



Data management and provision by Euro-Argo ERIC Argo ocean observing system

Thierry Carval – Ifremer - Euro-Argo – Blue Cloud



- Euro-Argo ERIC and Marine RIs make continuous efforts on FAIR principles with trainings and experts. Our yearly FIP exercise is a significant effort for Marine RIs FAIRness.
- Each Marine RI now have FAIR data experts working within Blue-Cloud community
 - In 2019 Marine RIs data management was FAIR for people but not for machine (GO-FAIR FIP, ENVRI-FAIR D9.1)
 - Now the Marine RIs are significantly more FAIR on the EOSC – Blue Cloud

- Published on Zenodo <https://doi.org/10.5281/zenodo.7505613>
 - Yearly FIPs exercises (2019 – 2022)
 - We use the FIPs SPARQL endpoint to analyze Marine RIs progress
 - Each RI improved its FAIRness within 3 years, **with an average of 11 new FERs**
 - The heatmaps of our common FERs illustrates our RIs convergence
FER: FAIR Enabling Resource

2019	SeaDataNet- CDI	lw-marine	ArgoGdac	EMSO	SeaDataNet- Sextant	ICOS
SeaDataNet-CDI		1	5		4	5
lw-marine	1		6		4	10
ArgoGdac	5	6			6	13
EMSO						
SeaDataNet-Sextant	4	4	6			10
ICOS	5	10	13		10	

2020	SeaDataNet- CDI	lw-marine	ArgoGdac	EMSO	SeaDataNet- Sextant	ICOS
SeaDataNet-CDI		4	7		6	10
lw-marine	4		7		6	13
ArgoGdac	7	7			6	16
EMSO						
SeaDataNet-Sextant	6	6	6			11
ICOS	10	13	16		11	

2021	SeaDataNet- CDI	lw-marine	ArgoGdac	EMSO	SeaDataNet- Sextant	ICOS
SeaDataNet-CDI		5	10	8	17	13
lw-marine	5		7	4	7	13
ArgoGdac	10	7		14	11	16
EMSO	8	4	14		8	10
SeaDataNet-Sextant	17	7	11	8		14
ICOS	13	13	16	10	14	

