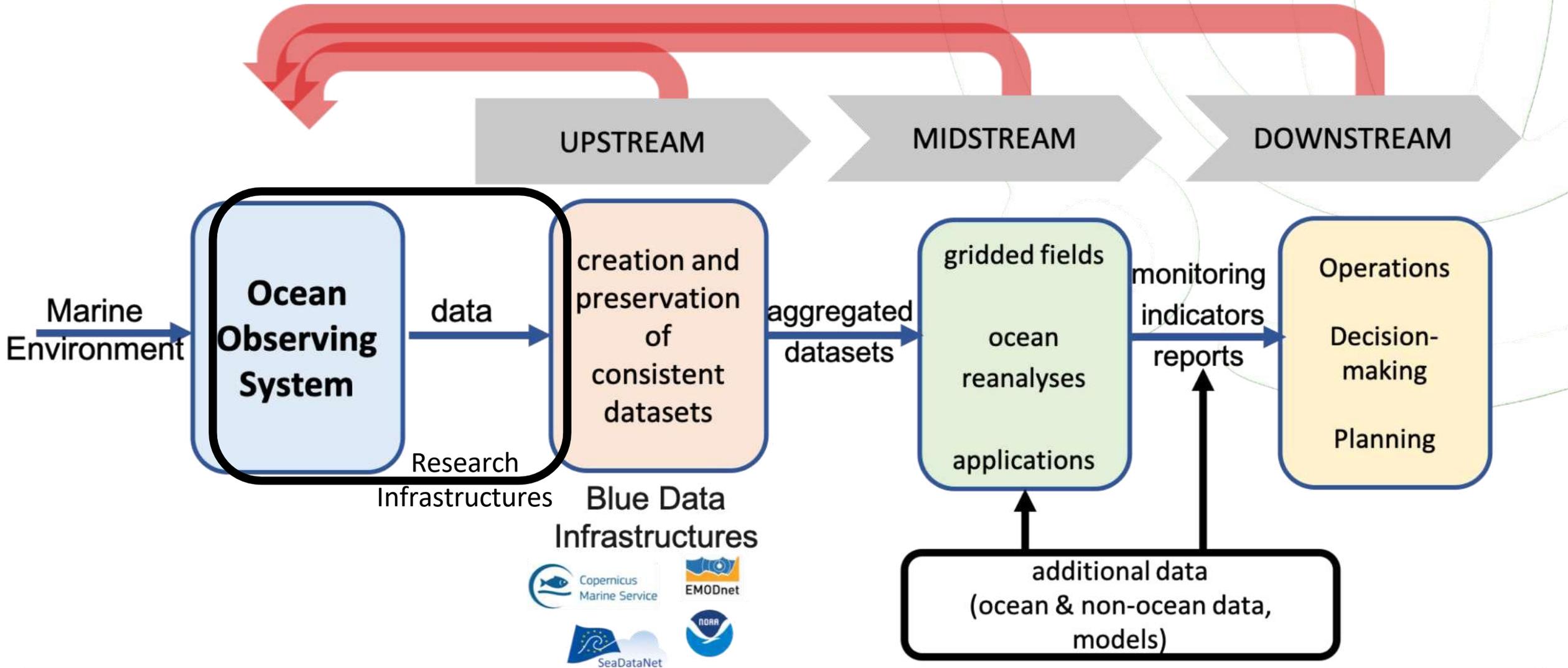
<https://doi.org/10.25607/brxb-kr45>

- 🌐 European component of the GOOS
- 🌐 48 members
- 🌐 5 Regional Systems
- 🌐 Working Groups → Data Management WG
- 🌐 networks of observing platforms (Task Teams)
- 🌐 cooperation to unlock quality marine data and deliver common strategies, priorities and standards towards an integrated, sustained and fit-for-purpose European Ocean Observing System (EOOS)
- 🌐 EOOS framework is driven by the European community to better coordinate Europe's ocean observing capacity



Marine Data Value Chain

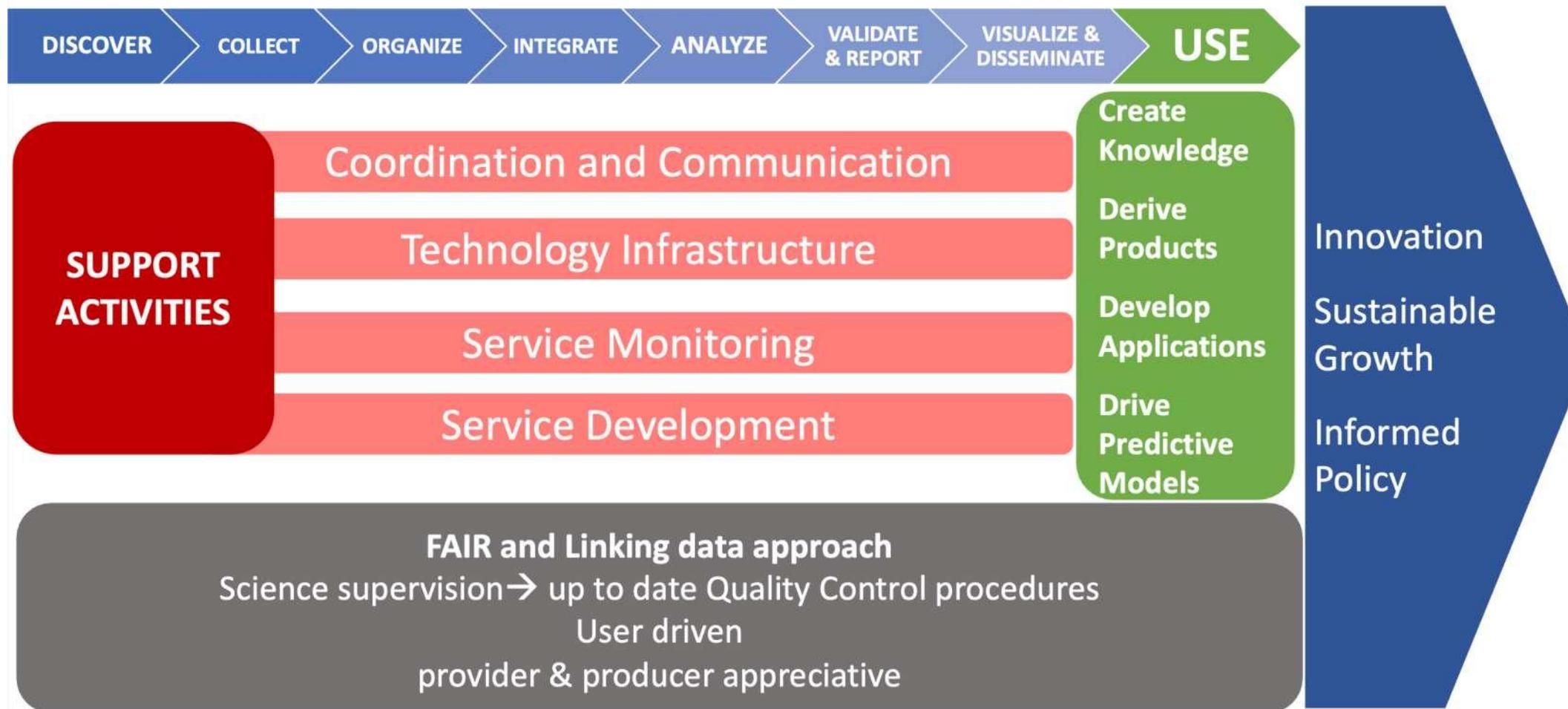
Simoncelli et al. 2022 <https://doi.org/10.1016/B978-0-12-823427-3.00001-3>



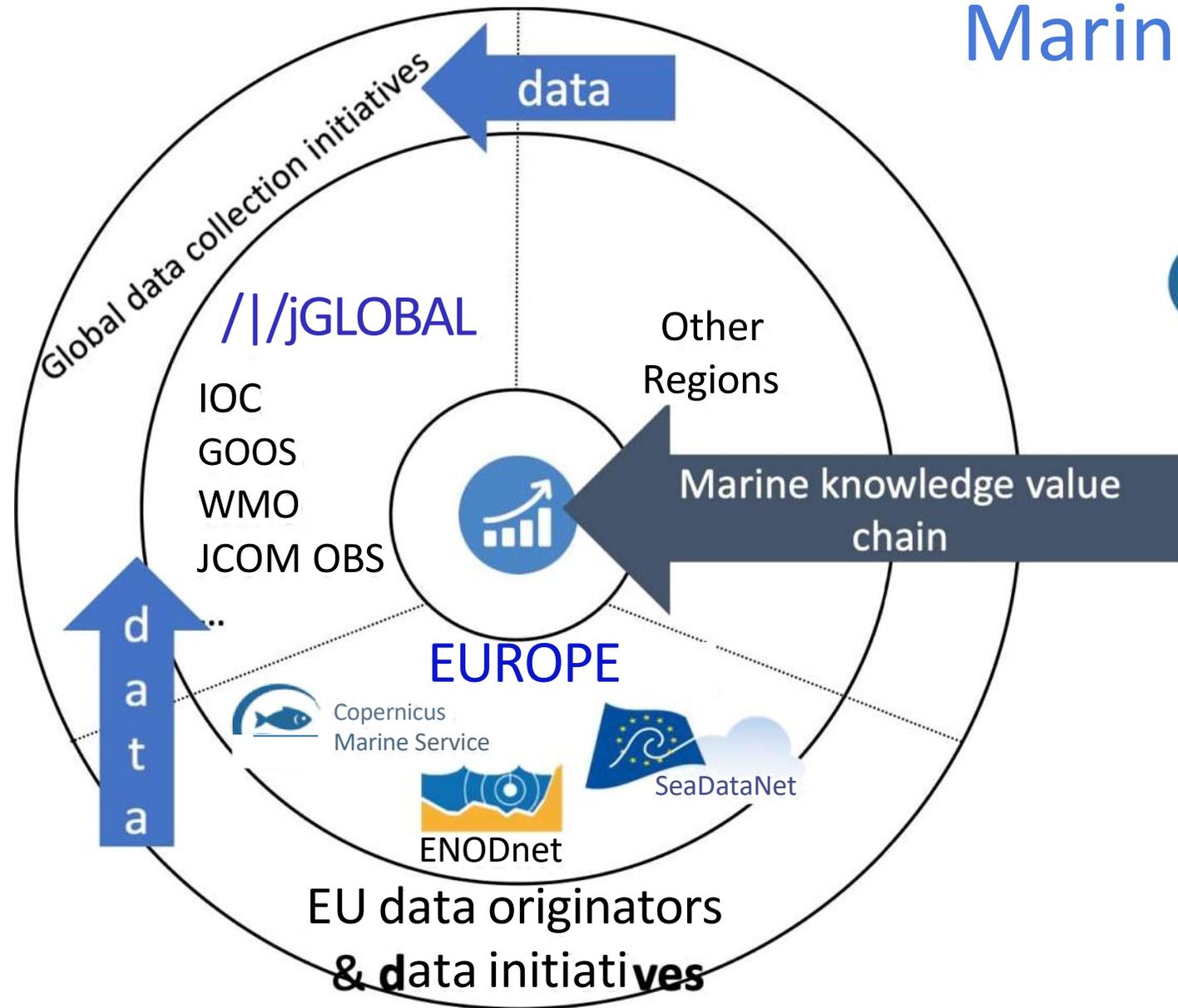
Blue Data Infrastructures

Marine data value chain: the necessary supporting activities to fulfill it and the underlying principles

