Blue-Cloud

Piloting innovative services for Marine Research & the Blue Economy

at #EOSCSymposium22

Blue-Cloud Open Science platform
as a model for a sustainable
Data federation in EOSC
Workshop

17 November 2022
11:00-12:30 CET
Prague, Czechia

eosc

EOSC Future

**Funded by
the European Union**

Why this session?

- **The Blue-Cloud Open platform is a great example of implementation of the three main components of the EOSC Federation - EOSC-Core, EOSC-Exchange and the Federation of Data & Data Services - into a unique Open Science platform** accessible and usable by other research communities.

Source: **Towards Sustainable Funding Models for the European Open Science Cloud**
Financial Sustainability Task Force Progress report, November 2022

- Of the three components, the **Data Federation is the raw material without which the EOSC cannot achieve the ambitious goals** set by its founders.

Blue-Cloud is a front runner of Data federation in practice, given its success in bringing together 9 data providers covering 10M datasets from the marine domain and making them available via the DD&AS and the VRE to users, allowing interdisciplinary interactions between disciplines, thus demonstrating the value of bringing together a variety of providers and users within EOSC.

Source: **Services Exploitation & Sustainability Plan**
Draft - November 2022

Today

We are here to analyse what scenarios for data federation are being tackled by different initiatives in Europe with respect to the main findings of the EOSC Task Force on Financial Sustainability and their inputs for funding the core, exchange & data federation components

Agenda

Time	What	Who
11.00 – 11.10	**Reflections on the approach and findings on the Task Force on Sustainability**	**Jessica Klemeier**, EMBL and member of the EOSC Sustainability Task Force representative
11.10 – 11.20	**Blue-Cloud smart federation model as a best practice for the EOSC**	**Sara Pittonet**, Trust-IT & Blue-Cloud coordinator
11.20 – 11.50	**Overview of Blue-Cloud best in class data management use case**	**Anton Ellenbroek**, FAO of the UN, Knowledge and Information Management Team of the Fisheries and Aquaculture Division **Karl Presser**, Managing Partner at Premotec GmbH and FNS-Cloud H2020 representative
11.50 – 12.00	**Future evolution of Interoperability & FAIR data management aspects in Blue-Cloud 2026**	**Alessandro Rizzo**, IRD and FAIR-EASE
12.00 – 12.30	**A sustainability model for data federation in EOSC**	**Panel** discussion moderated by **Jessica Klemeier**

Blue-Cloud

Jessica Klemeier

EMBL and alternate member of the EOSC Financial Sustainability Task Force

Reflections on the approach and preliminary findings on the Task Force on Financial Sustainability

Data Federation in EOSC

*"It is possible to envisage that the **EOSC Data Federation (DF)** should enable researchers to find and acquire data from multiple sources available at any of the levels of aggregation (local/institutional, national, thematic, European or international) through attribute-based discovery. A federation with these characteristics would lift the barriers to reusing data"*

Source: **Towards Sustainable Funding Models for the European Open Science Cloud**
Financial Sustainability Task Force Progress report, November 2022

TF Financial Sustainability

Financial Sustainability Task Force

Chairs

Jan Meijer
SIKT

Dale Robertson
EGI

Board Liaison

Bob Jones
CERN

Advisory Group Sustaining EOSC

Task Force Long-Term Data Preservation

Task Force Financial Sustainability

Members

Belso, Rene DeiC	**De Giacomo, Ornela** CERIC-ERIC	**Hacque-Cosson, Françoise** CEA	**Haley, Natalie** Instruct-ERIC
Hrušák, Jan Czech Academy of Sciences	**Ilijasic Versic, Ivana** ELIXIR	**Kaczmirek, Lars** University of Vienna	**Kraker, Peter** Open Knowledge Maps
Kurowski, Krzysztof Poznan Supercomputing and Networking Center	**Kvatadze, Ramaz** GRENA	**Lecarpentier, Damien** EUDAT	**Lueck, Rupert** EMBL
Matyska, Luděk CESNET	**Mergen, Patricia** Meise Botanic Garden	**Muscella, Silvana** Trust-IT Services	**Nordh, Valter** SUNET
Perez, Alberto Red.es	**Proudman, Vanessa** SPARC Europe	**Rey Mazón, Miguel** Graz University of Technology	**Roi, Arnaud** DARIAH ERIC
Szprot, Jakub University of Warsaw	**Klemeier, Jessica** EMBL	**Vilotte, Jean-Pierre** CNRS	

EOSC main components

- EOSC-Exchange
- EOSC-Interoperability Framework
- Federated Data
- EOSC-Core
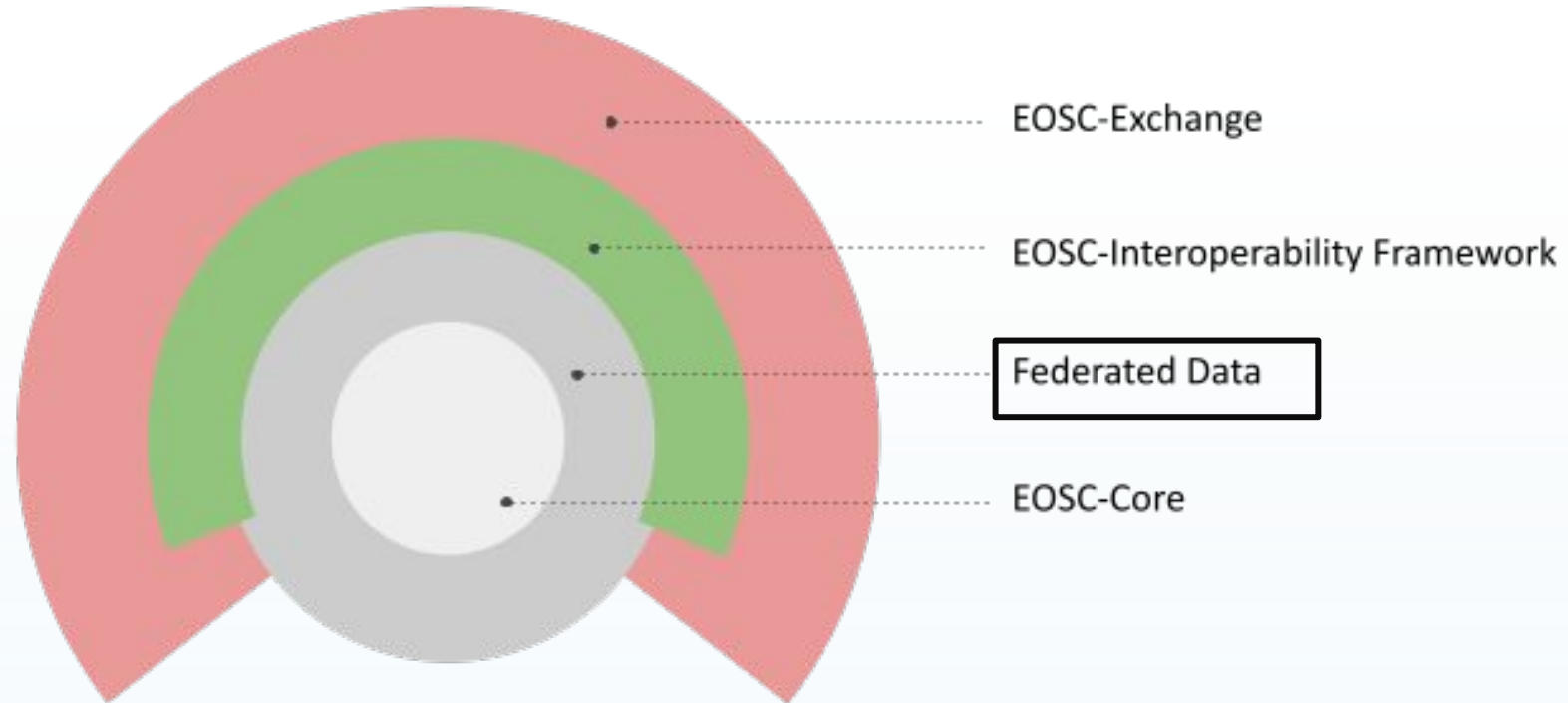
EOSC Data Federation
Lessons from use cases

Blue-Cloud

⇒ **Data and services must go together** in EOSC but have different financial requirements

⇒ Federating data implies **interoperability** between five levels of aggregation (local/regional/nat'l/EU/int'l)

⇒ EOSC DF must **rely on existing repositories and (FAIR) federations** & on all levels and make them discoverable

⇒ EOSC must **use existing infrastructures** and thematic ecosystems to avoid duplicating efforts

⇒ Need connect with **European Common Data Spaces, Gaia-X, Global Open Science Cloud**

⇒ **Establishment of EOSC globally** will help with funding, competitiveness int'l visibility

What could EOSC Data Federation be(come)?