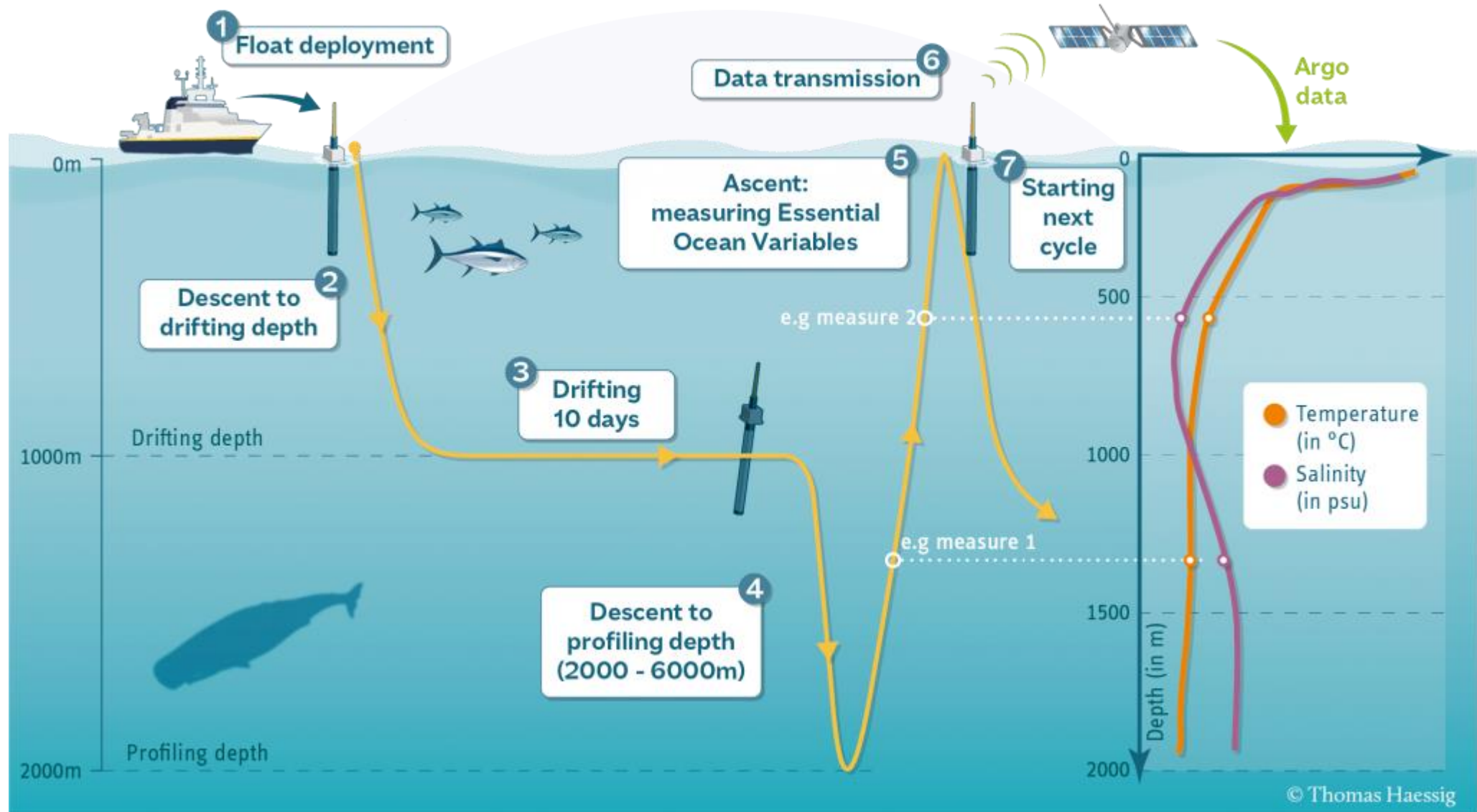
Vocabularies, a priority to cross domain and integration science

- We expose our vocabularies on vocabulary servers.
Before Blue-Cloud, marine ERICs did not use a vocabulary server, only conventional documentation:
 - our RIs were FAIR for humans (who would read the documentation)
 - our RIs are now FAIR for humans and for science cloud communities (machine-to-machine workflow through vocabulary servers)
- Each marine RI manages its vocabularies
 - With **smart mappings** between RIs vocabularies
 - Marine vocabs community activity
 - http://vocab.nerc.ac.uk/search_nvs/
 - <https://github.com/nvs-vocabs/ArgoVocabs>

