Fair principles and ENVRI-FAIR

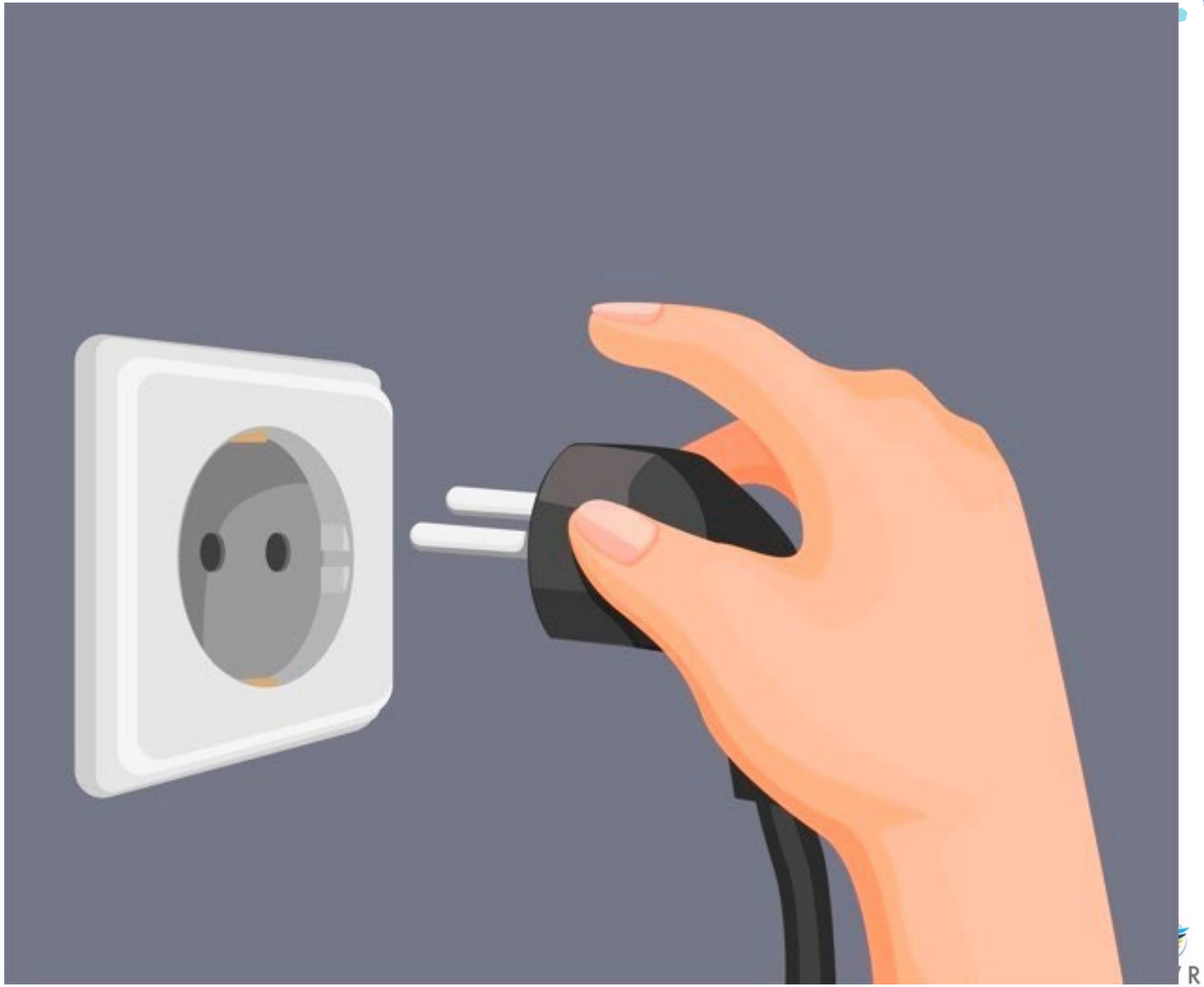
Sylvie Pouliquen

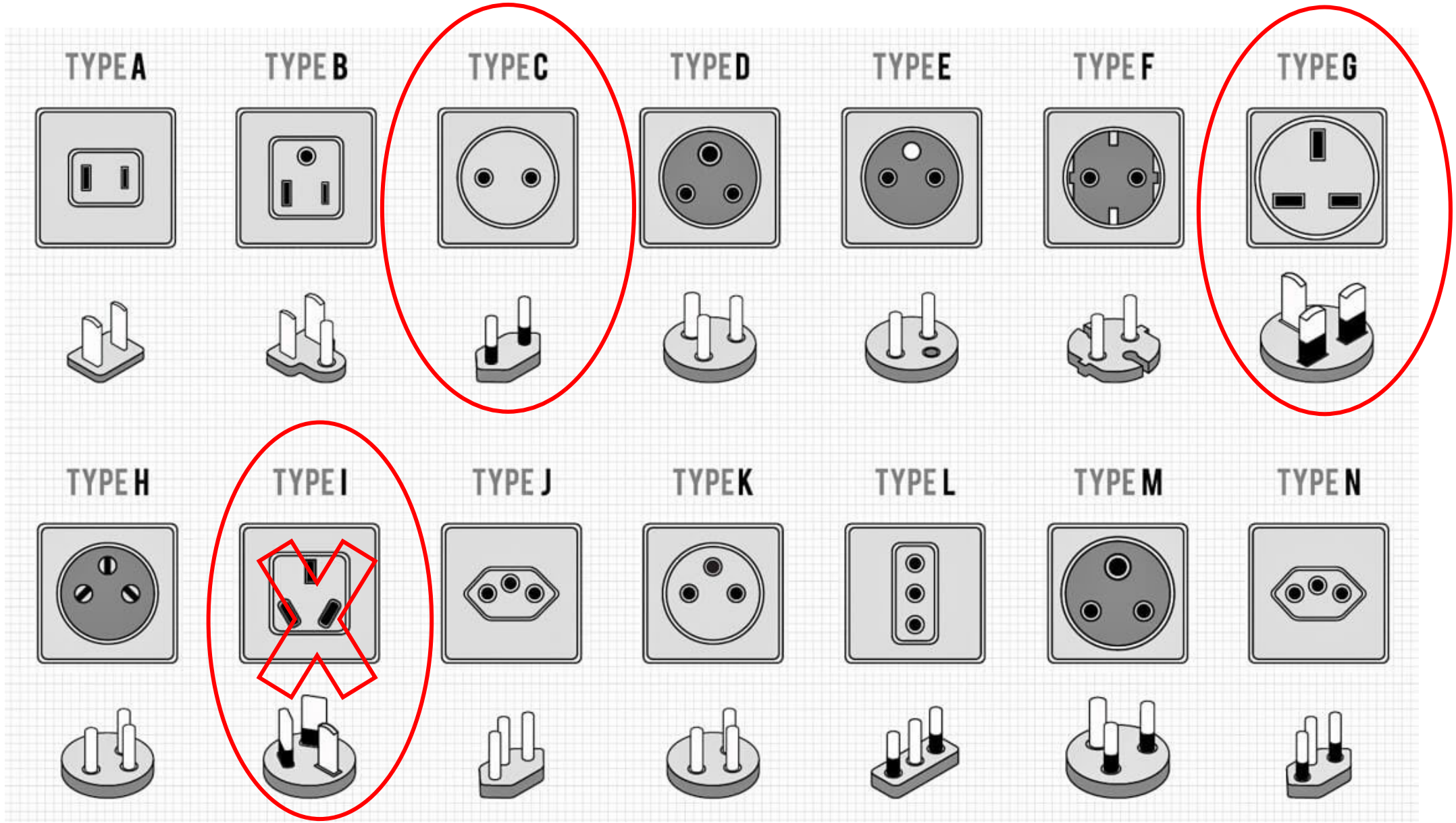
Ifremer Brest France

ENVRI-FAIR has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824068

The issue

The plug




TYPE A
TYPE B
TYPE C
TYPE D
TYPE E
TYPE F
TYPE G
TYPE H
TYPE I
TYPE J
TYPE K
TYPE L
TYPE M
TYPE N

The FAIR solution

The universal adapter

Best Practices Supporting Data and Innovation

What is FAIR?