Simoncelli et al. (2022) <https://doi.org/10.1016/B978-0-12-823427-3.00001-3>



Marine Data and Data Products Initiatives



DG-GROW

provides NRT&DM observations and predictions on the status and dynamics of the ocean and the ecosystem



DG-RTD

distributed Marine Data Infrastructure for the management of DM in situ data sets; it develops standards and vocabulary; it provides services/tools and data products



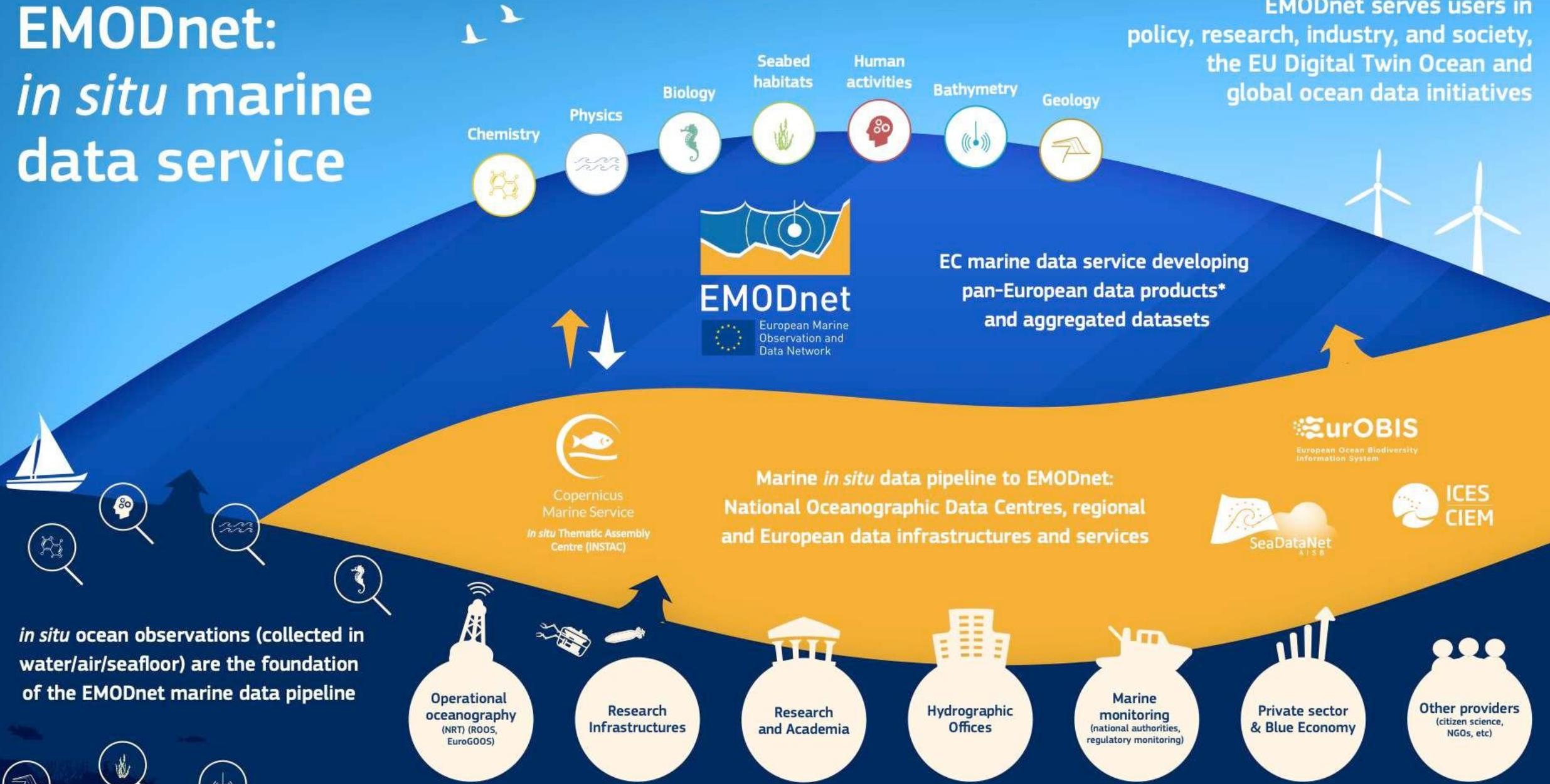
DG-MARE

organizes marine data in a unique infrastructure, deriving information from data, creating products targeted to multiple user communities

adapted from *Martin Miguez et al. (2019)*

EMODnet: *in situ* marine data service

EMODnet serves users in policy, research, industry, and society, the EU Digital Twin Ocean and global ocean data initiatives



*All EMODnet data products are published under an open data licence

One central map viewer

to visualise all EMODnet data



1 OCEAN 1 EMODnet

One single portal

One central metadata catalogue

to enhance data search and discovery

120
partners

+100
use cases

Discover, visualise and
download marine data and products
across 7 thematics and hundreds of parameters



BATHYMETRY



HUMAN ACTIVITIES



PHYSICS



GEOLOGY



SEABED HABITATS



CHEMISTRY



BIOLOGY

EMODNET.EC.EUROPA.EU

YOUR GATEWAY TO *IN SITU* MARINE DATA IN EUROPE AND BEYOND



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.





Copernicus Marine Service

Providing free and open marine data and services to enable marine policy implementation, support Blue growth and scientific innovation.

Access Data >

DATA

OCEAN PRODUCTS

A robust ocean data catalogue, to download or visualise data including hindcasts, nowcasts and forecasts.

EXPERTISE

OCEAN STATE REPORT

Extensive annual analysis on the state of the ocean over nearly 20 years and severe/notable annual events.

TRENDS

OCEAN CLIMATE TRENDS

Monitoring the health of the ocean.
[Ocean Monitoring Indicators](#)
[Ocean Climate Portal](#)

EXPLORATION

OCEAN VISUALISATION

Dive into our 4D digital oceans through our 3 visualisation tools for beginner, intermediate and advanced users



INSTAC Dashboard

Platforms Type & Status

Choose the type of platform you are looking for:

High Frequency Radars (HF)

Moorings (MO) River Flows (RF)

Tide Gauges (TG)

Profilers (PF) Gliders (GL)

Drifters (DB) Drifters (DC) Saildrones (SD)

Thermistor chains (TX)

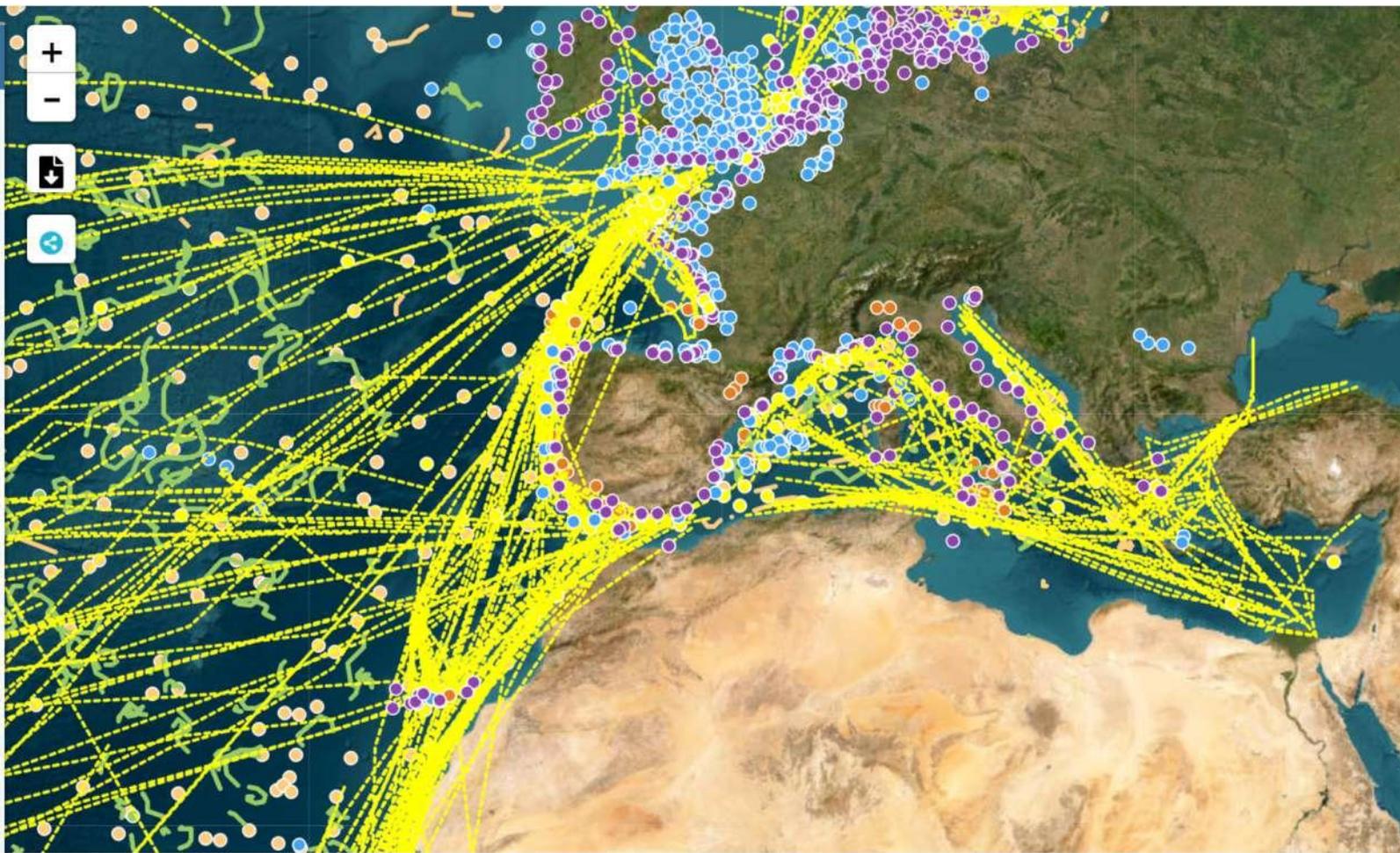
Ferrybox (FB) XBTs (XB) Mini Loggers (ML)

CTDs (CT) Thermosalinometer (TS)

Bottles (BO) Sea mammals (SM)

Filter also by status of the platform:

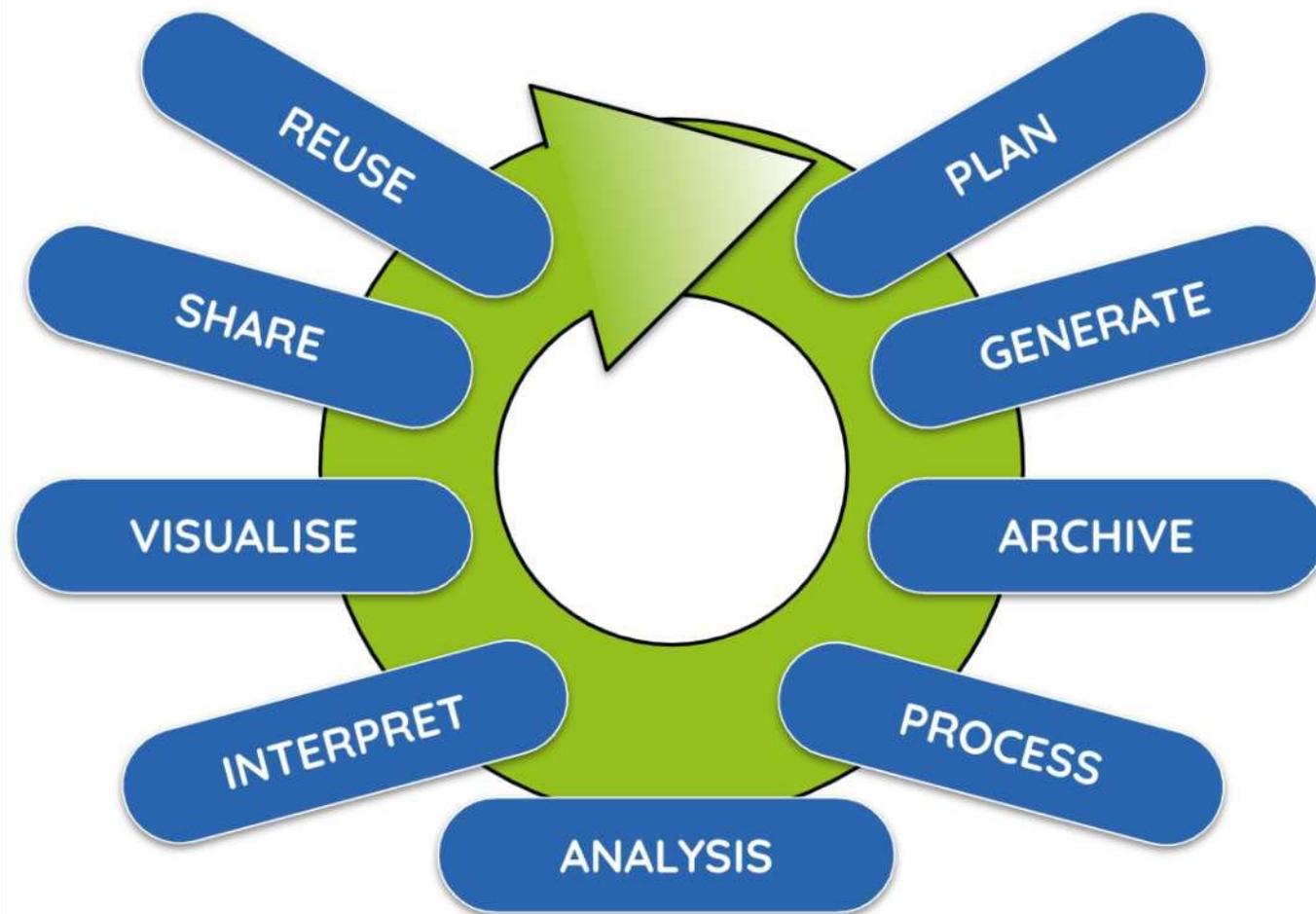
Copernicus In Situ TAC



<p>Volume of data</p> <p>19708 MB</p> <p>From last 30 days</p>	<p>Data providers</p> <p>0</p> <p>From last 30 days</p>	<p>Number active platforms</p> <p>0</p> <p>From last 30 days</p>	<p>Services availability</p> <p>~ 99.9%</p> <p>From last 30 days</p>
---	--	---	---

Research Data Life Cycle

adapted from Locati & Cacciola (2024) Research Data Management
<https://istituto.ingv.it/ufficio-gestione-dati>



Data Provider (PI) → data generation and responsible for QA/QC

Data Curators → data curation (annotation, publication, presentation and preservation)

Data Stewards → making data Interoperable

Data Managers → Governance and Data Management Plan



Data Management

from Locati & Cacciola (2024) Research Data Management <https://istituto.ingv.it/ufficio-gestione-dati>

Research Data Management (RDM)

Organization, storage, preservation, and sharing of data in a research environment throughout the **entire data life cycle and beyond**.

RDM ensures the efficiency, transparency, and reproducibility of research and include **practices and policies that aim to improve the quality and value of research data**.

Data Management Plan (DMP)

A formal document that outlines **how data is handled during and after a research project**.

Details how data is collected, analysed, preserved, and shared, including details about data formats, metadata standards, data storage, and data sharing policies.

It is **often a required component of grant proposals**, emphasizing the importance of responsible data management for funders and institutions.



Data collection: Quality Assurance and Best Practices

- 🌐 Traditional view of data (collecting, processing, analyzing, and publishing of results) substituted by a life cycle approach that highlights the importance of finding, storing, and sharing data
- 🌐 **Prior planning** has become mandatory to document data creation, content, context, but also to fulfill data quality requirements
- 🌐 Data quality requires predefined Quality Assurance (QA) strategies based on the selection of internationally validated methodologies for sampling and analysis, the mandatory use of reference materials

The data providers (PI)

- 🌐 are responsible of the adequacy of the sampling strategy to the scope
- 🌐 must follow specific QA procedures and protocols applied before and during the dataset creation
- 🌐 have responsibilities in terms of documentation, calibration/intercalibration exercise, sampling strategy, admissible ranges of data, algorithms used, corrections, and flags

QA/QC protocols and Standard Operating Procedures



Best practices have been developed from expert groups to have agreed and broadly adopted methods across ocean research, operations and applications

Best Practice:

- is a methodology that has repeatedly produced superior results relative to other methodologies with the same objective
- is method that has been adopted and employed by multiple organizations

<https://www.oceanbestpractices.org/>

A screenshot of the Ocean Best Practices System website. The page features a dark blue header with the "Ocean best practices" logo on the left and a search bar on the right. Below the header is a navigation menu with links for "ABOUT US", "NEWS AND EVENTS", "REPOSITORY", "COMMUNITY AND DEVELOPMENT", "OUR WORK", and "RESOURCES". The main content area has a background image of a sunset over water. The text reads "OCEAN BEST PRACTICES SYSTEM" in large white letters, followed by "Providing technological advances and community approaches for all ocean methods to better understand and sustain our oceans". Below this is a link for "OBPS WORKSHOP VIII, 14-18 OCT 2024 HERE : RECORDINGS AVAILABLE HERE". At the bottom, there are three white boxes with blue icons and text: "SEARCH FOR PRACTICES" (with a magnifying glass icon), "SUBMIT A PRACTICE" (with a group of people icon), and "EXPLORE OUR PROGRAMMES" (with a database cylinder icon).

Metadata

- 🌐 Records of monitored parameters must include a minimum set of information mapped through metadata
- 🌐 overview of the sensors and the methodologies (platform, instrument type, sensor's accuracy, calibration info)
- 🌐 measurement position, date and time
- 🌐 units
- 🌐 quality information (quality flags)

Standard and Formats

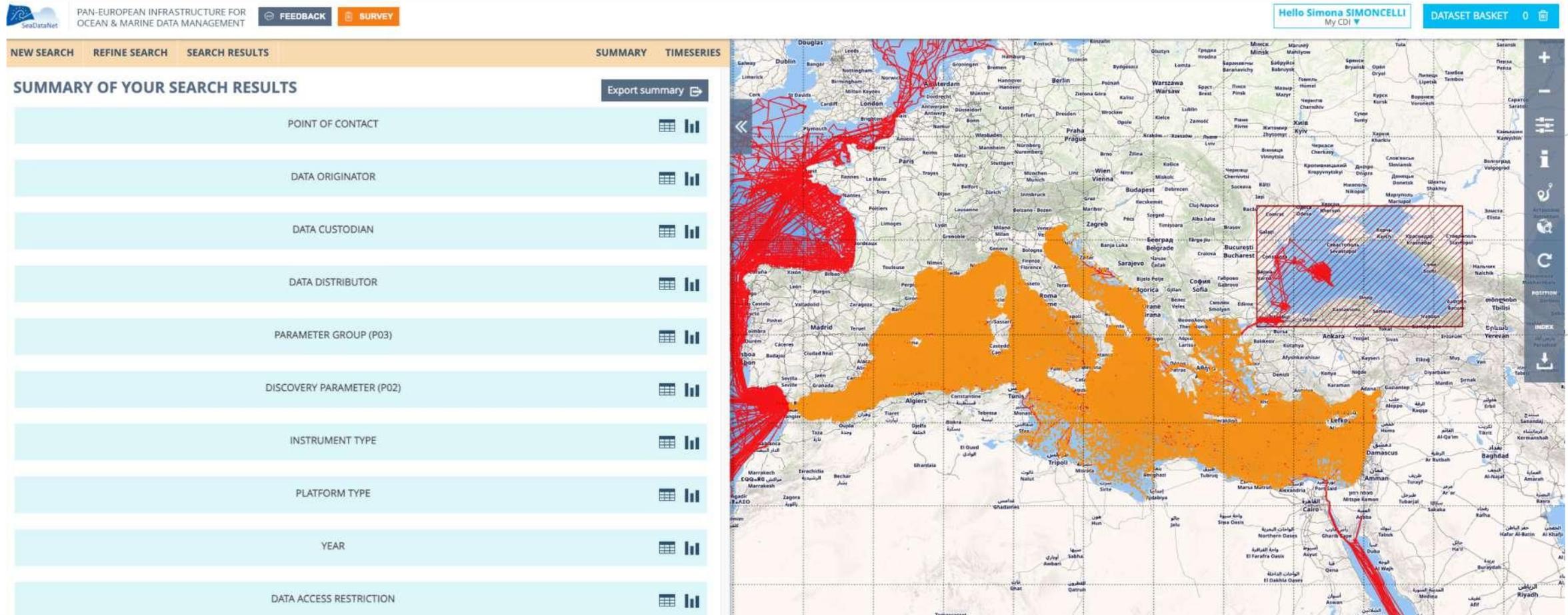
- 🌐 controlled vocabularies
- 🌐 ISO 19115 metadata standards
- 🌐 Data Transport Formats
- 🌐 common QC protocols and flag scales

The Power of Metadata

<https://cdi.seadatanet.org/search>



Metadata analysis can be used to monitor the **data sharing monitoring landscape**: data originator/distributor, instrument type, platform type, year, data access restrictions