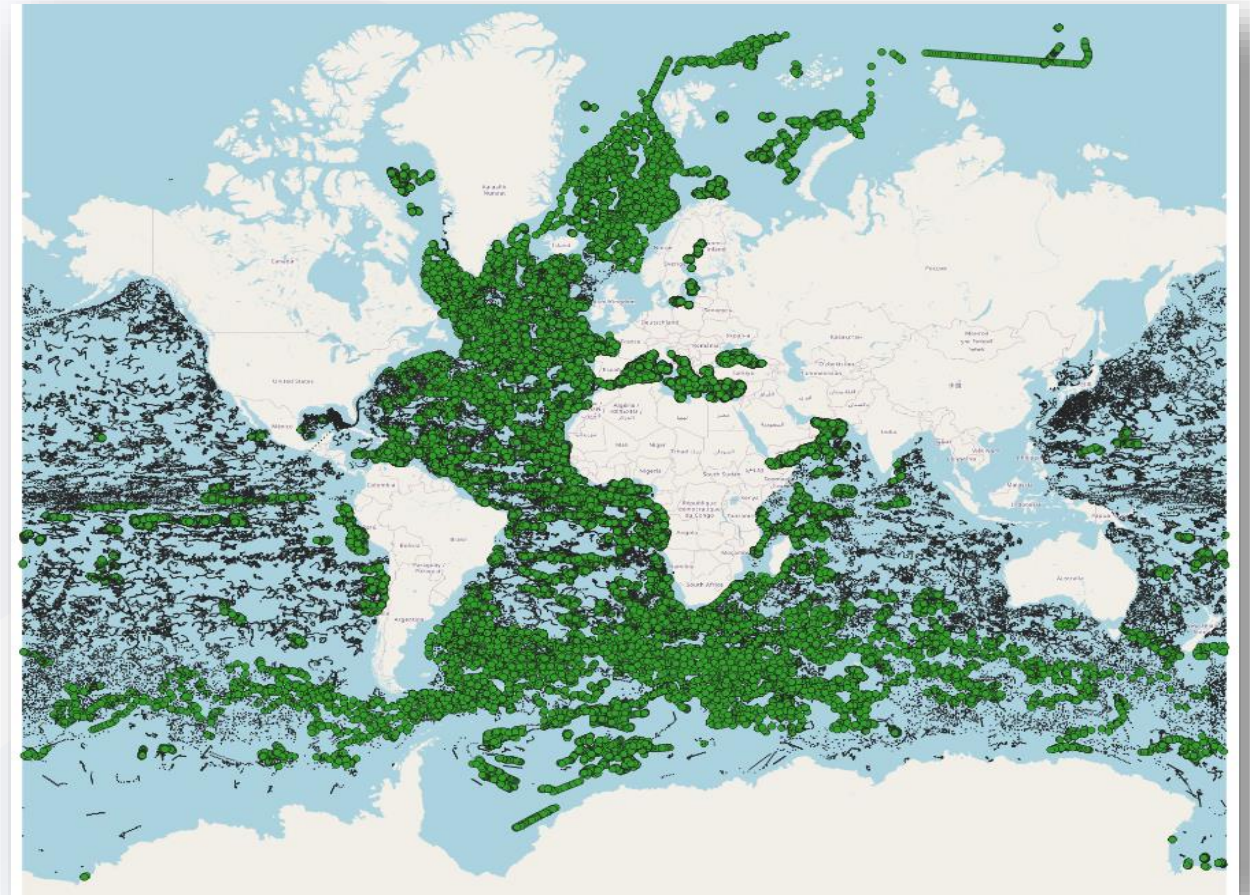
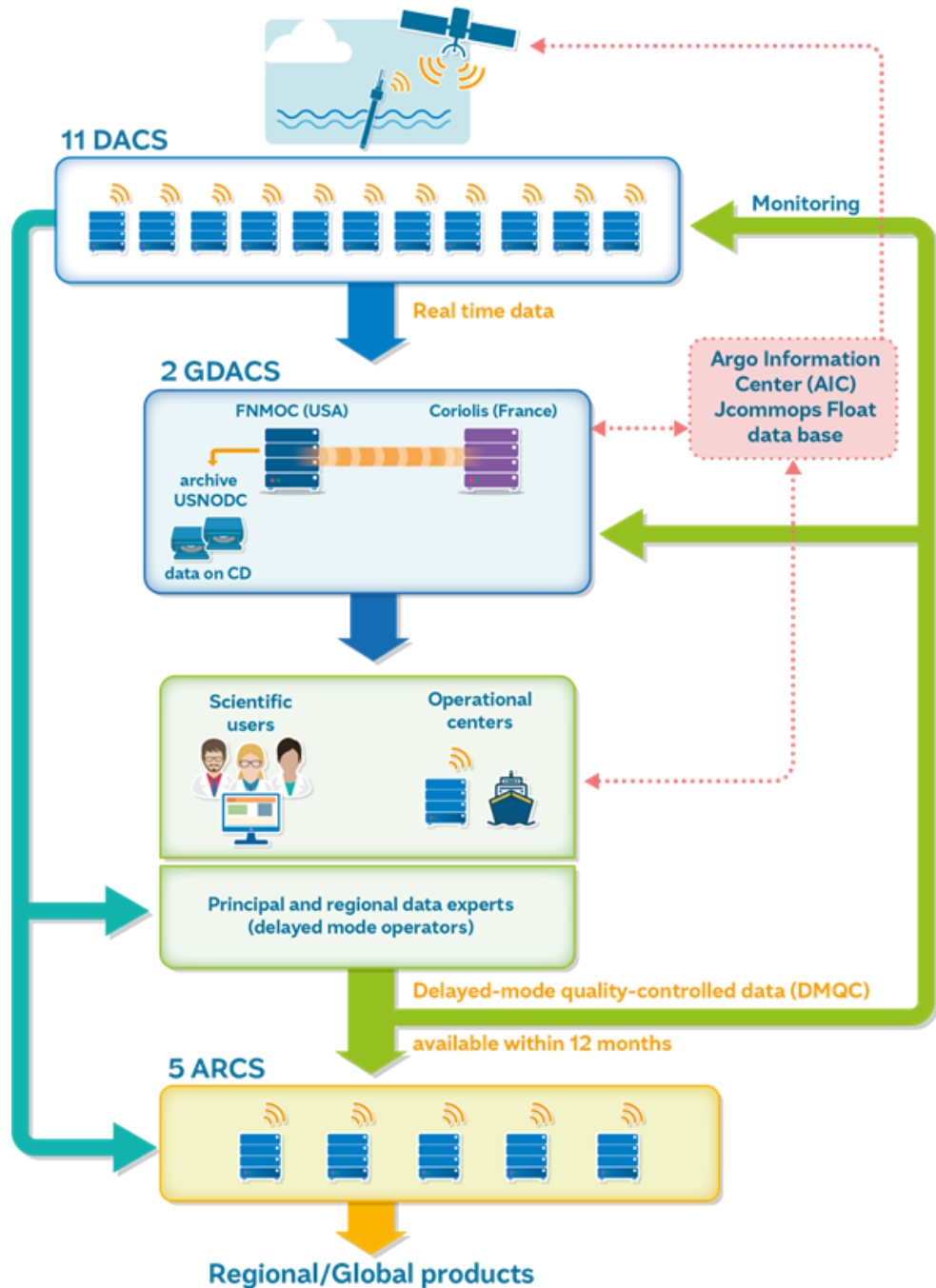
Semantic web: a major contribution for cross domain and integration science