Blue-Cloud

Solution	Main characteristics
Overlay Intermediary software layer in addition to any federation layers existing in a scientific domain or geographical area	Centrally managed and maintained by EOSC Enables data discovery from any provider Ensures data interoperability to the highest degree possible
Metadata catalogue System to track metadata organised into a catalogue	**Can** achieve interoperability
Catalogue of data providers and platforms Compilation of providers, platforms, and how to access them	**Cannot** achieve interoperability
Natural evolution Default solution in the absence of EOSC	Existing data infrastructures remain as go-to sources Convergence expected to still occur but uncoordinated and slow; involves significant costs

Identified additional costs

Making data FAIR	Making experiments reproducible	Ensuring long-term access to data	Federating data to EOSC
Operational Costs ● Data Stewardship ● Control Systems to acquire metadata ● Storage ● Data transfer ● Access to computing ● Making data findable Development costs ● Capture metadata ● Improve data reduction Legacy → FAIR-data	Workflows and software Data analysis environments Software catalogues Computing capacity to run data analysis Data transfer protocols	Storage & archival costs for sustainable repositories Update of metadata and APIs Connect endpoints and make data findable via data catalogues across Europe Tools & services to make data FAIR	Creating links & enabling interoperability between data resources and EOSC Operational costs of interoperability Data deduplication Data harmonisation Alignment of metadata schemas Certification & validation of repositories Legal and ethical (sensitive data)

Interim conclusions, open questions and next steps

Blue-Cloud

- Several options available for EOSC DF (from "Overlay" to keep "natural evolution")

- Costs & responsibilities must be assigned to the right level (EOSC, national, etc.) depending on the DF model chosen

- Further exploration of case studies (e.g. to obtain figures), or new ones, may be required

- Compare/combine TF analysis with EC-commissioned study on characterisation of European Research Data Landscape & outputs from TFs Long-term Data Preservation & FAIR Metrics and Data Quality

- Seek alignment with the Data Spaces Support Centre about Smart Middleware Platform

- As of now, EOSC won't replace thematic portals, e.g. Blue Cloud - need to find ways to complement each other

Blue-Cloud

Sara Pittonet, Trust-IT & Blue-Cloud coordinator

Blue-Cloud smart federation model as a best practice for the EOSC

Blue-Cloud Key products and services

Developed to showcase the potential of the European Open Science Cloud

Blue-Cloud Virtual Research Environment to provide a range of services and to facilitate orchestration of computing and analytical services for constructing, hosting and operating Virtual Labs for specific applications

EOSC CORE

Blue-Cloud Virtual Labs, configured with analytical workflows to serve 5 demonstrators, which can be adapted to support other thematic communities.

EOSC EXCHANGE

Blue-Cloud Data Discovery & Access service, federating key European data management infrastructures, to facilitate users in finding and retrieving multi-disciplinary datasets from multiple repositories

EOSC Data Federation

[Services available in the EOSC Portal](#)

Blue-Cloud overarching concept

- Developing and deploying Virtual Research Environment (VRE) with an array of services for configuring and running Virtual Labs for specific analytical workflows, use cases, and demonstrators

- Applying common standards and interoperability solutions for providing harmonized metadata and data

- Developing and deploying harmonized discovery and access to established European marine data management and processing infrastructures

Added-value services and Applications
VRE – Cloud Platform
Downstream services

Standards
OGC, ISO, W3C
Vocabularies

Upstream services

Discovery and Access to data sets from many resources
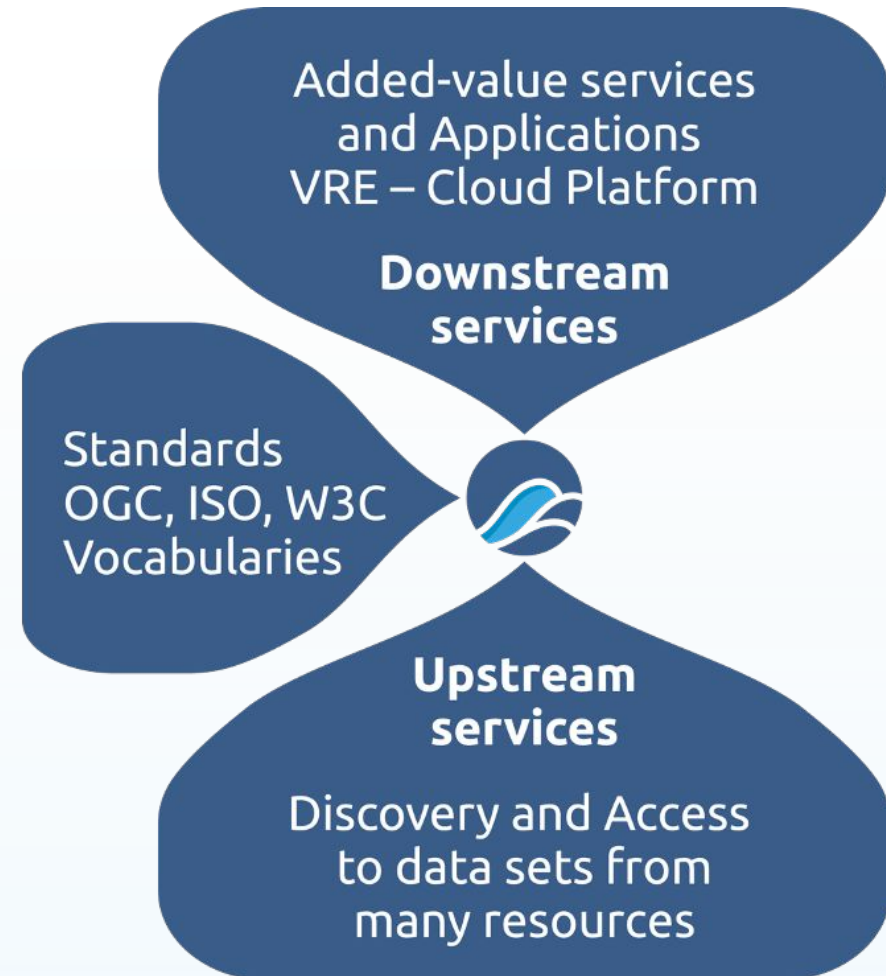
Expectations & Opportunities from EOSC

- **Blue-Cloud understands the win-win where EOSC offers the frame or "container" to implement the data technical harmonisation of the Blue-Cloud VRE**. Allowing interactions between different disciplines in EOSC will feed interdisciplinarity interactions as data become accessible, demonstrating the value of bringing together the variety of users within EOSC.