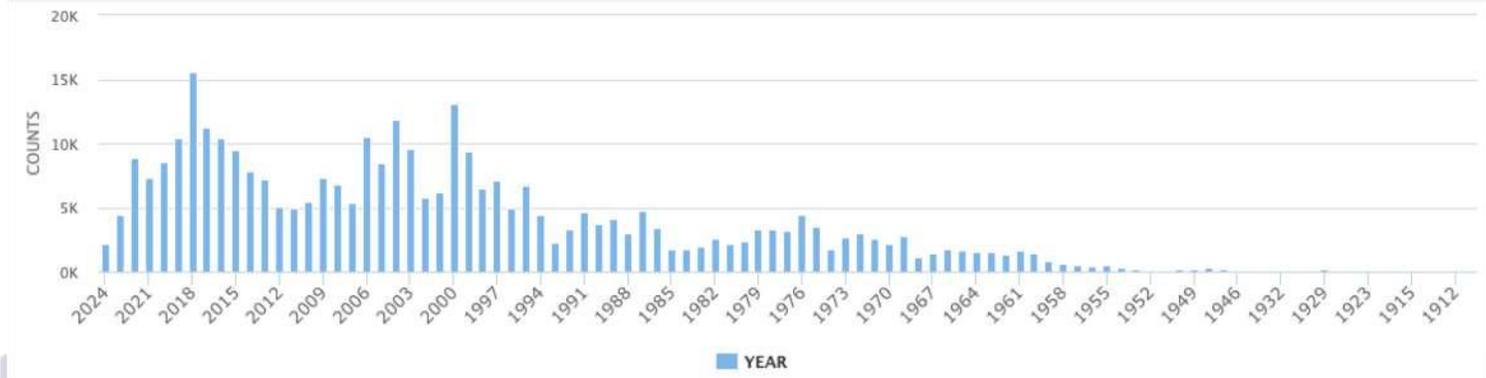
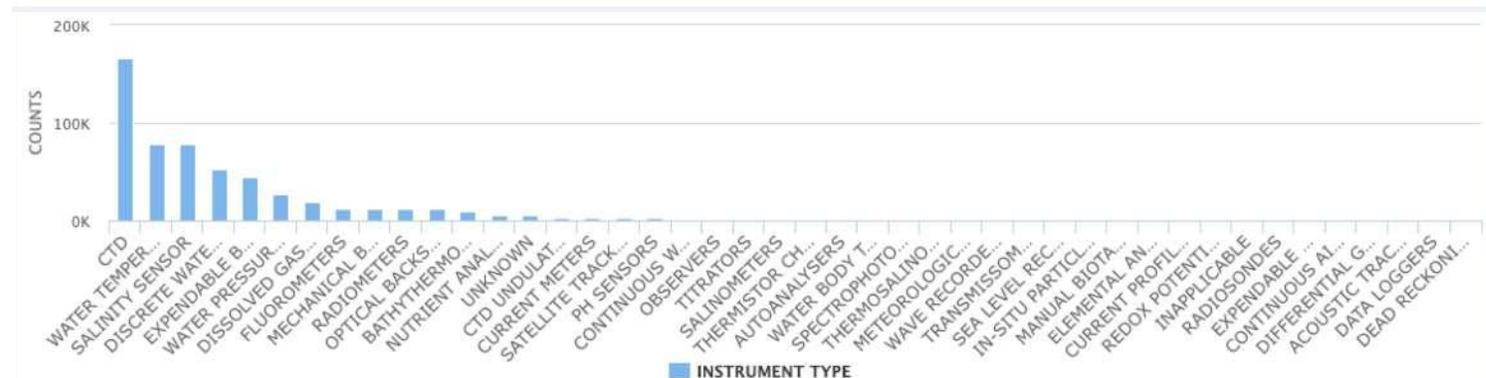
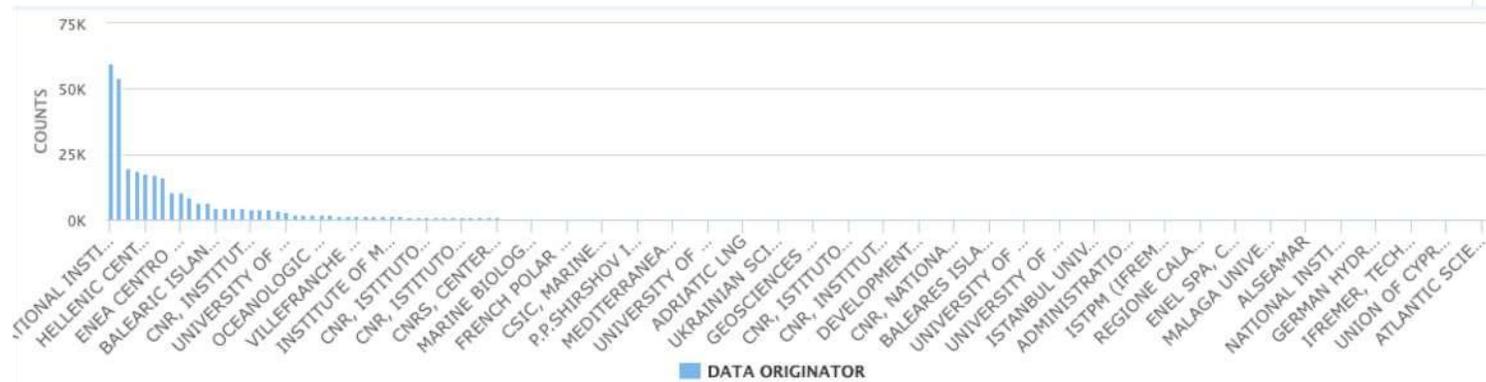


The Power of Metadata

<https://cdi.seadatanet.org/search>

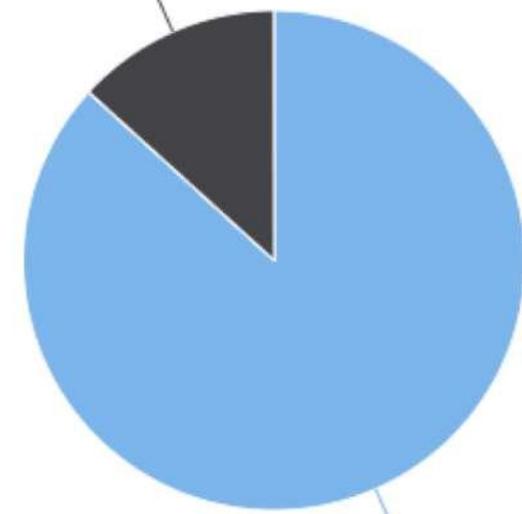


Metadata analysis from SeaDataNet data access portal on the specific data query results (i.e selected space-time domain, parameter)



DATA ACCESS RESTRICTION

43668: (43668)



286380: (286380)

FAIR data principles

from Locati & Cacciola (2024) Research Data Management <https://istituto.ingv.it/ufficio-gestione-dati>

Findable **A**ccessible **I**nteroperable **R**eusable

Wilkinson et al. (2016)



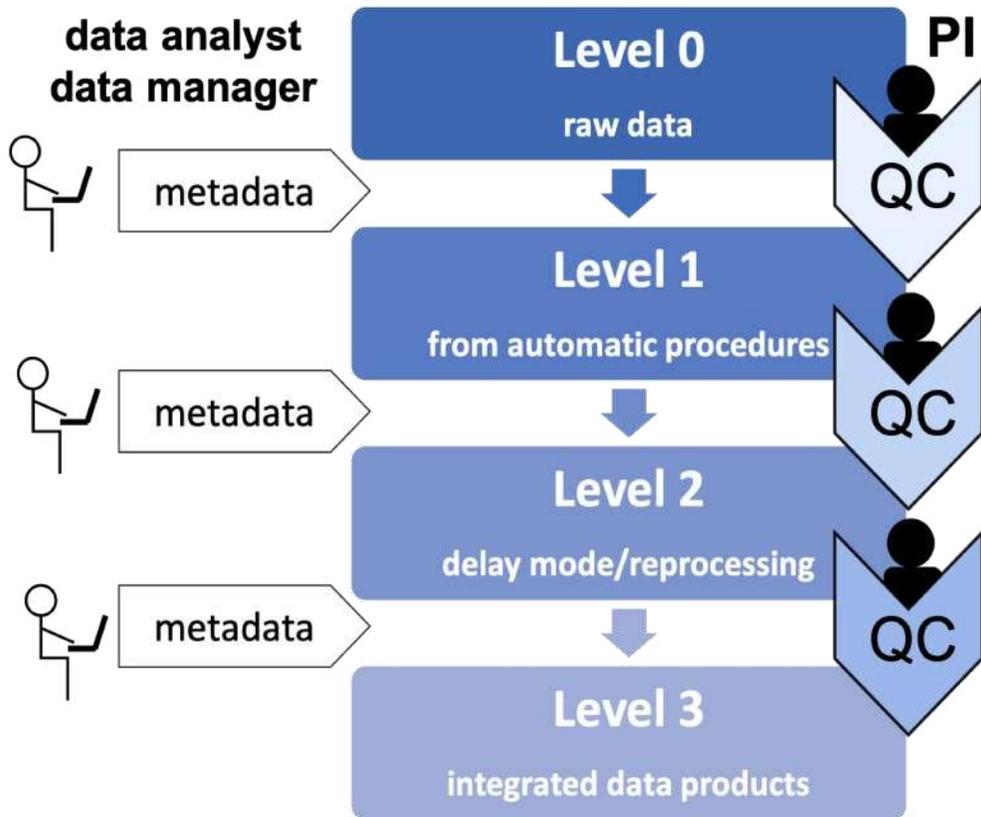
Findable: easy to find by humans and computers thanks to metadata and unique persistent identifiers

Accessible: stored for easy access and downloading

Interoperable: ready to be combined with other datasets by humans and computer systems

Re-usable: ready for reuse thanks to detailed, accurate documentation and clear usage license

Data Processing Levels (DPL)



Data Latency

Real Time

Near Real Time

Delay Mode



Purpose

- ocean prediction
- early warning
- ocean state assessment
- climate studies

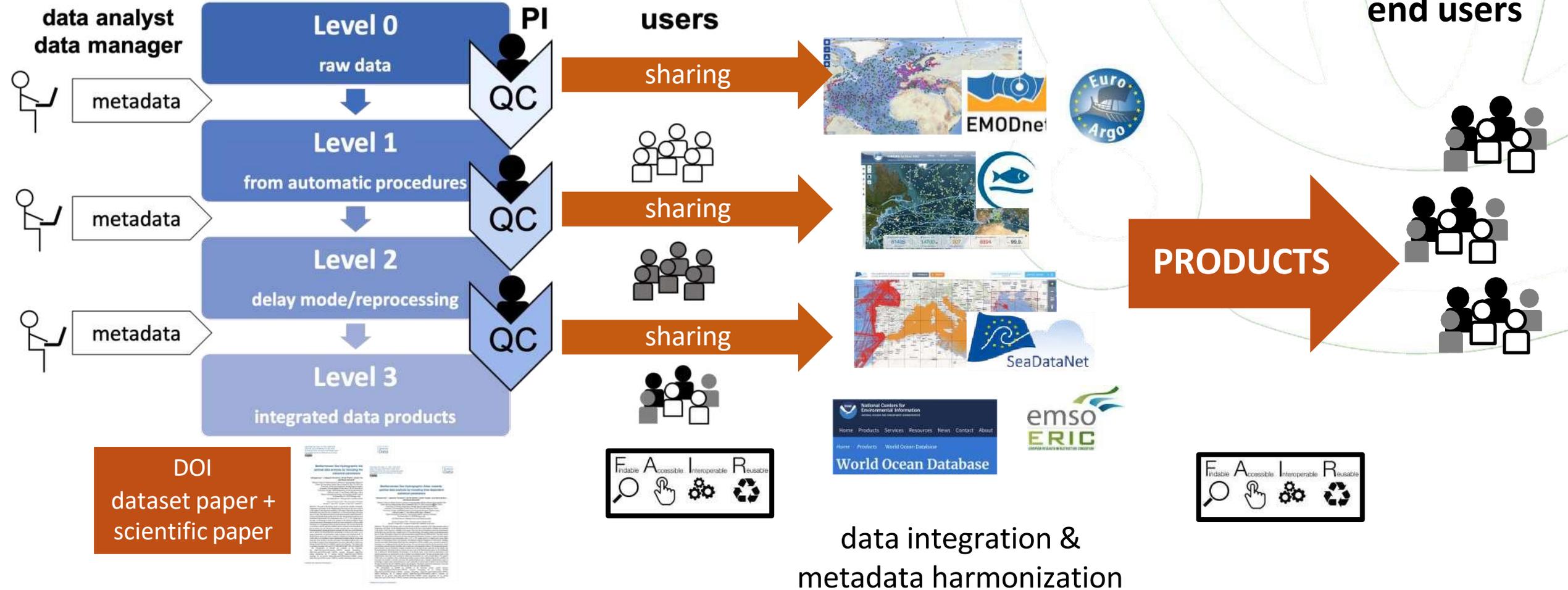
This could generate issues related to different data versions and duplicates within data infrastructures