- Beyond the vocabularies, ontologies are managed for RI specific domain
- A triple store contains the metadata described with the ontology. A SPARQL endpoint queries this ontology.
- We now provide semantic queries on our metadata with 7 SPARQL endpoints, and more to come.

- Argo is an international program to observe the interior of the ocean with a fleet of profiling floats drifting in the deep ocean currents (<https://argo.ucsd.edu>). Argo GDAC is a dataset of 5 billion in situ ocean observations from 18.000 profiling floats (4.000 active) which started 20 years ago. Argo GDAC dataset is a collection of 18.000 NetCDF files. It is a major asset for ocean and climate science, a contributor to IOCCP reports.
- Argo data are available with a one click download <https://doi.org/10.17882/42182>



Argo data system



Argo performed 160 000 ocean profiles in 2023 (Euro-Argo dac in green)

- Argo data management is performed on 3 levels
 - PI : principal investigator for a series of floats
 - DAC : data assembly centre for a series of PIs
 - GDAC : global data assembly centres for all the DACs
- The Argo global data set is continuously and homogeneously managed, improved, in real-time and delayed mode

• Each DAC

- Produces Argo NetCDF meta, profile, traj and tech files
Argo NetCDF format and vocabularies <https://doi.org/10.13155/29825>
- Pushes Argo NetCDF files on GDAC

• The GDAC

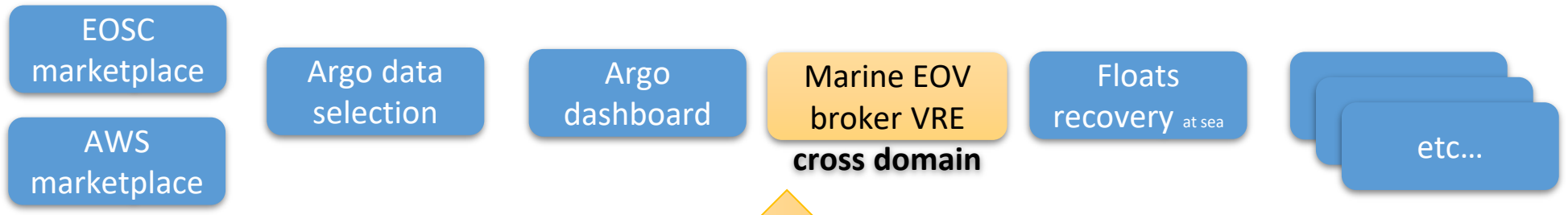
- Collects the Argo files
- Applies the file format checker to accept/reject files
<https://github.com/euroargodev/ArgoNetCDF>

There are 11 DACs, their role is to:

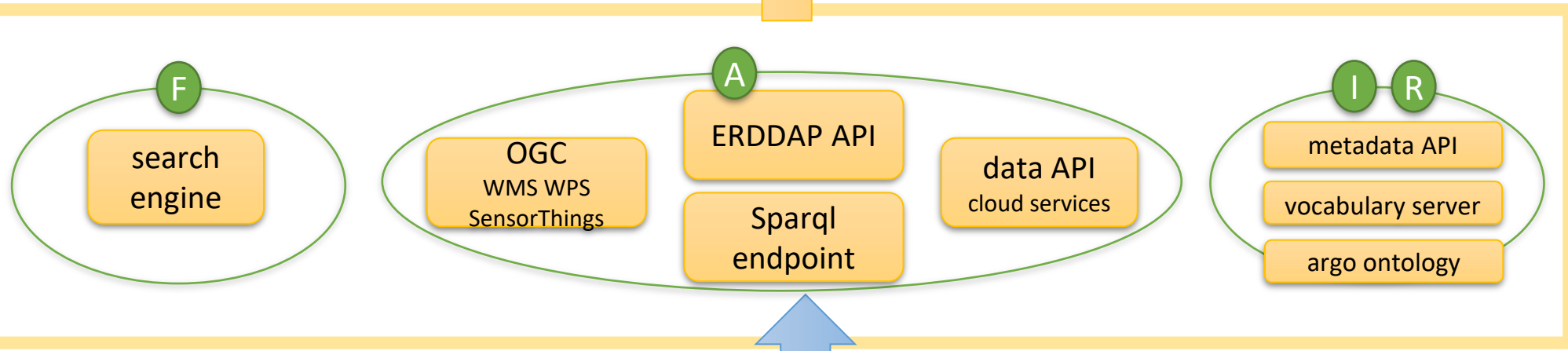
- Collect and archive real-time data (iridium, argos)
- Manage metadata published on OceanOPS
- Decode, format, apply RTQC
- Collect and check DMQC data
- Push NetCDF files on GDAC
- Push real-time data on GTS as BUFR bulletins