The **FAIR Data Principles** are a set of guiding **principles** in order to make **data** findable, accessible, interoperable and reusable

Tanhua T, Pouliquen S, Hausman J, O'Brien K, Bricher P, de Bruin T, Buck JJH, Burger EF, Carval T, Casey KS, Diggs S, Giorgetti A, Glaves H, Harscoat V, Kinkade D, Muelbert JH, Novellino A, Pfeil B, Pulsifer PL, Van de Putte A, Robinson E, Schaap D, Smirnov A, Smith N, Snowden D, Spears T, Stall S, Tacoma M, Thijsse P, Tronstad S, Vandenberghe T, Wengren M, Wyborn L and Zhao Z (2019) Ocean FAIR Data Services. Front. Mar. Sci. 6:440. doi: 10.3389/fmars.2019.00440

Snowden D, Tsontos VM, Handegard NO, Zarate M, O' Brien K, Casey KS, Smith N, Sagen H, Bailey K, Lewis MN and Arms SC (2019) Data Interoperability Between Elements of the Global Ocean Observing System. Front. Mar. Sci. 6:442. doi: 10.3389/fmars.2019.00442

ENVRI FAIR

Best Practices Supporting Data and Innovation

What are the benefits of FAIR?

Example:

Booking multi-destination tickets via different airlines through a single travel website

Image: https://www.indianeagle.com/traveldiary/booking-flights-through-travel-agents-benefits/

Best Practices Supporting Data and Innovation

What are the benefits of FAIR?

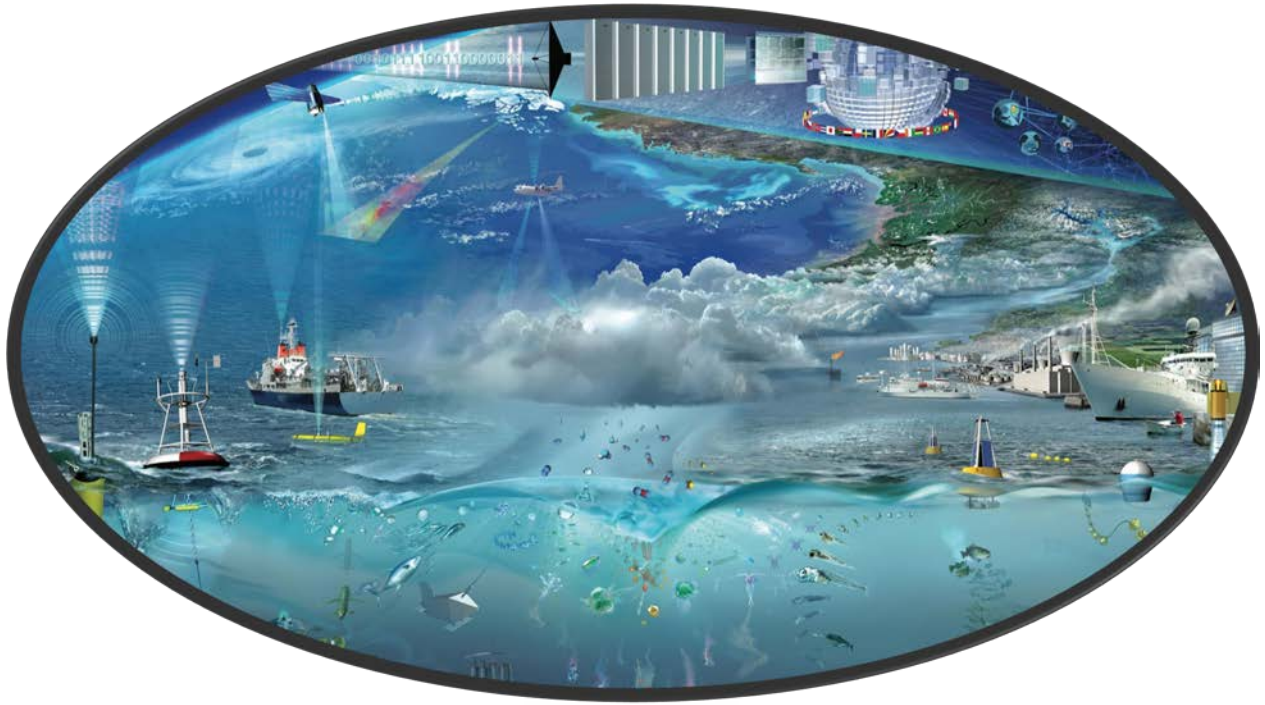


Image: https://www.atlantos-h2020.eu/ocean-observation/

For Environment community:

Aggregate observations from multiple systems with as few human interactions as possible to concentrate efforts on information development

Best Practices Supporting Data and Innovation

How do I comply with FAIR?