- **The Blue-Cloud DD&AS aligns its position with the European infrastructures** and ensures that the Blue Data Infrastructures and major initiatives such as EMODnet, COPERNICUS, and SeaDataNet receive added-value in EOSC.

Source: **Blue-Cloud Exploitation and Sustainability Plan**
Draft - November 2022

Blue-Cloud

Blue-Cloud

Name

Blue-Cloud – Piloting innovative services for Marine Research & the Blue Economy

Grant agreement ID

862409

Website

blue-cloud.org

CORDIS

cordis.europa.eu/project/id/862409

Brief description of the project

Blue-Cloud is the thematic marine EOSC, a collaborative Open Science platform in support of the EU Mission Ocean. Blue-Cloud innovative core services are deployed through a smart federation of leading European marine data and infrastructures, bringing an unprecedented amount of multidisciplinary data repositories, analytical tools, and computing facilities to the EOSC.

Relevance of the KERs to the EOSC Advisory Groups (AGs) and Task Forces (TFs)

AG	Task Force	KER #1	KER #2	KER #3	KER #4	KER #5	KER #6
AG 1	1.1 PID Policy and Implementation	✓	✓				
	1.2 Researcher Engagement and Adoption	✓	✓				✓
	1.3 Rules of Participation Compliance Monitoring	✓					
AG 2	2.1 FAIR Metrics and Data Quality	✓	✓	✓	✓	✓	✓
	2.2 Semantic Interoperability		✓	✓	✓	✓	
AG 3	3.1 Data Stewardship Curricula and Career Paths						
	3.2 Research Careers, Recognition, and Credit	✓	✓				✓
	3.3 Upskilling Countries to Engage in EOSC	✓	✓	✓			
AG 4	4.1 AAI Architecture	✓		✓			
	4.2 Infrastructure for Quality Research Software	✓				✓	✓
	4.3 Technical Interoperability of Data and Services	✓		✓	✓	✓	
AG 5	5.1 Financial Sustainability	✓					
	5.2 Long-term Data Preservation						

AG1: Implementation of EOSC AG2: Metadata and data quality AG3: Research careers and curricula
AG4: Technical challenges on EOSC AG5: Sustaining EOSC

KER #1 Blue-Cloud Virtual Research Environment (VRE)

The Blue-Cloud Virtual Research Environment is a cloud-based analytical and publishing framework working as a federation and orchestration of computing platforms and analytical services for constructing, hosting and operating Virtual Labs for specific applications in the marine and ocean domain.

Exploitability

Operational service

Link: blue-cloud.d4science.org

KER #2 Blue-Cloud catalogue

The Blue-Cloud Catalogue features datasets and products resulting from the Blue-Cloud VLabs and provenance metadata on the methods, data sets and workflows used to generate them.

Exploitability

Operational service

Link: blue-cloud.d4science.org/catalogue-bluecloud

Blue-Cloud Data Discovery & Access Service

The Blue-Cloud Data Discovery & Access Service facilitates federated discovery and retrieval of data sets and data products from multiple Blue Data Infrastructures in a common discovery and access interface, both for external users in stand-alone mode, and for users of the Blue-Cloud VRE.

Exploitability

Operational service

Link: data.blue-cloud.org

Blue-Cloud Virtual Lab for Fisheries data

The Virtual Lab on Fisheries data allows users to explore all oceans and regions of the world with the Fisheries Atlas. Features range from global fisheries maps, statistics and overviews, aggregate records about major world fisheries, and the Global Record on Stocks and Fisheries.

Exploitability

Operational service

Link: blue-cloud.org/vlabs/global-record-stocks-and-fisheries

Blue-Cloud Virtual Lab for Aquaculture Monitor

A tool to produce national aquaculture sector overviews via OGC compliant data services to monitor country aquaculture sector. The tool is built on interoperable services where teams can compute and publish reproducible experiments.

Exploitability

Operational service

Link: blue-cloud.org/vlabs/aquaculture-monitor

KER #6 Blue-Cloud Virtual Lab for Zoo and Phytoplankton EOV products

The Zoo and Phytoplankton EOV demonstrator provides a description of the current state of the plankton communities and forecasts their evolution. This provides valuable information for the modelling, assessment and management of the marine ecosystem.

Exploitability

Operational service

Link: blue-cloud.org/vlabs/zoo-and-phytoplankton-eov-products

KERs Categories

- Technical Harmonisation
- Policy Harmonisation
- Discovery/Access Platform
- Virtual Research Environment (VRE)
- Training Resource
- Knowledge Centre
- Authentication and Authorization Infrastructure (AAI)
- Validation Tool or Other

17 November 2022   EOSC Symposium 2022   17

Is funding for Blue-Cloud assets secured until 2030?

EU and other public funding

Funding coming directly from research institute and academia

Private contracts and partnerships, other commercial agreements

Source: **Services Exploitation & Sustainability Plan**
Draft - November 2022

Blue-Cloud VRE, including Data Analytics Framework

Vision by 2030	To consolidate as one of Europe's leading collaborative fit-for-purpose analytical & modelling environments accelerating data interoperability, modelling and Big Data analytics in the marine & freshwater domains.
Developments needed by 2030	Achieving this vision will require developing and adding to the D4Science e-infrastructure the possibility to read and process very large data sets at high performance (HPC services). This would have significant impact on the attractivity and wider use of the Blue-Cloud VRE for data intensive research, bringing yet unknown analytics and simulations within reach.
Replicability	Yes The Blue-Cloud VRE is configured to exploit services, computing and storage that are all virtualised and can be dynamically assembled to deliver an integrated platform to the community.
Is funding by 2030 secured?	At least till middle of 2028, while extension to 2030 is very likely, since CNR is continuing to manage and operate the underlying D4Science e-infrastructure
	Blue-Cloud VRE is powered by D4Science managed by CNR-ISTI. Its business model is based on diversification of the funds, ranging from EU and national projects, institution funds, and commercial contracts. This result can be replicated with an affordable effort and offered to other communities, even from different domains
What needs to be funded to continue the service?	New development for optimisation and expansion of the common (core) services enabling the operation of the infrastructure
Level of investment needed	€250K-€300K per year to sustain the enhancement and evolution of the common services enabling the infrastructure powering the VRE. While expanding into HPC capabilities will require much higher investments and funding levels.

Blue Cloud Virtual lab for Fisheries data

Vision by 2030	**Fisheries Atlas**: Coverage of more fisheries and better mash-up with data from other domains to support Digital Fisheries. **GRSF**: Fully traceable fisheries resources with identifiers informing on provenance, safety (for the fish and the consumer) sustainability and impact
Developments needed by 2030	Maintenance of the core SDI OS components (Geonetwork, geoserver etc) Improvements to bespoke SDI components such as GEOFLOW to better manage dataflows into and out of the system Improve support to time dimension and regional atlases Analytical capacity for geostatistics and time series Computing capacity for ecological modelling
Replicability	End products of the Fisheries Atlas are meant to be reproducible since we generate time series of fisheries data. The main goal of our work is to ensure our ability to easily run a yearly update of our time series by relying on a stable infrastructure. GRSF is a product *per se* - doesn't need replicability.
Funding by 2030 secured?	At least till middle of 2028, while extension to 2030 is very likely, since CNR is continuing to manage and operate the underlying D4Science e-infrastructure, and there are bilateral arrangements between CNR and FAO.
	The results depend on data availability which is managed by the FIRMS partnership to ensure a long-term collaboration between tuna RFMOs, FAO, and IRD. FAO re-uses many components, bringing down the costs of maintenance and on-boarding similar services. For what concerns the GRSF, operation and maintenance of the service is secured thanks to a great global collaboration with leaders in academia and industry.
What needs to be funded to continue the service?	● The underlying infrastructure to ensure reproducibility of the work within a stable a runtime environment ● ETP / engineers hired by projects to improve the quality of existing code and datasets and to create new products ● Engineers (permanent positions) to reproduce the work of previous projects (yearly updates of datasets) even if no projects is bringing funding.
Level of investment needed	2M €