Dataflow

Data Provider

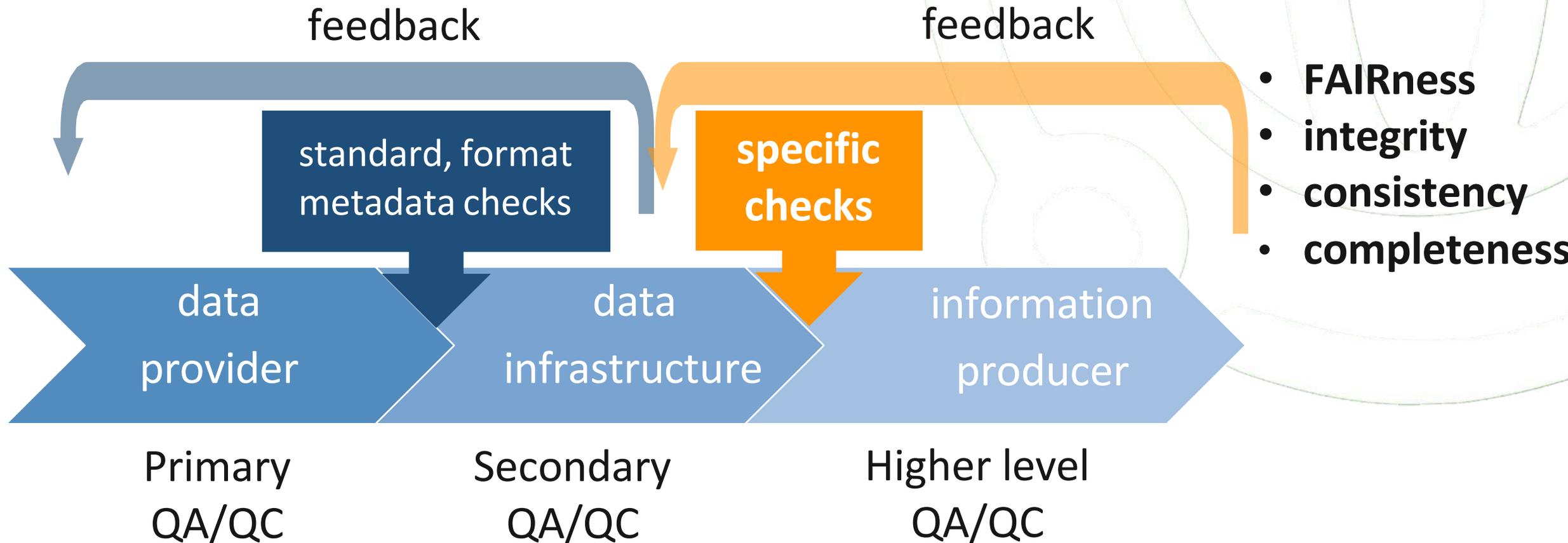
Blue Data Infrastructures

end users



Quality Control

Simoncelli et al. (2022) <https://doi.org/10.1016/B978-0-12-823427-3.00001-3>

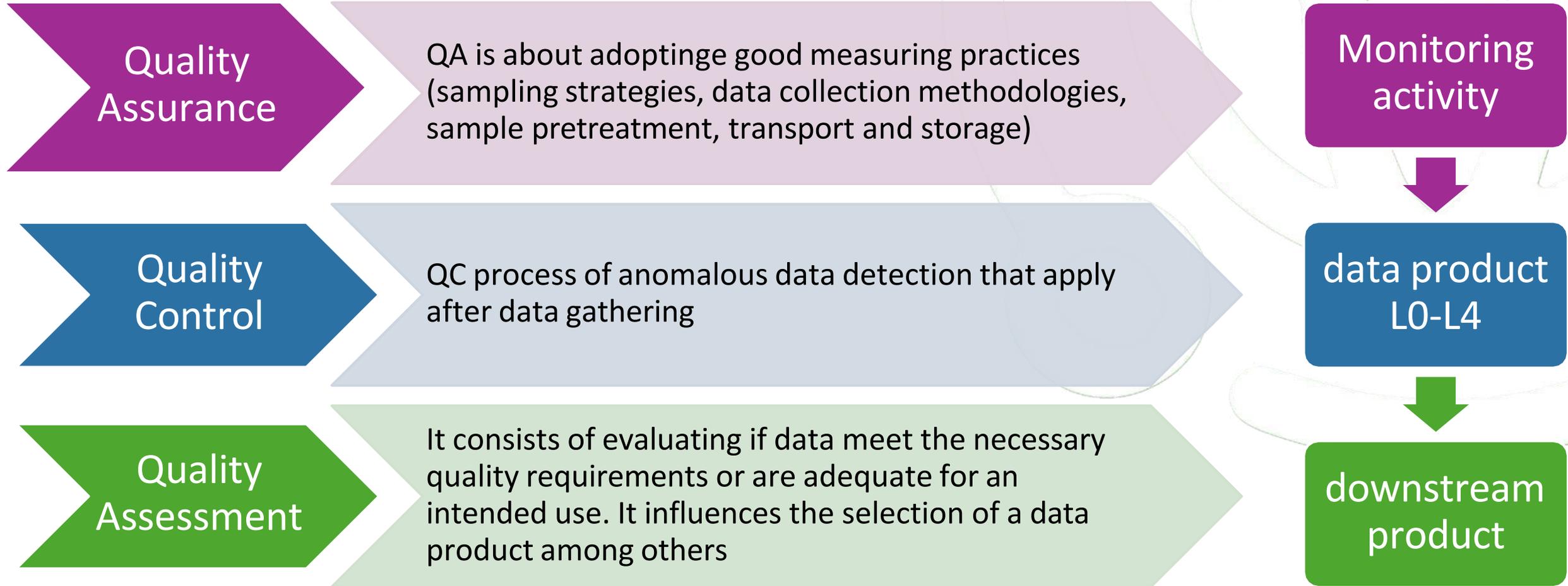


- **FAIRness**
- **integrity**
- **consistency**
- **completeness**

→ several actors apply QA/QC procedures along with the data value chain

→ **provenance** and **lineage information** are key elements to preserve

about Quality



Data accuracy and uncertainty are key quality elements of data reliability

Data uncertainty determination and its propagation along with the data value chain is still a present challenge

It is very important to annotate data uncertainty in the metadata

Quality Control

- 🌐 QC is vital for data reuse, and without it data from different sources cannot be combined to gain value
- 🌐 Scientific, analytical and statistical evaluations must determine if data present adequate quality to support the intended data usage, resulting in labeling each numerical value with a Quality Flag (QF) and avoid modifying the original data record following a harmonized scheme of QFs
- 🌐 QFs ensure that the quality of the data is apparent to the user, who holds sufficient information to decide the suitability for a specific task applying the proper data filtering

QC practices include:

- data integrity checks (e.g., format)
- data value checks (range checks, spikes and outliers checks, neighbor checks, climatology checks)
- QC highly depend on the data thematic, sensor type and the available amount of time for the analysis (RT vs DM)

Automatic QC (RT procedures) → algorithms

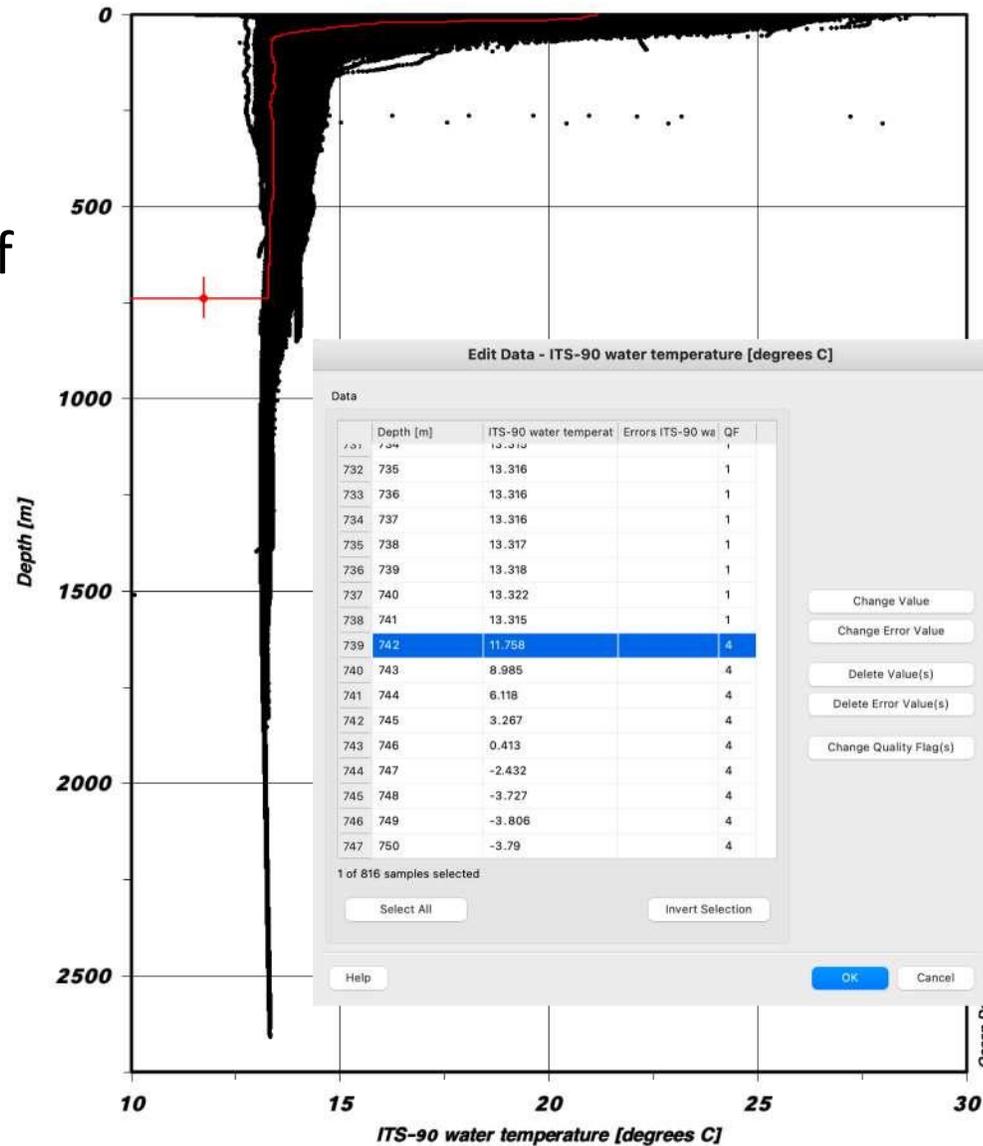
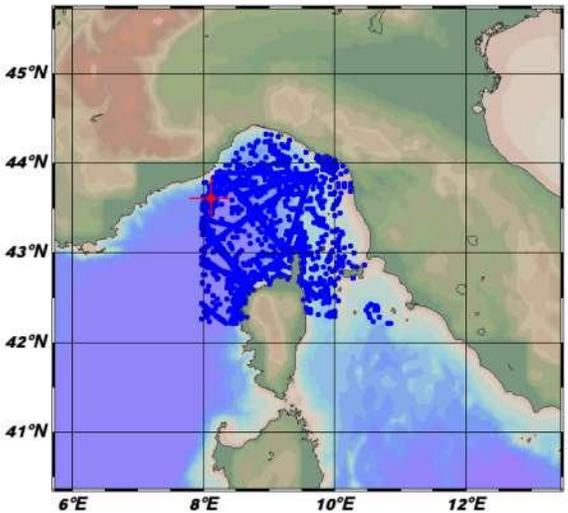
Visual QC (DM procedures) → visual tools (i.e. OceanDataView) and graphical interfaces