- ***Adhering to standards and best practices*** in data management will lead to higher levels of FAIR compliance and ensure the establishment of data frameworks capable of leveraging future technological advances (Cloud, Machine Learning, Big Data)
- ***Do not reinvent the wheel*** and, if possible, work closely with **professional data centers**
- Using standards and best practices may incur **a small initial cost but *fixing issues in the future will be much more costly*, if not impossible.**

In addition to being FAIR, Data needs to be free and openly available if we want to move from data to information development.

Best Practices Supporting Data and Innovation

The Argo Example

Current

Findable:

Detailed metadata

- ***but some are only described in the manual***

Accessible

Services

- ***All data in a single point FTP architecture described in the manual***

Pretty FAIR for people, still work to be done for machines

Ideal

Findable:

Detailed metadata

- ***Metadata fully described in manual AND community vocabulary server***

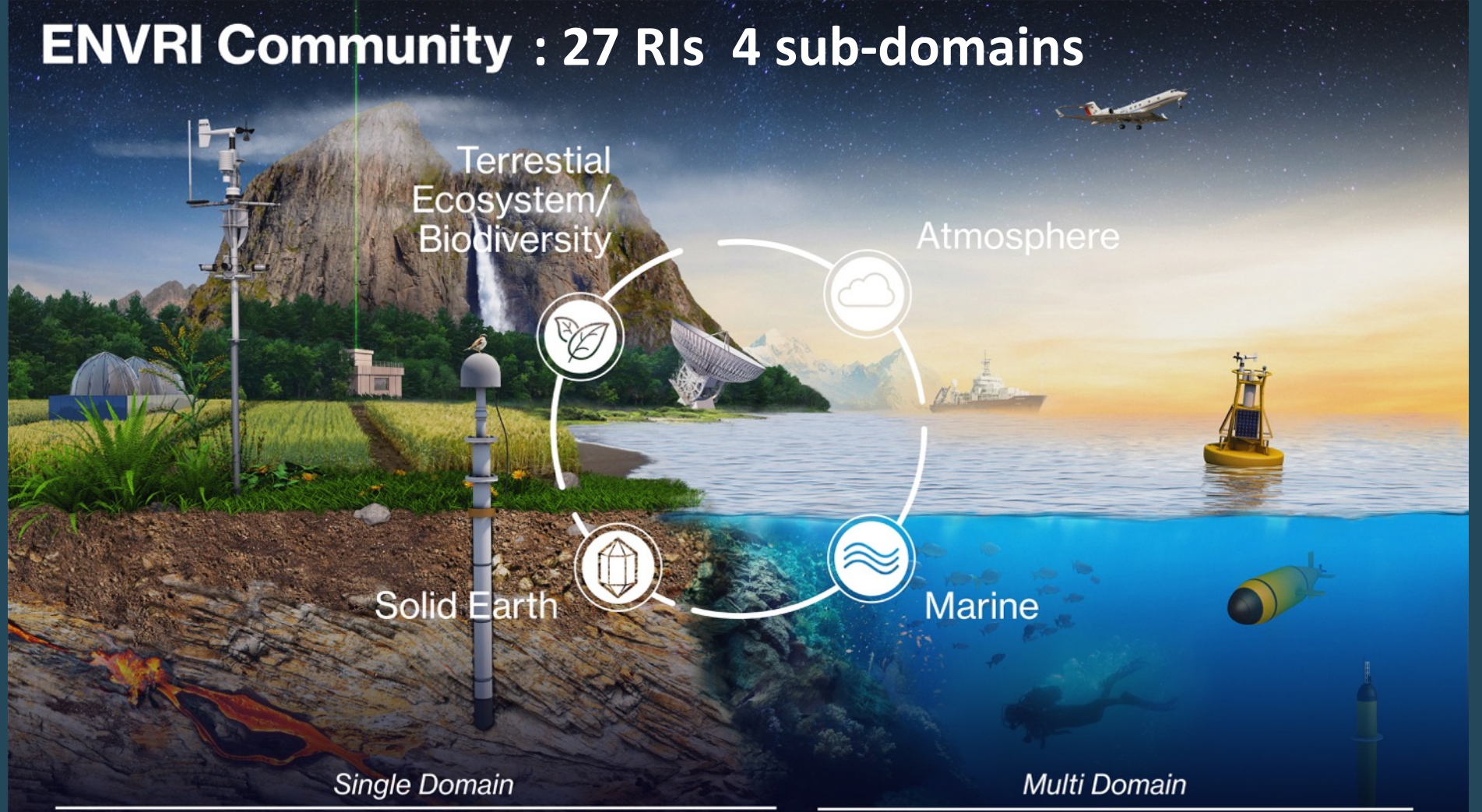
Accessible

Services

- ***Implement data and metadata API allowing interoperable access***

FAIR for people AND machines

ENVRI Community : 27 RIs 4 sub-domains



Single Domain

AnaEE
DiSSCO
ELIXIR
EMPHASIS
INTERACT

EPOS

EUROFLEETS
EURO-ARGO
JERICO-RI
SEADATANET

ACTRIS
ARISE
EISCAT_3D
EUFAR
EUROCHAMP 2020
HEMERA
IAGOS

Multi Domain

EuroGOOS
ICOS
IS-ENES
SIOS

DANUBIUS
eLTER
EMBRC
LifeWatch
AQUACOSM

EMSO

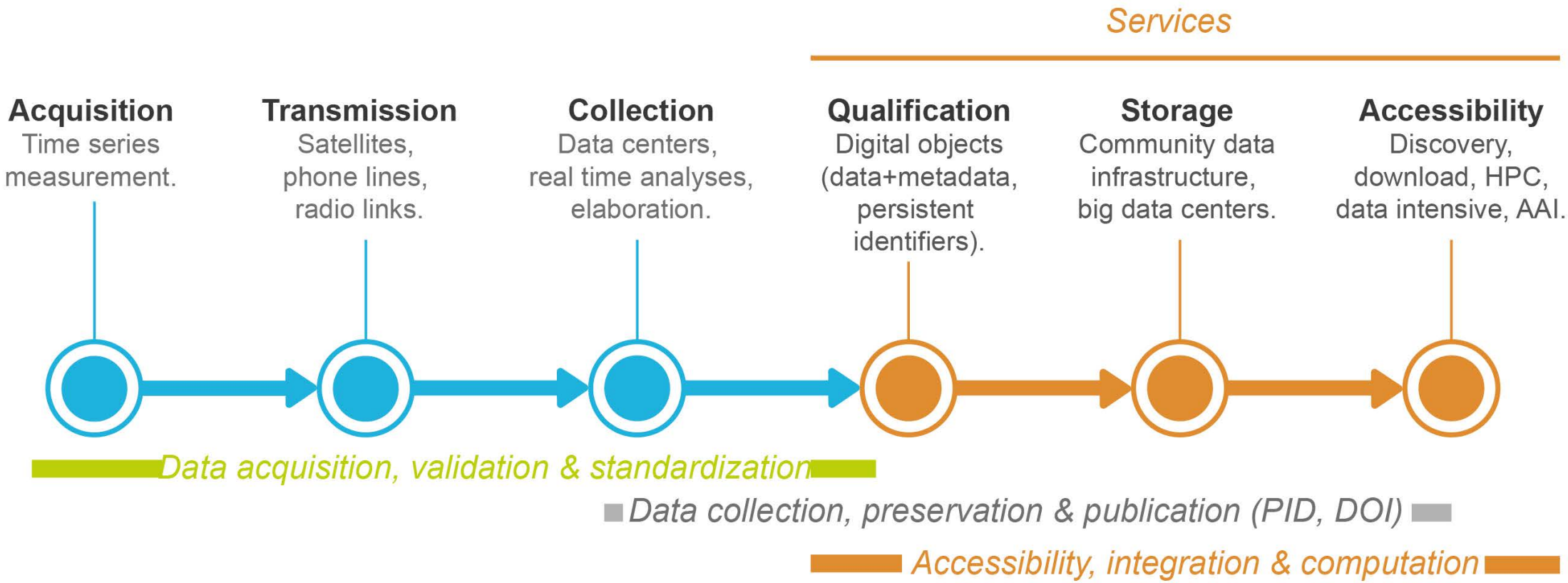
Environmental Research Infrastructures

- provide data and research products from all four sub-domains of the Earth system;
- data are crucial European contributions to global monitoring of the state of the Earth system and climate;
- data are vital for assessing past and defining future policies, as well as for the development of environment-friendly innovations.