Data Discovery and Access Service

Blue-Cloud

Vision by 2030	Become the reference service for harmonised discovery and access to a range of multi-disciplinary repositories in the marine community
Developments needed by 2030	The functionality of the service needs to be further developed and expanded, e.g. with data subsetting and data lake techniques, more Blue Data Infrastructures (BDIs) need to be federated, the machine-to-machine exchange with BDIs needs to be streamlined concerning functionalities and FAIRness, the interfacing towards the future DTO needs to be developed, tuning with EMODnet and Copernicus Marine.
Replicability	**Partially** Technology is replicable; organisation is more difficult as it requires a lot of networking
Funding by 2030 secured?	Not yet; Funding is secured till middle of 2026
	MARIS and the other owners of exploitation aim for continued **EU funding** for further developments and for qualifying as an essential service for a large community of users as well as of increasing number of connected BDIs
What needs to be funded to continue the service?	● New developments for optimisation and expansion of offer, functionality and services. ● Promotion and marketing activities to increase number of users. ● Operational costs of the underlying infrastructure
Level of investment needed	2-4M € (also needed for upgrading of web services of BDIs)

Blue-Cloud Data Discovery and Access Service

Blue-Cloud developed the Data Discovery and Access Service as a search engine that facilitates discovery and retrieval of data sets and data products from multiple Blue Data Infrastructures.

Circa **25.000 entries** are indexed in B2FIND at the level of data collections, while using additional search criteria specific per each data infrastructure, users can access more than 10 million data sets from a wide variety of disciplines (physics, biology, geology, bathymetry, chemistry).

data.blue-cloud.org

Architectural models for the EOSC Data Federation

1) **"Overlay":** to enable data discovery from any provider, EOSC could put in place an **intermediary (software) layer**, centrally managed and maintained by EOSC, that ensures data interoperability to the highest degree possible. This would come "on top of" (i.e. in addition to) any DF layers which already exist, e.g. in a specific scientific domain or geographical area.

2) **Metadata catalogue:** Instead of linking the data, **a system to track metadata** (i.e. not data itself, but their attributes or characteristics), **organised into a catalogue**, could be adopted by EOSC. This catalogue could then be used to achieve data interoperability, since having a coherent and consistent metadata classification scheme for data of very different nature allows to find and eventually use diverse types of data.

3) **Catalogue of data providers and platforms:** If a common interface to access the (meta)data cannot be implemented, an alternative solution would be to **compile a list of all data providers and the platforms they use**, indicating how to access single or combined data sources depending on the scientific discipline or geographic area.

4) **Natural evolution**: In the absence of the EOSC, existing data infrastructures will remain as the go-to sources for researchers. Convergence between them would continue to occur but it would be uncoordinated and slow, involving significant costs and implying the extension of the current suboptimal (inefficient) use of resources and siloed data landscape.

Source: **Towards Sustainable Funding Models for the European Open Science Cloud**
Financial Sustainability Task Force Progress report, November 2022

Blue-Cloud DD&AS vs EOSC Data Federation Architectural models

2) Metadata catalogue: At its level 1 the Blue-Cloud Data Discovery & Access service (DD&AS) maintains a common "metadatabase" of data collections as managed and offered by each of the Federated Blue Data Infrastructures (BDI). This metadatabase follows the ISO 19115 - 19139 metadata standard, which is also used as model for the EU INSPIRE Directive. This common metadatabase, at present circa 25.000 entries, is generated using a metadatabroker (DAB) which has plug-ins towards the metadata service of each BDI . This level 1 metadatabase is also integrated into EOSC as we are exchanging its content regularly from DD&AS to the EUDAT B2FIND catalogue service, which is provided at the EOSC Marketplace. From each collection entry a reference link is given back to the DD&AS, which then allows users to drill down in collections with additional metadata search criteria, specific for each BDI, and then to access the associated data sets from the BDIs by downloading.

1) "Overlay": The Blue-Cloud Data Discovery & Access service (DD&AS) federates multiple BDIs and this is done at two levels. Level 1 at collection level facilitates to discover which BDI has what interesting data collections (see also above), while at level 2 users can downdrill into the offer of each BDI at granule level using additional search criteria which might be specific per BDI. The results are selected granule metadata entries which include data download links. These are used by the Blue-Cloud Data Broker to facilitate a shopping mechanism, whereby data packages (incl metadata) from multiple BDIs can be put in a shopping basket for downloading.

Anton Ellenbroek, FAO of the UN, Knowledge and Information Management Team of the Fisheries and Aquaculture Division

FAIRderation of fisheries data

Fisheries and Aquaculture Community

Out of the Blue

- Are fisheries and aquaculture data brought into one federated environment? (No)
- Is there a partnership to assist domain data owners to harmonize and standardize data (Yes)

Fisheries Data FAIRderation

BlueCloud; bring fisheries data to the surface; data management services on top of public data

- Global Fisheries Atlas
- Global record of Stocks and Fisheries
- Blue Cloud FNS Cloud Proof of Concept: uFish2

All services use co-designed VRE's (community of practice)

Fisheries Data FAIRderation

Fisheries Atlas; a family of VRE's for SDG14.4.1
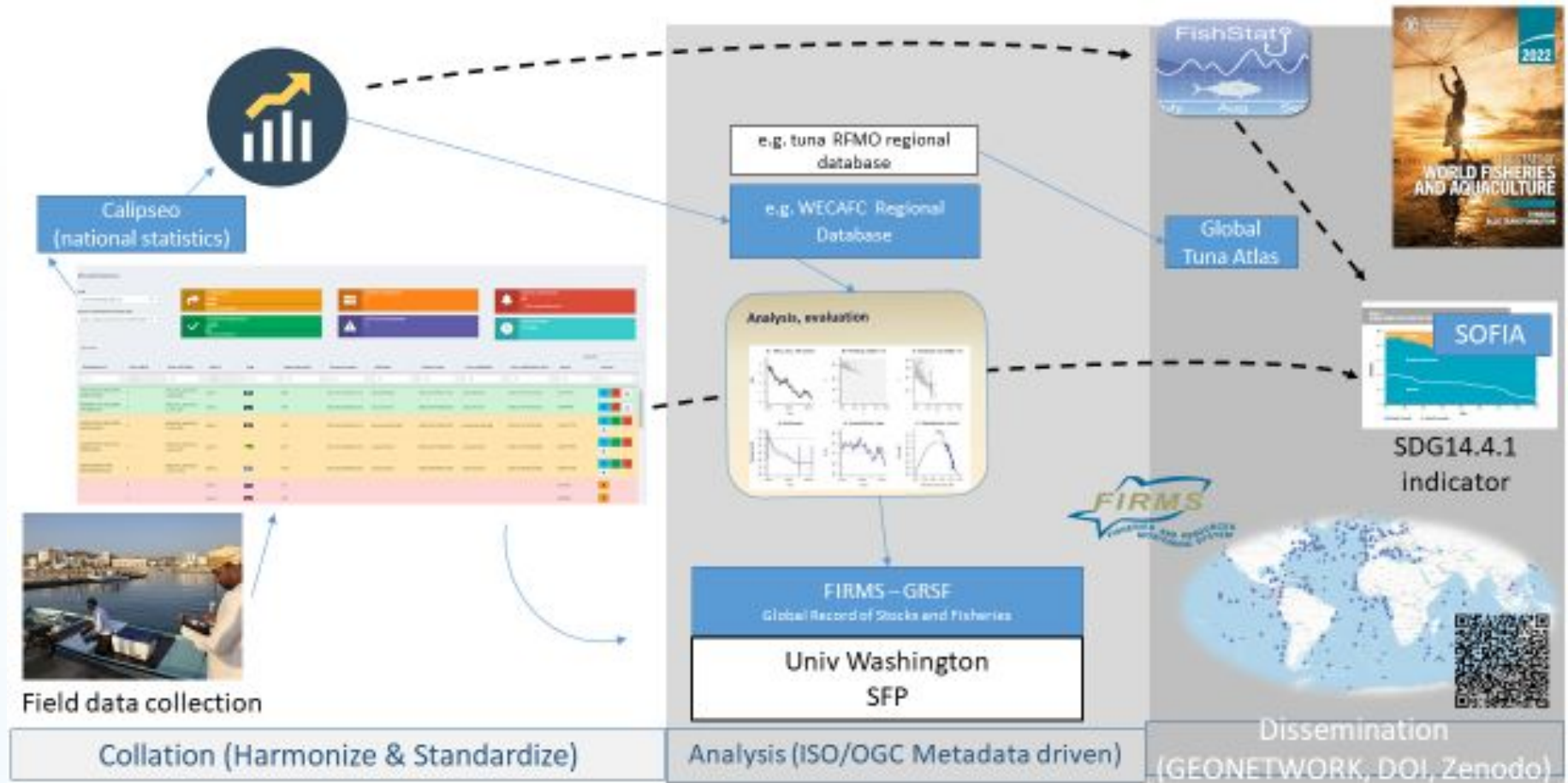
Community of practice