Quality Flags

- 🌐 Labels associated to each measurement that the user can use to filter data according to the needs
- 🌐 SeaDataNet harmonised scheme of QC Flags to be used to label individual data values
- 🌐 QC Flag scale is available in the SeaDataNet Common Vocabularies as list L20

Key	Entry Term	Abbreviated term	Term definition
0	no quality control	none	No quality control procedures have been applied to the data value. This is the initial status for all data values entering the working archive.
1	good value	good	Good quality data value that is not part of any identified malfunction and has been verified as consistent with real phenomena during the quality control process.
2	probably good value	probably_good	Data value that is probably consistent with real phenomena but this is unconfirmed or data value forming part of a malfunction that is considered too small to affect the overall quality of the data object of which it is a part.
3	probably bad value	probably_bad	Data value recognised as unusual during quality control that forms part of a feature that is probably inconsistent with real phenomena.
4	bad value	bad	An obviously erroneous data value.
5	changed value	changed	Data value adjusted during quality control. Best practice strongly recommends that the value before the change be preserved in the data or its accompanying metadata.
6	value below detection	BD	The level of the measured phenomenon was too small to be quantified by the technique employed to measure it. The accompanying value is the detection limit for the technique or zero if that value is unknown.
7	value in excess	excess	The level of the measured phenomenon was too large to be quantified by the technique employed to measure it. The accompanying value is the measurement limit for the technique.
8	interpolated value	interpolated	This value has been derived by interpolation from other values in the data object.
9	missing value	missing	The data value is missing. Any accompanying value will be a magic number representing absent data.
A	value phenomenon uncertain	ID_uncertain	There is uncertainty in the description of the measured phenomenon associated with the value such as chemical species or biological entity.

Ocean Data View (ODV) software package for the interactive exploration, analysis and visualization of oceanographic and other geo-referenced profile, time-series, trajectory or sequence data



Edit Data - ITS-90 water temperature [degrees C]

Depth [m]	ITS-90 water temperat	Errors ITS-90 wa	QF
732	735	13.316	1
733	736	13.316	1
734	737	13.316	1
735	738	13.317	1
736	739	13.318	1
737	740	13.322	1
738	741	13.315	1
739	742	11.758	4
740	743	8.985	4
741	744	6.118	4
742	745	3.267	4
743	746	0.413	4
744	747	-2.432	4
745	748	-3.727	4
746	749	-3.806	4
747	750	-3.79	4

1 of 816 samples selected

Select All Invert Selection

Change Value Change Error Value Delete Value(s) Delete Error Value(s) Change Quality Flag(s)

OK Cancel

Station ID: 644562

Accession Num...	636844
Cruise	MOON2013
Station	2103 (C)
Longitude	8.133°E
Latitude	43.6°N
Date	21 September 2013
Time	18:05:00
Depth Range [m]	[4 - 819]
LOCAL_CDI_ID	XO20130921003_136_H13
EDMO_code	136
Bot.Depth [m]	0
Instrument Info	
P01 Codes in O...	SDN:P01::ADEPZZ01 SDN:P01::TEMPET01
P35 Contributo...	SDN:P35::WATERTEMP = [SDN:P01::TEMPET01]
References	

Sample: 739 / 816

1: Depth [m]	742	1
2: ITS-90 water temperature [degrees ...	11.76	4
3: Water body salinity [per mille]		9
drvd: Potential Temperature θ [degC]		9
drvd: Potential Density Anomaly σ_0 [kg/ ...		9
drvd: Stability Ratio R_p		9
drvd: Dynamic Height-700 [dyn m]		9

Isosurface Values

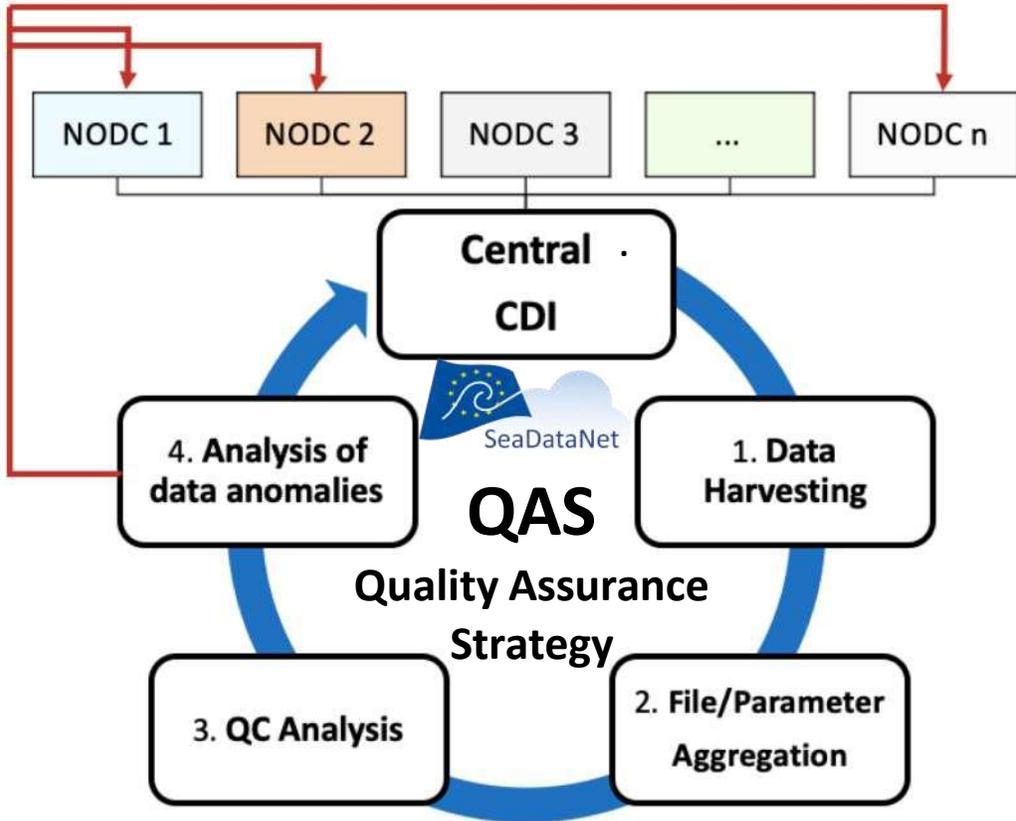
Longitude	8.133
Latitude	43.600
Time [yr]	2013.723
Day of Year	264
ITS-90 water temperature [degrees C] @ Depth [m]=150.00	13.48
ITS-90 water temperature [degrees C] @ Depth [m]=300.00	13.43
ITS-90 water temperature [degrees C] @ Depth [m]=600.00	13.35
ITS-90 water temperature [degrees C] @ Depth [m]=1000.00	
ITS-90 water temperature [degrees C] @ Depth [m]=2000.00	

<https://odv.awi.de/>

Data Products (example)



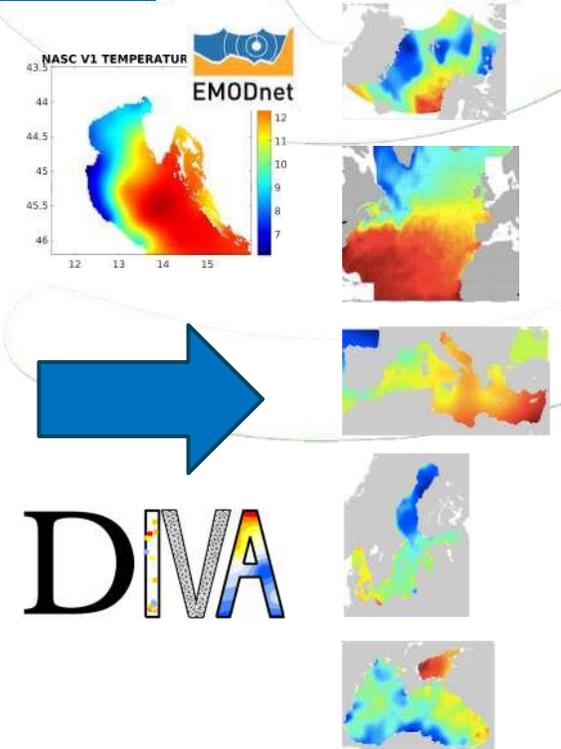
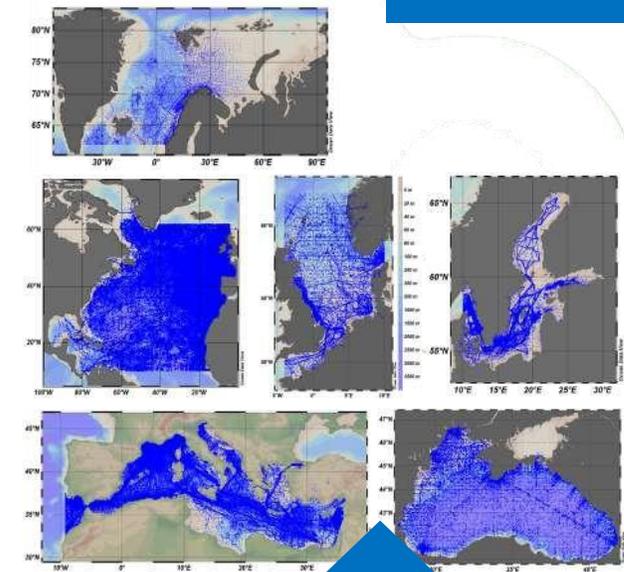
network of SeaDataNet data centers



**T&S
aggregated dataset**

SDN catalog
DOI
Product
Information
Document (PIDoc)