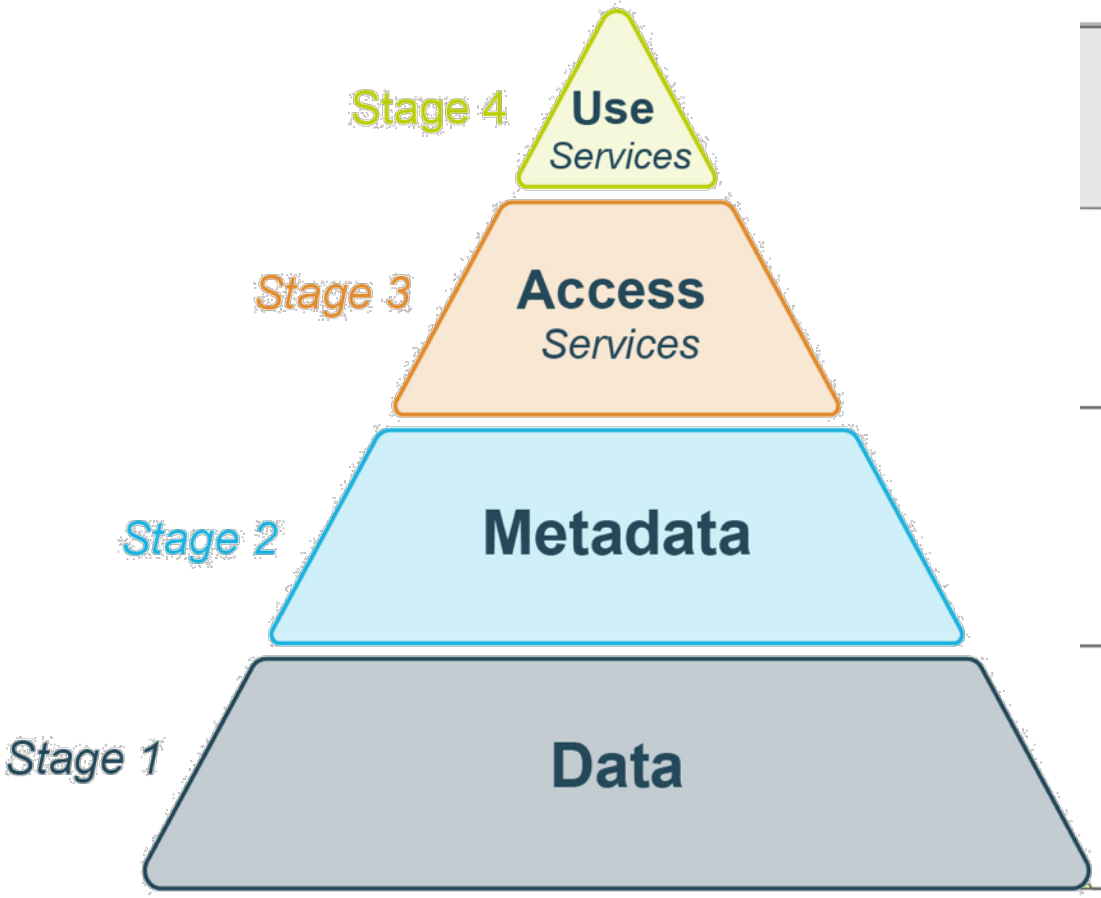
ENVRI-FAIR (14 RIs on ESFRI Roadmap)

- develops FAIR-based tools and resources for easy and seamless access to environmental data and services provided by ENVRIs.