Virtual Research Environments

> Web-based, community-oriented, comprehensive, flexible, and secure working on-line environments.

- Communities are provided with applications to interact with the VRE services
- Client services are provided both with APIs (Java, R) and simple HTTP-REST interfaces

| The community | The data | The analytics |

FAIRderation in Fisheries Atlas

Co-designed with / funded by a flex community of practice

- Driven by SDG-needs to train / inform on global fisheries
  - SDG14.4.1 => 750 registered users
  - FAO-TunaAtlas => 68 registered users => [fao tuna atlas](#)
  - SOFIA-TAF => 75 registered users
  - GRSF => 42 registered users => [catalogue](#) and [map-viewer](#)
  - WECAFC-FIRMS => 28 registered users
- With a managed data workflow from Attribute to Zenodo

FAIRderation in Fisheries Atlas

Tuna Atlas WF

- Pilot plugin to be drafted to facilitate submission by tRFMOS for yearly updates of the FIRMS Global Tuna atlas

- **Triggers a *geoflow* to publish spatialized dataset**

- **Described by standard ISO 19115/19139 dataset metadata**

- **Structurally described by standard ISO 19110/19139 structure metadata**

- **Backed by OGC data services enabling rich multi-dimensional filtering**

- Feeds the FIRMS Global Tuna Atlas map viewer application

- Data stored in repository, when relevant published

WECAFC FIRMS WF

- Compiles country datasets on data task basis, into a regional dataset (timeseries) that progressively expands over the years

- **Triggers a *geoflow* to publish spatialized dataset**

- **Described by standard ISO 19115/19139 dataset metadata**

- **Structurally described by standard ISO 19110/19139 structure metadata**

- **Backed by OGC data services enabling rich multi-dimensional filtering**

- Feeds the WECAFC-FIRMS map viewer application

- Data stored in repository, when relevant published

The Global Record of Stocks and Fisheries

Funded by BlueCloud, FAO, Univ Washington, SFP, UNDP

FAIRderation with the GRSF

Three important components of an FDBS are autonomy, heterogeneity and distribution

- Autonomy; 3 global data providers
- Heterogeneity; Management / Science / Traceability data
- Distribution; a truly global fisheries information system

Data, data everywhere, not a byte to eat …. 2000 stocks with very heterogeneous data tales and tails

GRSF collaboration

Community demand for interoperability > federation

1. Register stocks and fisheries as part of a global repository (Semantic KB)
2. Publish UUID's as stable identifiers
3. Federate knowledge on status/trends of stocks and fisheries across various sources
4. Disseminate information through e.g. Repository / Sparql / QR
5. Test (2023) with traceability companies
6. Go fishing
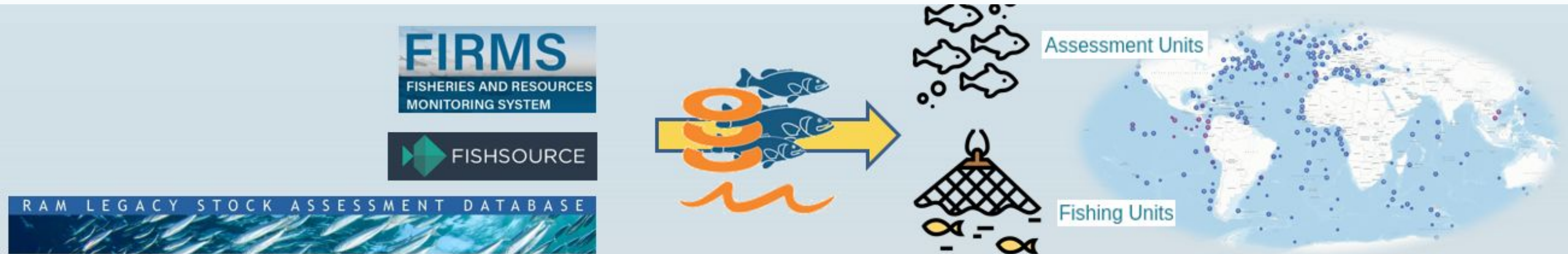
uFish2 - Fooderation

Sharing of data between fisheries domain and FNS domain

- Shared concepts, vocabularies, QA/QC??? No …
- Shared vision on provenance and traceability?? No …
- Overlap in Food descriptions => Yes
  - FoodEx2; standardised classifying and describing food
  - uFish2; a BlueCloud fish-food composition tool
  - Angular / JAVA application dockerized in D4Science
  - Now considered for 2023 funding through FAO

uFish2 for Foodsystems

FAO interest

FNS interest

Experiments on the infrastructure
Farm to Fork or
Server farm to Git fork
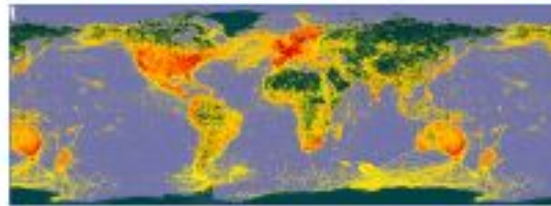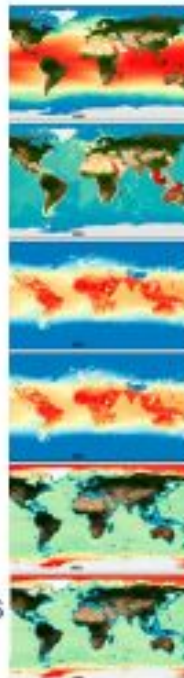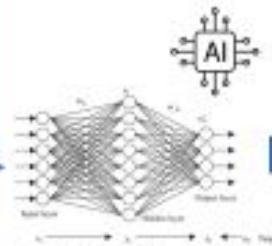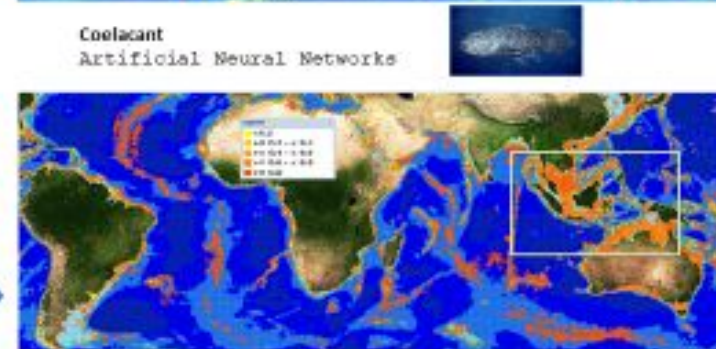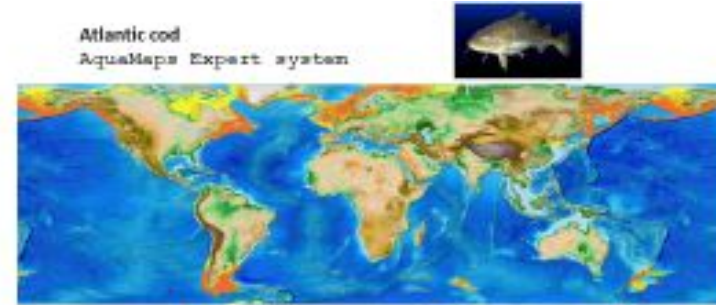
Karl Presser, Managing Partner at Premotec GmbH and FNS-Cloud H2020 representative