**T&S
gridded
climatologies**



**merging data
from other BDIs**



**+ OHC, MLD, density,
stratification
products**



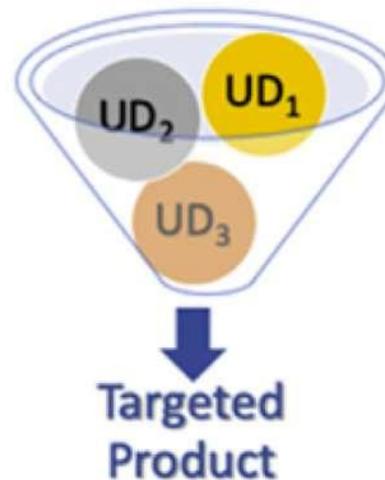
Quality Assessment

Input Data



(expected)

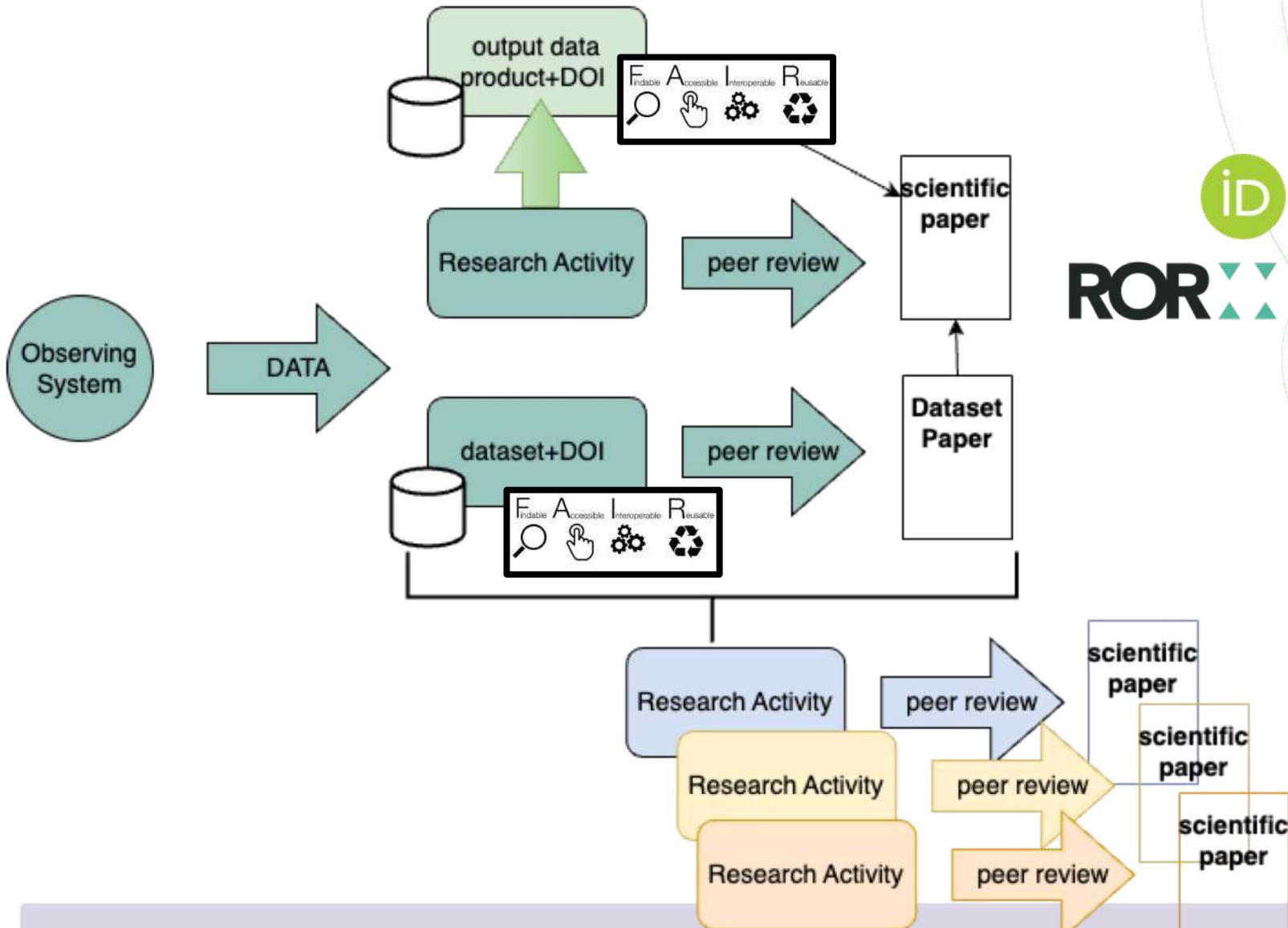
available Upstream Data



(realized)

- User requirements define the **Data Product Specification** and the necessary **Input Data** to generate it
- the realized **Targeted Product Description** compared to the preliminary specification provide a measure of fit-for-use
- **Upstream Data** are fit-for-purpose if adequate for the targeted product realization
- Quality assessment process might provide a feedback on the **adequacy of observing system** for specific products/applications

Data Publishing Strategy



- a correct data management maximizes research impact for the scientist (ORCID) and the Research Institute (ROR)
- it promotes and enhances knowledge generation and innovation
- it guarantees transparency, traceability and reproducibility of scientific outcomes

Digital Object Identifier (DOI)

🌐 DOI metadata schema

<https://schema.datacite.org/meta/kernel-4.5/metadata.xsd>

<https://datacite-metadata-schema.readthedocs.io/en/4.5/>

🌐 Citation made by: creators, title, issue data → no changes are advisable

🌐 landing page should contain How To Cite instruction and Data license

🌐 landing page best practices at

<https://support.datacite.org/docs/landing-pages>

The screenshot shows the PANGAEA website interface. At the top, there is a navigation bar with the PANGAEA logo, the text "Data Publisher for Earth & Environmental Science", and links for "SEARCH", "SUBMIT", "ABOUT", and "CONTACT". A "Not logged in" status is visible in the top right corner. The main content area displays a dataset entry with the following text: "Hüning, Anne K; Melzner, Frank; Thomsen, Jörn; Gutowska, Magdalena A; Krämer, Lars; Kohno, Mika; Rosenstiel, Philip; Pörtner, Hans-Otto; Philipp, Eva E R; Lucassen, Magnus (2013): Seawater carbonate chemistry and mantle gene expression patterns of the Baltic Sea blue mussel in a laboratory experiment. PANGAEA, doi: <https://doi.org/10.1594/PANGAEA.833670>". Below this text, there are instructions: "Always quote above citation when using data! You can download the citation in several formats below." and a dropdown menu with options like "Download", "View", and "Print". At the bottom of the dataset entry, there is a "Download Data" section with two options: "Download dataset as tab-delimited text (use the following character encoding: UTF-8: Unicode (PANGAEA default))" and "View dataset as HTML (shows only first 2000 rows)".

Annotations on the left side of the screenshot:

- Full bibliographic citation**: Points to the full citation text.
- DOI displayed as URL**: Points to the DOI link in the citation.
- Access to the item**: Points to the "Download Data" section.

Citing source data is a compulsory requirement by most data licenses (the “BY” component in a Creative Commons license)

A bibliographic citation of any digital object associated to a DOI can be obtained entering the code at

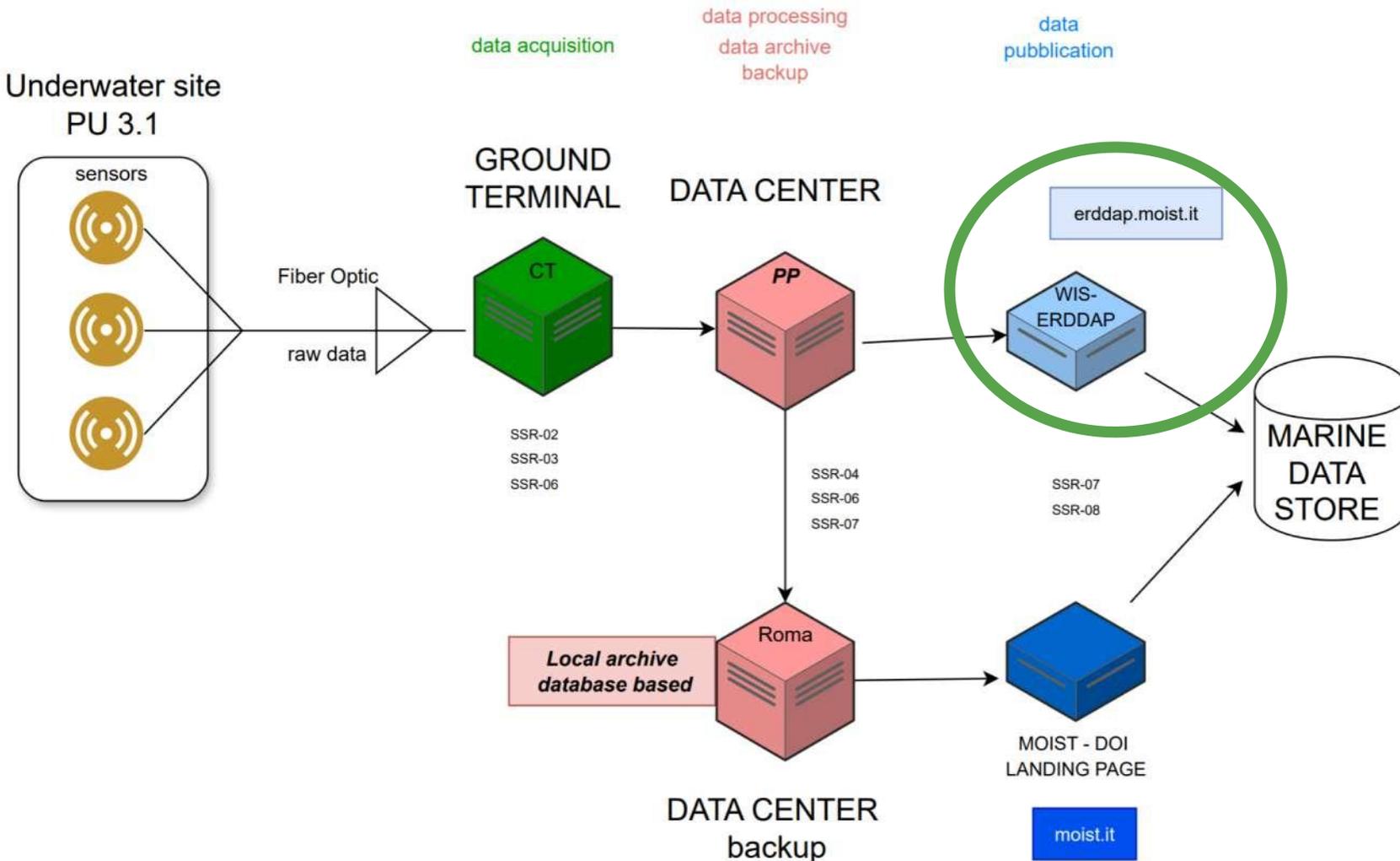
<https://citation.crosscite.org/>

from Locati & Cacciola (2024) Research Data Management
<https://istituto.ingv.it/ufficio-gestione-dati>