FAIR Data and Services

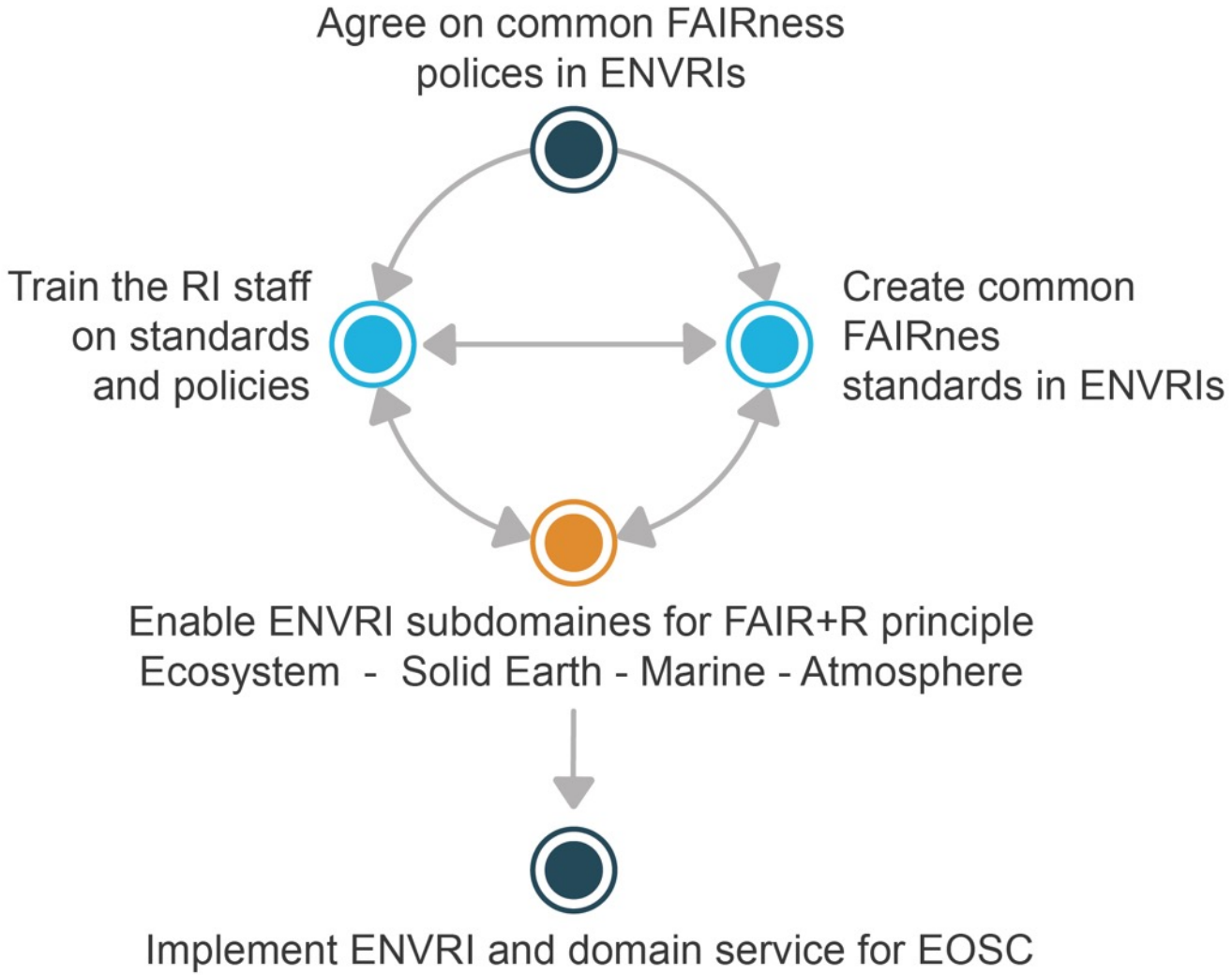


Maturity Stages to FAIR Data and Services



FAIR PRINCIPLES
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. **F4.** (meta)data are registered or indexed in a searchable resource.
F1. Metadata are assigned a globally unique and eternally persistent identifier. **F2.** data are described with rich metadata. **F3.** metadata specify the data identifier. **F4.** metadata are registered or indexed in a searchable resource. **A2.** metadata are accessible, even when the data are no longer available. **I1.** metadata use a formal, accessible, shared, and broadly applicable. **I2.** metadata use vocabularies that follow FAIR principles. **I3.** metadata include qualified references to other metadata. **R1 (R1.1 - R1.2 - R1.3)** Metadata are richly described with a plurality of accurate attributes.
F1. Data are assigned a globally unique and eternally persistent identifier. **F4.** Data are registered or indexed in a searchable resource. **I1.** Data use a formal, accessible, shared, and broadly applicable language for knowledge representation. **I2.** Data use vocabularies that follow FAIR principles. **I3.** Data include qualified references to other (meta)data. **R1.1.** Data are released with a clear and accessible data usage license. **R1.3.** Data meet domain-relevant community standards.