Data Federation within FNS-Cloud

What is FNS-Cloud?

- FNS = Food and Nutrition Security
- thematic EOSC cloud for food and nutrition
- Also called "Food Cloud"
- Sister project of Blue-Cloud
- Goal: Develop a first-generation "food cloud" by federating existing and new resources
- 35 beneficiaries
- 11 EU Member States (AT, BE, BG, DE, DK, ES, GR, IE, IT, NL, and SI), UK, Serbia, & Switzerland
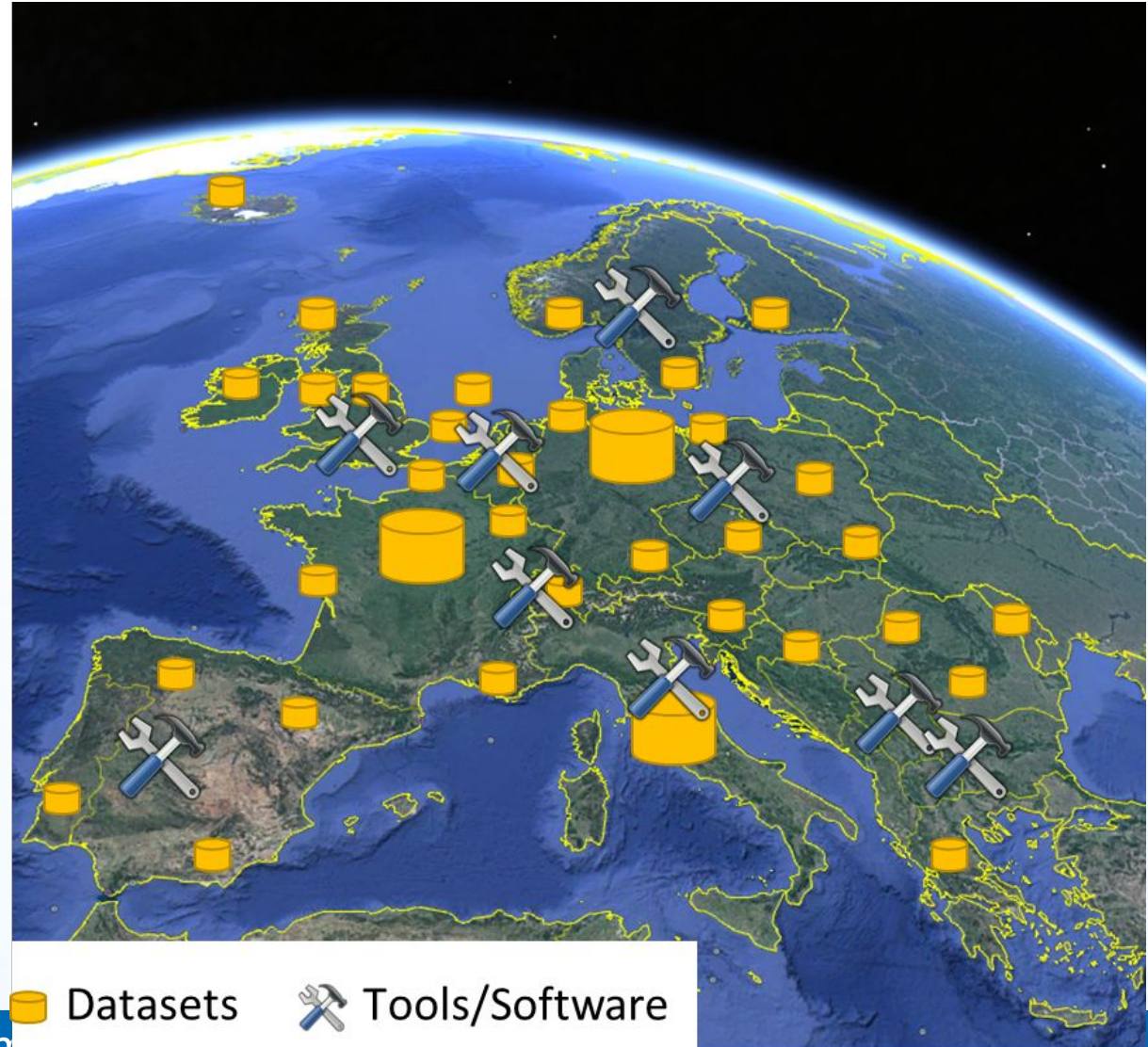
Fragmentation

- Highly fragmented
- National/institutional repositories
- Reason: National data, regional data, seasonal data, many small studies, living data
- Not many "central" databases

-> Findability is challenging and federation as well

Datasets 🛠 Tools/Software

Data Federation

- Map with FNS topics
- For each topic define data transfer models and API definitions
- Recommendation to build central databases
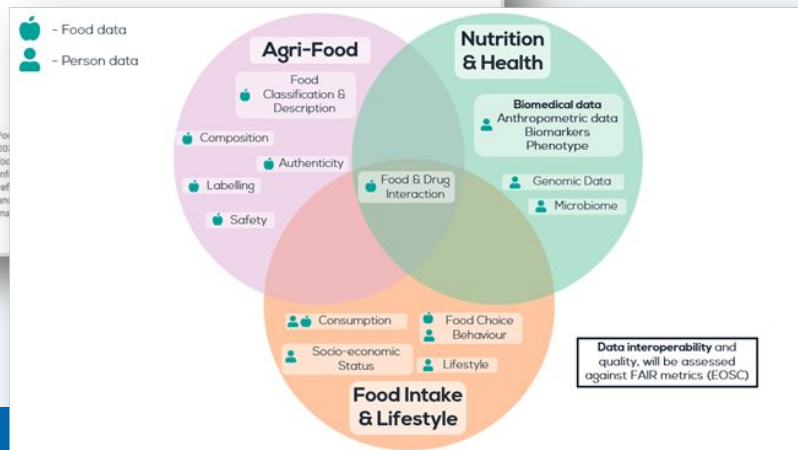- First generation ontology

5.2.5 Thesauri

Name	Reference
Country Code	https://www.iso.org/obp/ui/#search https://www.iso.org/iso-3166-country-codes.html
CODEX	http://www.fao.org/fao-who-codexalimentarius/codex-texts/list-standards/en/
GS1	https://www.barcodefaq.com/1d/gs1-id-key/
E-Number	https://www.food.gov.uk/business-guidance/approved-additives-and-e-numbers
INS	http://www.fao.org/tempref/codex/Meetings/CCFAC/ccfac31/INS_e.pdf
Component	http://www.eurofir.org/eurofir-thesauri/
Unit	http://www.eurofir.org/eurofir-thesauri/
Matrix Unit	http://www.eurofir.org/eurofir-thesauri/
Value Type	http://www.eurofir.org/eurofir-thesauri/
Reference Type	http://www.eurofir.org/eurofir-thesauri/
Method Type	http://www.eurofir.org/eurofir-thesauri/