This is documented on <http://www.argodatamgt.org/Documentation>

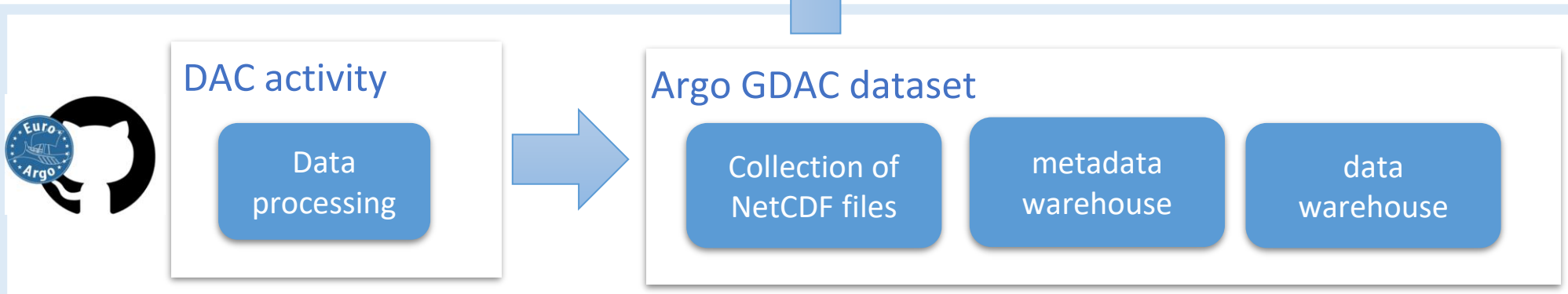
Argo is well represented on EOSC Blue-Cloud services



Front office
Stakeholders services:
EOSC Blue cloud,
Copernicus, Argo...



Back office
Machine-to-machine
services Implementation
plan within ENVRI-FAIR
for **F A I R**
enhancements



Marine RI infrastructures
Argo GDAC, EMSO,
Lifewatch, ICOS,
SeaDataNet data
repositories

Argo metadata lake, a combination of NetCDF – Parquet – Elasticsearch

Argo NetCDF
2 million files

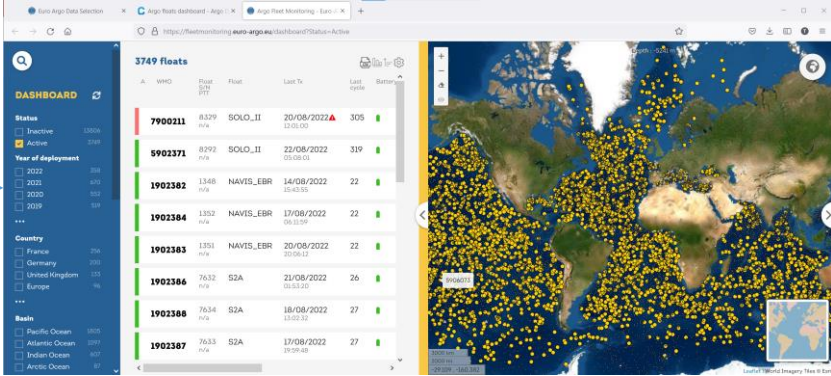
co04060109
NetCDF to parquet
on spark cluster

Argo parquet
3000 files

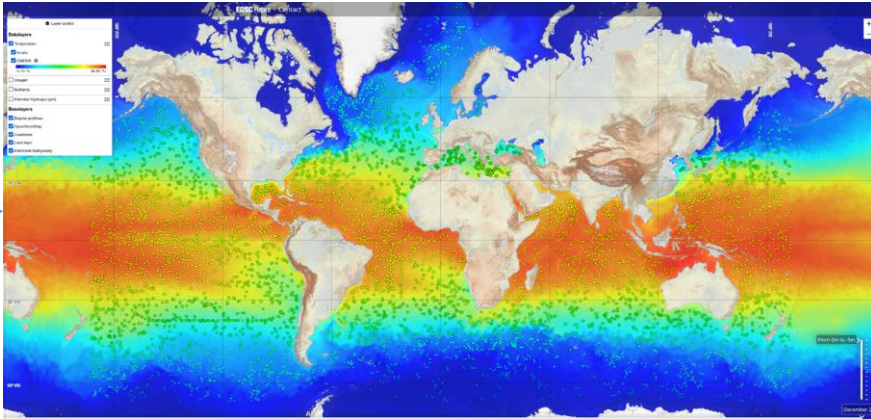
co050147
parquet to Elastic
on Elastic cluster