WIS dataflow and ERDDAP data access service

ERDDAP → FAIR-compliant data access service in line with the GOOS Observations Coordination Group strategy



- supports dozens of popular formats
- provides standards-based metadata and formats
- supports federated access of distributed ERDDAP data services;
- supports both human and machine interactions
- supports sub-setting of large datasets
- provides improved discovery of datasets through commercial search engines
- provides support for archival of datasets

Workflows

Workflow → a series of activities/processes that are necessary to complete a task

→ its **automation and management process** increase efficiency, optimize the results, allows its replicability and monitoring

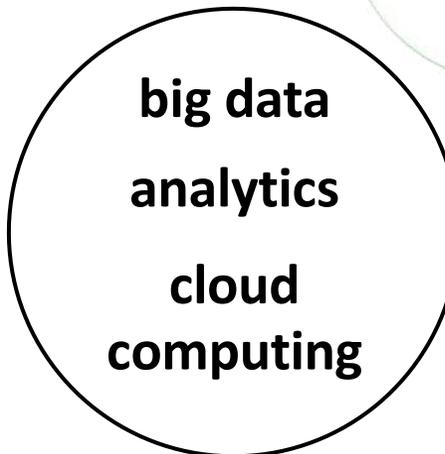
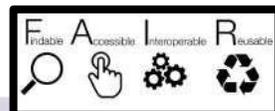
→ its implementation builds on advanced services and FAIR principles

Data Workflow



Digital Ecosystem

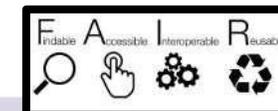
**Data
Discovery
&
Access
Services**



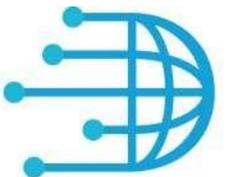
**Virtual Research
Environment**

data products

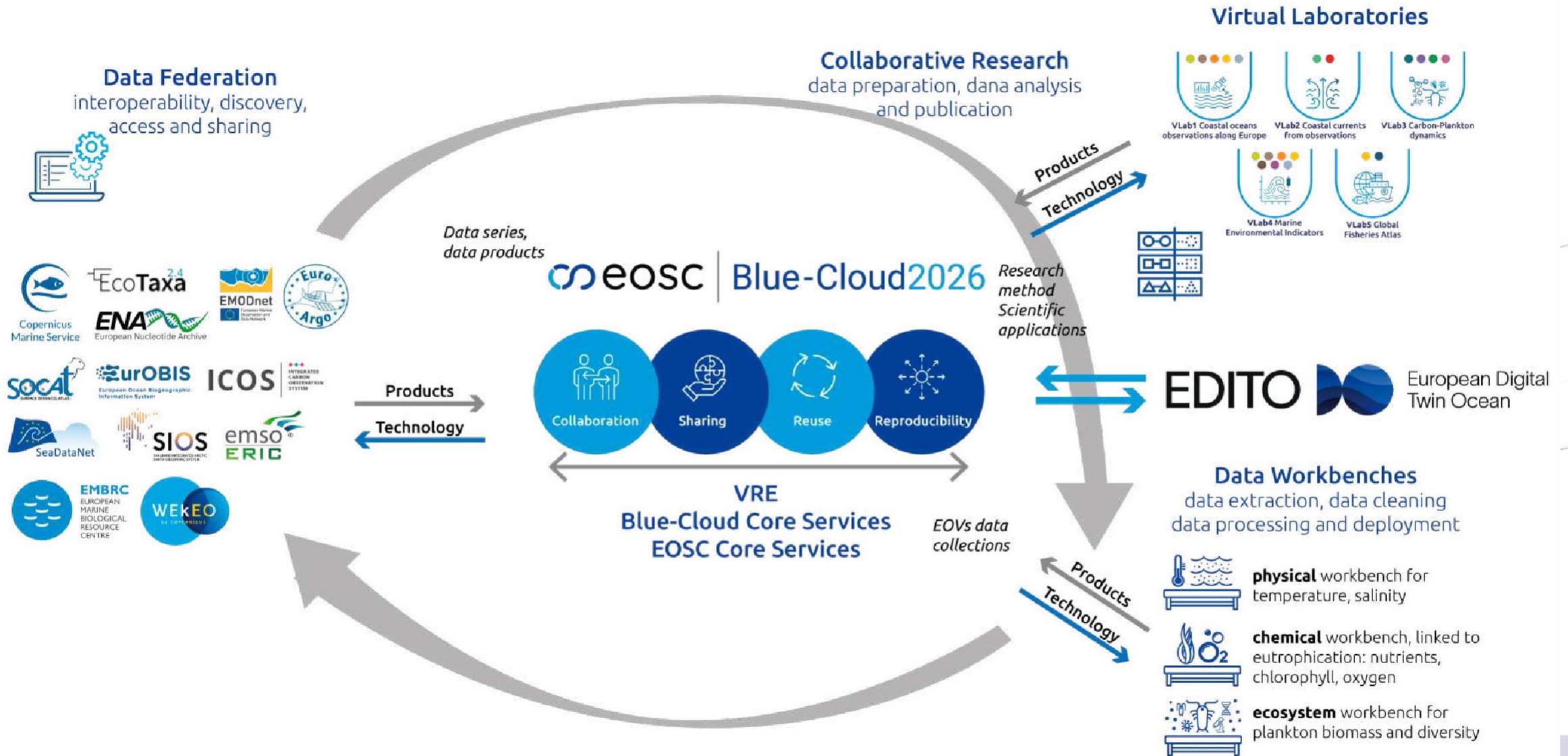
applications



**Digital
Twins**



Blue Cloud in a nutshell



FAIRness Assessment

Findable
Metadata and data should be findable for both humans and computers

Interoperable
Data needs to work with applications or workflows for analysis, storage and processing

F A I R

Accessible
Once found, users need to know how the data can be accessed

Reusable
The goal of FAIR is to optimise data reuse via comprehensive well-described metadata

FAIR proposed by the community <https://force11.org/info/the-fair-data-principles/>

Findable	
F1	(meta)data are assigned a globally unique and eternally persistent identifier.
F2	data are described with rich metadata.
F3	(meta)data are registered or indexed in a searchable resource.
F4	metadata specify the data identifier.
Accessible	
A1	(meta)data are retrievable by their identifier using a standardized communications protocol.
A1.1	the protocol is open, free, and universally implementable.
A1.2	the protocol allows for an authentication and authorization procedure, where necessary.
A2	metadata are accessible, even when the data are no longer available.
Interoperable	
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.
Re-usable	
R1	(meta)data have a plurality of accurate and relevant attributes.
R1.1	(meta)data are released with a clear and accessible data usage license.
R1.2	(meta)data are associated with their provenance.
R1.3	(meta)data meet domain-relevant community standards.

FAIRness Assessment



Data FAIRness is useful for data providers to learn and adopt FAIR principles during data generation or to implement data FAIRification:

1. Data Maturity Model (FDMM) <https://doi.org/10.15497/rda00050>
2. F-UJI tool (automatic evaluation through DOI or URL) developed within FAIRsFair project <https://www.f-uji.net/>



FAIR principles and assessment is expanding to all «digital research objects» (data, software, semantic artifacts, VRE, ...)

Software FAIRness

RDA FAIR4RS working group's principles → indicators under development

<https://doi.org/10.15497/RDA00068>

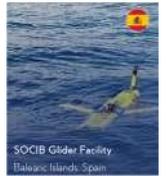


- It provides access to a comprehensive and diverse suite of integrated RIs to address challenges and explore opportunities for the long-term sustainability of our marine and freshwater ecosystems
- through **transnational Access (TA) funding calls** targets and supports research and innovation activities that contribute to the objectives, regional scope and implementation of the *EU Mission 'Restore our Ocean and Waters by 2030'*
- 4 lighthouse regions:** Baltic and the North Sea Basins, Black Sea, Atlantic/Arctic, Mediterranean Sea along with their associated major rivers
- brings together an **online catalogue of RIs** and support research and innovation projects via TA calls
- Open data policy**, implemented via a dedicated **Data Management Plan**, to ensure that all gathered and generated metadata and data are managed in line with the FAIR principles

<https://aquarius-ri.eu/>



data gathering



DMP template
Data Summary Log



expert data centers

feeding



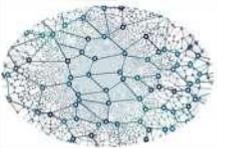
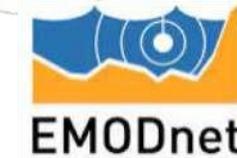
<https://www.emodnet-ingestion.eu/>

SEANOE <https://www.seanoe.org/>



eosc
Blue-Cloud2026

feeding



Ocean Decade
Ecosystem

All metadata/data become part of the archives managed and operated by leading EU data infrastructures for QA/QC, long term stewardship, wide access and use

57 Ris

- research vessels
- marine mobile obs platforms
- fixed marine facilities
- aircrafts
- drones
- ...

References

Simoncelli, S., Manzella, G.M.R., Storto, A., Pisano, A., Lipizer, M., Barth, A., Myroshnychenko, V., Boyer, T., Troupin, C., Coatanoan, C., Pititto, A., Schlitzer, R., Schaap, D.M.A., Diggs, S., 2022. A collaborative framework among data producers, managers, and users. <https://doi.org/10.1016/B978-0-12-823427-3.00001-3> In: Manzella, G., Novellino, A. (Eds.), **Ocean Science Data: Collection, Management, Networking and Services**. Elsevier, pp. 197–280. ISBN: 9780128234273

UNESCO-IOC (2023). Ocean Decade Data & Information Strategy. Paris, UNESCO. (The Ocean Decade Series, 45)

Calewaert et al. (2024). Ocean Decade Vision 2030 White Papers – Challenge 8: Create a Digital Representation of the Ocean. Paris, UNESCO-IOC. (The Ocean Decade Series, 51.8.). <https://doi.org/10.25607/bxhy-ra59>

Miloslavich et al. (2024). Ocean Decade Vision 2030 White Papers – Challenge 7: Sustainably Expand the Global Ocean Observing System. Paris, UNESCO-IOC. (The Ocean Decade Series, 51.7.). <https://doi.org/10.25607/brxb-kr45>

Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. 2016. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data* 3:60018 <https://doi.org/10.1038/sdata.2016.18>

The Creative Commons family of licences <https://creativecommons.org/about/ccllicenses/>

Schaap, D., Le, H. M., Iona, S., Diez, S., Partescano, E., Sagen, H., Buga, L., Schepers, L., Hedblom, M., O'Grady, E., Van Iseghem, S., Krijger, T., & Giorgetti, A. (2024). AQUARIUS Data Management Plan (Version 1). Zenodo. <https://doi.org/10.5281/zenodo.13642869>

Locati & Cacciola (2024) Research Data Management <https://istituto.ingv.it/ufficio-gestione-dati>



THANKS!

IR0000032 – ITINERIS, Italian Integrated Environmental Research Infrastructures System
(D.D. n. 130/2022 - CUP B53C22002150006) Funded by EU - Next Generation EU PNRR-
Mission 4 "Education and Research" - Component 2: "From research to business" - Investment
3.1: "Fund for the realisation of an integrated system of research and innovation infrastructures"