5.3.1.1.1 Food entity

Attribute	Description	(M)andatory/ (O)ptional (U)nique	Type
country	ISO 3166 code	M	THS
id	Primary key	M;U	KEY
name	Food name	M;U	STR
englishName	Food name in English	M;U	STR
timestamp	date and time of data collection	M	DAT(TIM)
technicalName		O	STR
ingredientList	List of ingredients listed on the package	O	STR
additions	Additions listed on the package	O	STR
allergenInformation		O	STR
servingsPerPack		O	NUM
servingSuggestions		O	STR
finalPreparation		O	STR
samplingPlace	place of data collection, for example a store name	O	STR
producer	Producer name	O	STR
dietaryClaims	Health&nutrition claims	O	STR
netWeight	Net weight value	O	NUM
netWeightUnit	Net weight unit	O	THS
marketShare	market share value	O	FRC
physicalState	Cooled, Frozen, Dried, Uncooled, Conserved/can, Others	O	THS
packagingMaterial	Plastic, Metal, Paper	O	THS

![Blue-Cloud]
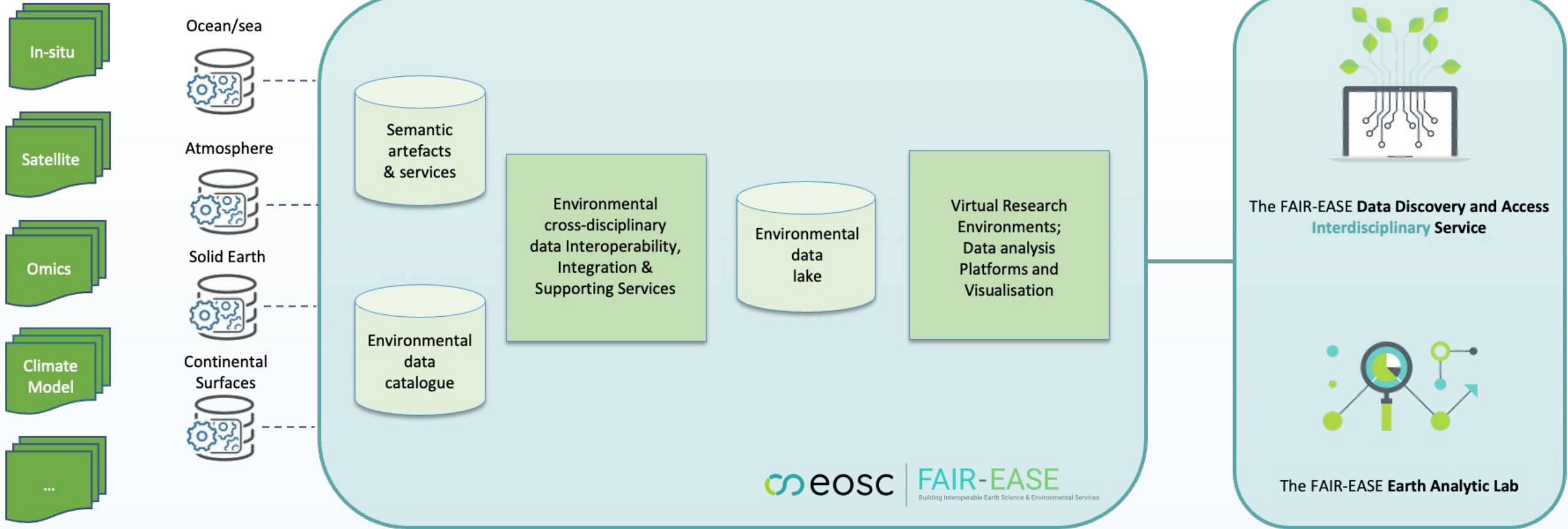
Alessandro Rizzo, IRD and FAIR-EASE

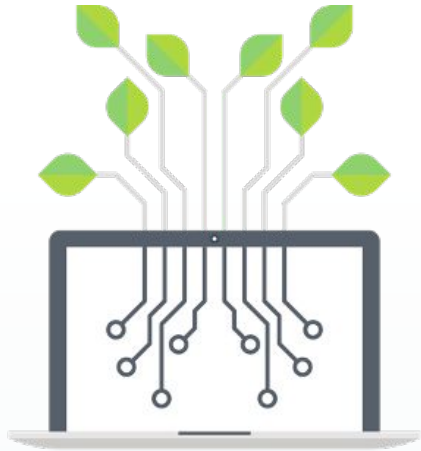# Future evolution of Interoperability & FAIR data management aspects in Blue-Cloud 2026

FAIR-EASE services... a way to federate data for Earth System and Environment

The FAIR-EASE **Data Discovery and Access Interdisciplinary Service** will provide users with an easy and FAIR service for discovery of and access to multidisciplinary and aggregated data sets.

The FAIR-EASE **Earth Analytic Lab** will provide users with an easy way to visualise, analyse and process environmental data on-demand, that is, according to their specific objectives, thematic, geographical areas and time periods of interest.

FAIR-EASE and Blue-Cloud 2026: from an interdisciplinary approach to synergy effects

EARTH AND ENVIRONMENTAL DYNAMICS

ENVIRONMENTAL BIO-GEOCHEMICAL ASSETS

BIODIVERSITY OBSERVATIONS
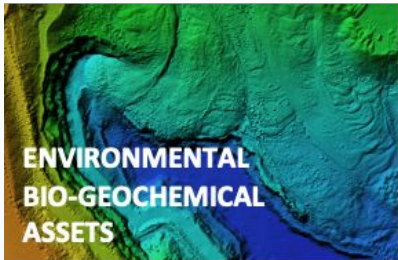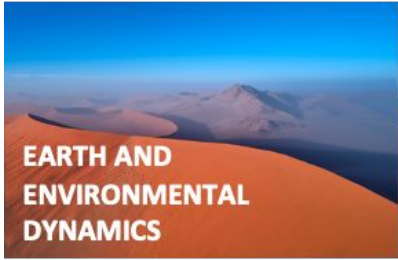
Interdisciplinary