Elasticsearch
18 000 items
10 million
metadata

Argo metadata
API
on tomcat server



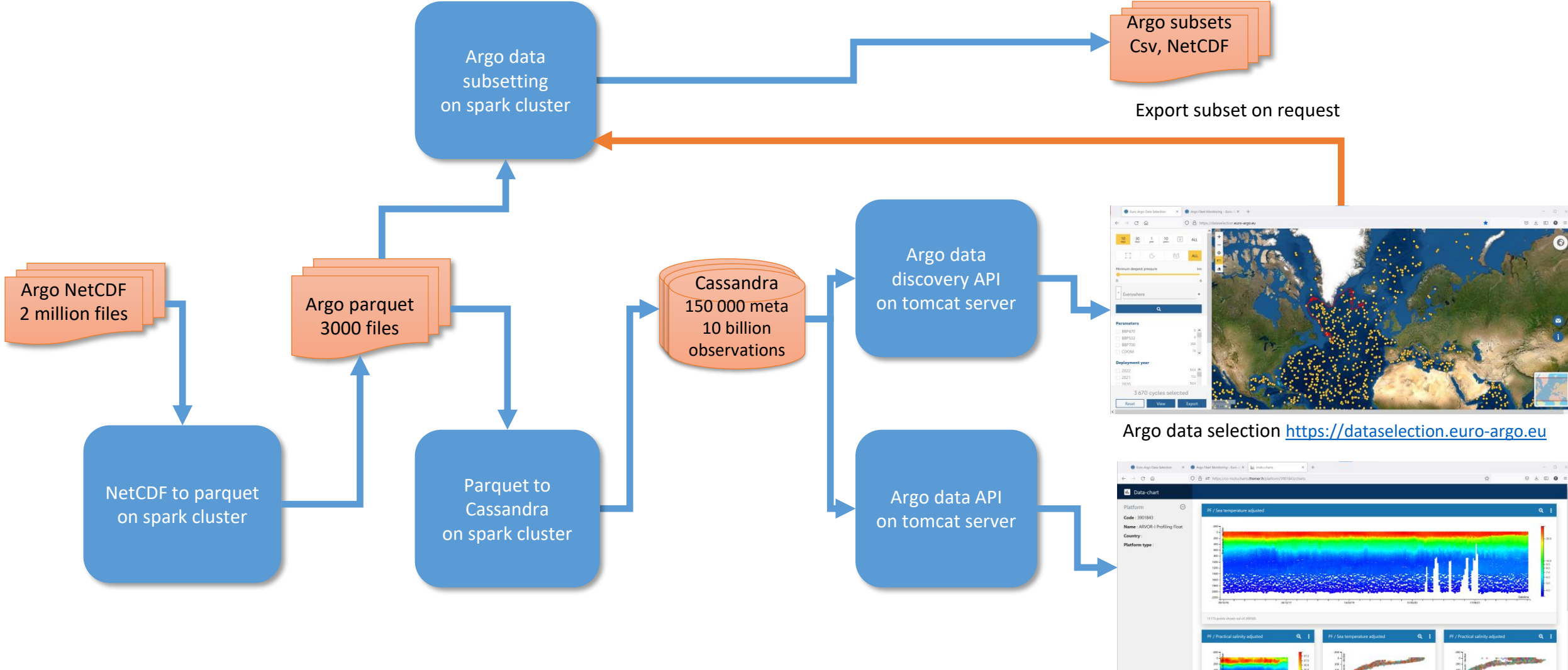
Argo floats dashboard <https://fleetmonitoring.euro-argo.eu>