Involving two or more areas of knowledge

Synergy
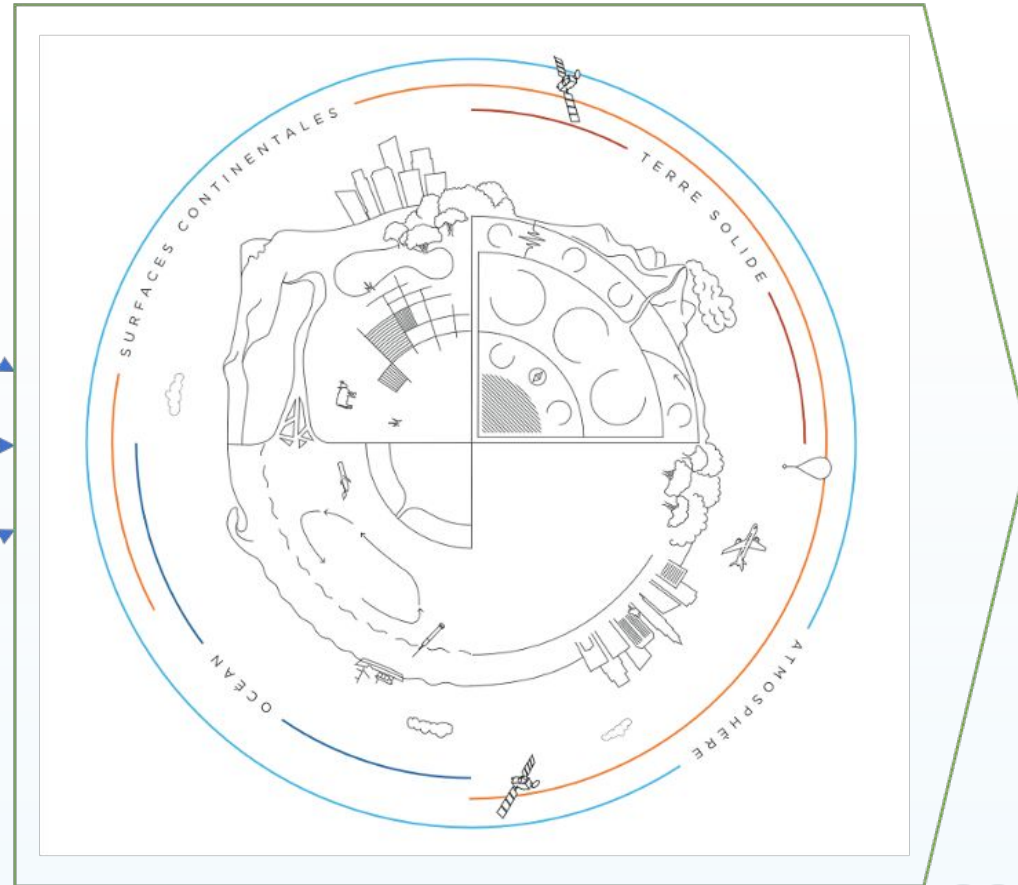
Working together

Credits: www.data-terra.org
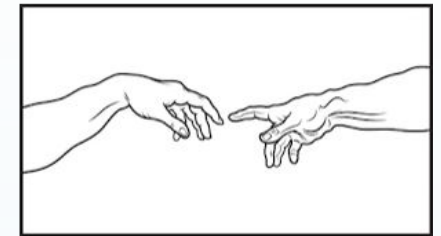
Panel discussion

A sustainability model for data federation in EOSC : menti.com/ 8466 1243

Question to Alessandro Rizzo

Participation of Blue Data Infrastructures in these initiatives has benefitted them by further developing their sustainability base and helping propagate standards and services. This shows a clear need and opportunity for cooperation and synergy between RIs (i.e. BDIs) and the EU-lead initiatives.

Source: **Towards Sustainable Funding Models for the European Open Science Cloud**
Financial Sustainability Task Force Progress report, November 2022

What are your recommendations with respect to cross-disciplinary, cross-national and cross-project collection and use of data? How can cross-project collaboration stimulate FAIRness of marine data and boost data management sustainability?

Questions to
Anton Ellenbroek and Karl Presser

Public funding is the most visible part of the funding landscape. However, in a situation that resembles that of an iceberg, public and private funding at MS level underneath is actually much larger.

Source: **Towards Sustainable Funding Models for the European Open Science Cloud**
Financial Sustainability Task Force Progress report, November 2022

Given your experience of interaction with commercial entities, what are your recommendations on how to build a federation that also takes into account the needs and exploitation opportunities of the private sector?

Open question

What should a data federation look like?

Blue-Cloud use case demonstrates that services & data go hand-in-hand.

Do you see Blue-Cloud as a replicable model for Data Federation?

menti.com/ 8466 1243

Next steps

We are going to have a closer discussion with Blue-Cloud data providers in early 2023 on how much accessing their data on an yearly basis cost, to deliver some funding model scenarios.

MISSION: To develop further the European federation of marine and inland water data management infrastructures & high quality services

A1. DD&AS
A FAIR compliant Data Discovery & Access Services > access to 10+ million open data sets & products by 13 major BDIs

A2. VRE
An Open Science Virtual Research Environment (VRE) federating multiple e-infrastructures > supporting Analytical Big Data Workbenches & VLabs

A3. EOVs
3 EOV Workbenches for highly qualified data collections

3.000 DATA ANALYTICS SESSIONS PER MONTH - 5,000 HTC DATA ANALYTICS JOBS PER MONTH

A4. VLABs - FIVE DOMAIN-BASED VIRTUAL LABS

Coastal Ocean observations along Europe

Coastal currents from observations

Carbon-Plankton Dynamics

Marine Environmental Indicators

Global Fisheries Atlas

A7. COMMUNITY
- **All EU countries** engaged
- **3k+** engaged Blue- Cloud community users
- **5k+** followers across all the platforms
- **10+** External Stakeholders

OUTREACH
- **1** Blue-Cloud Hackathon
- **1** Blue-Cloud TV
- **18** Newsletter issues
- **11** Webinars on Blue-Cloud VRE, DDAS & EOV Workbenches
- **3** Blue-Cloud Annual Impact Events
- **3** Ocean Literacy Webinars
- Videos & Interviews

A6. TRAINING ACADEMY & CATALOGUE
- **3** Online training course on Best Practices for FAIR data principles
- **3** Info session & course on the EOV Workbenches
- **2** online webinars dedicated to the BlueCloud VRE
- **2** dedicated to the DDAS and the innovations introduced
- A series of training sessions on how to use the VLabs

POLICY
- Scientific papers & articles
- Restoring healthy oceans, seas, coastal & inland waters in Europe
- Strategic Roadmap 2030 **A5. ROADMAP**
- Cross-domain expansion factsheets
- Sustainability Business model

DTO Task Force

In Blue-Cloud 2026 further optimisation is planned by including 1) semantic brokerage as BDIs might use different vocabularies for several metadata fields and 2) streamlining the web services of each BDI towards a more common functional model.

Thank You!

Data Services (in-kind from MS)

MS provide many in-kind data services through their national activities as deployed by research institutes and national bodies. Data collection is mostly done at national level, as part of scientific activities or of governmental monitoring for environmental management or safety purposes. On a second level, MS deploy monitoring programs at national and regional level to support national marine governance goals, EU policies, and growing demands for sustainable use of resources. This data are complemented by further data gathered by research institutes through e.g. research vessels, and furthermore by data collected by private organisations in support of economic activities. The accompanying data management is also done mostly at national level . MS/AC are also heavily involved in the implementation of EU directives that rely on the data to determine the current environmental status of their marine ecosystems, identify threats and establish measures to restore or maintain ecosystem health.

Current Sources of Funding

In kind contributions from MS/AC coexist with funding by the EU for exchange systems and common standards at European level that allow overview of and access to the data collections (this is the most visible part of the funding landscape; however, in a situation that resembles that of an iceberg, public and private funding at MS level underneath is actually much larger). Furthermore, most BDIs have been developed and built 'bottom-up' from community initiatives with co-funding from EU DG RTD and the ESFRI programme, complemented since 2008 by funding from other EU Directorates, launching the 'top-down' initiatives EMODnet by EU DG MARE, COPERNICUS by EU DG DEFIS, as well as MSFD-related data projects by EU DG ENV and EEA—now further complemented by EOSC, which has received funding from DGs RTD and CNECT.

The role of BDIs

Participation of BDIs in these initiatives has benefitted them by further developing their sustainability base and helping propagate standards and services. This shows a clear need and opportunity for cooperation and synergy between RIs (i.e. BDIs) and the EU-lead initiatives for the benefit of both sides: RIs get access to networks of data originators and can manage pipelines for structured sharing of their data, while top-down initiatives focused on deriving data products and services for the community are thus able to obtain data from in-situ and remote sensing resources. For the Blue-Cloud DD&AS it is instrumental to align its position with the European infrastructures and to ensure that the BDIs and major initiatives as EMODnet, COPERNICUS, and EOSC receive added-value.