Preview EOSC-Future demo application: Explore co-location of SeaDataNet CDI - Argo - Copernicus marine data

Argo data lake

A combination of NetCDF – Parquet – Cassandra - Elasticsearch



Argo data selection <https://dataselection.euro-argo.eu>

Argo data charts <https://co-insitucharts.ifremer.fr/platform/3901843/charts>

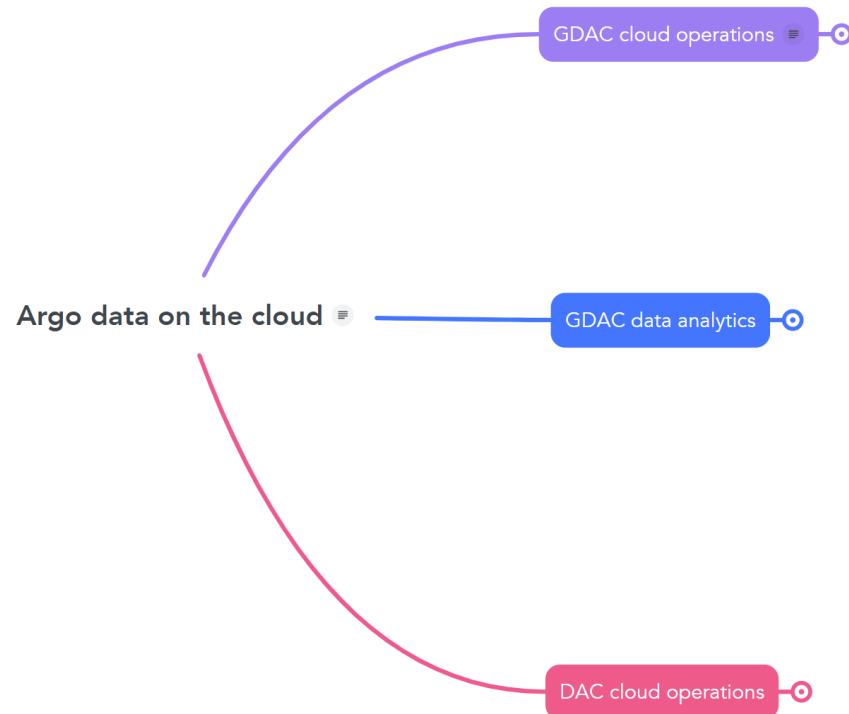
- OneArgo : a new global design is under development
 - towards spatial completeness to include Polar sea-ice zones and marginal seas
 - increasing regional resolution in key areas like the Western Boundary currents and equatorial regions
 - major new missions: BioGeoChemical BGC-Argo and Deep-Argo missions
- Challenges
 - Significant new resources are needed – roughly 4-5 times the current investment
 - Requires strong community support and advocacy
 - Ensure our **data management teams are ready and able to handle the new data streams**
 - Continue to grow and strengthen our national and international partnerships

The Argo DAC-GDAC infrastructure, based on FTP transfers, was conceived a quarter of a century ago. A significant transition towards a cloud-based infrastructure is needed, driven by the following key objectives:

- **Scalability:** to accommodate the evolving demands of Argo data management and ensure seamless growth.
- **Reliability:** enhancing the trustworthiness and consistency of data operations, minimizing downtime.
- **Robustness:** strengthening the infrastructure's resilience and capacity to withstand challenges and disruptions.

This shift to a cloud-based approach will improve the efficiency of Argo data management and align it with contemporary technological standards.

- To have a clear idea of what is needed and where we go, here is a mind map listing the functionalities of **Argo cloud services**.



eOSC | Blue-Cloud2026



blue-cloud.org



[@bluecloudeu](https://twitter.com/bluecloudeu)



[blue-cloud org](https://www.linkedin.com/company/blue-cloud-org)



Funded by
the European Union