ENVRI-FAIR Workflow



Implementation Strategy

Implementation by the research infrastructures

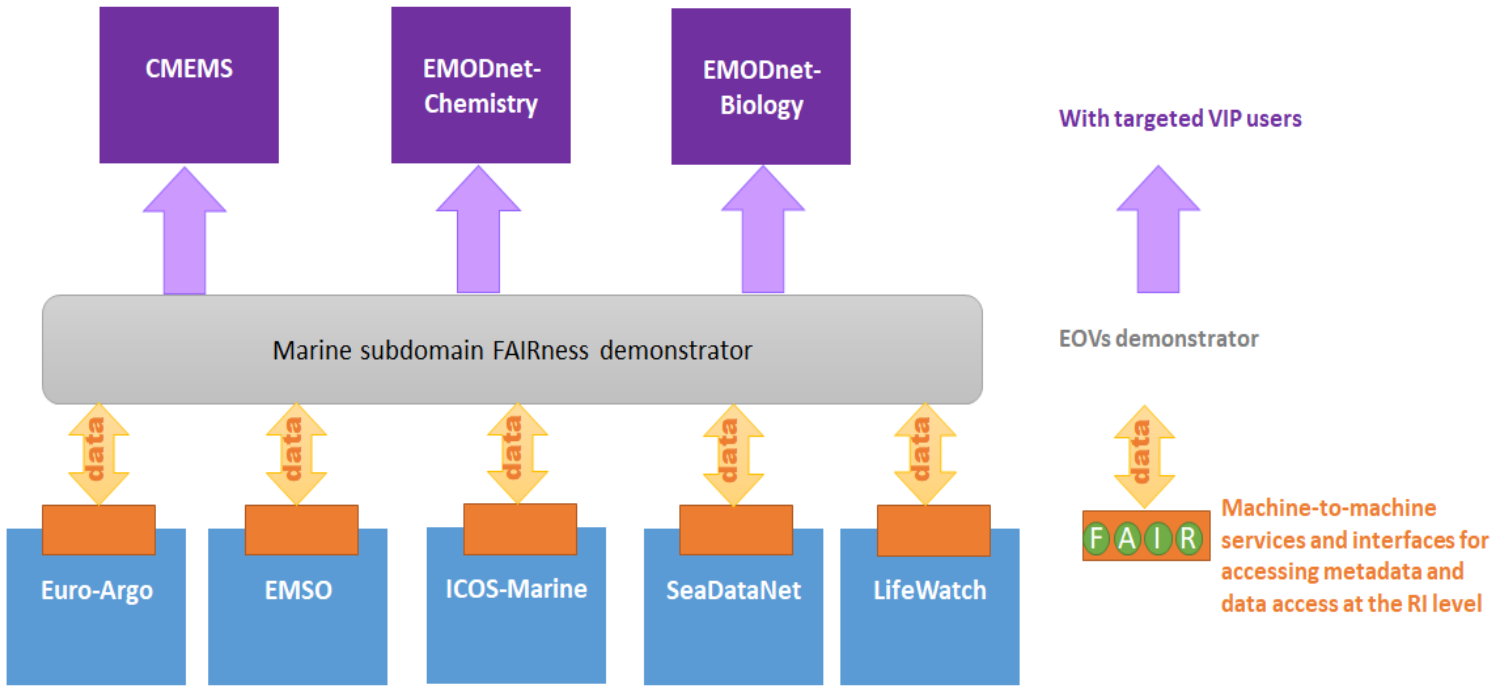
Common policies are agreed, and are the ”border control” of what is considered FAIR

Standards and implementation are hierarchical

- Cluster level
- Subdomain level
- RI level

Common Strategy for implementation in Ocean domain

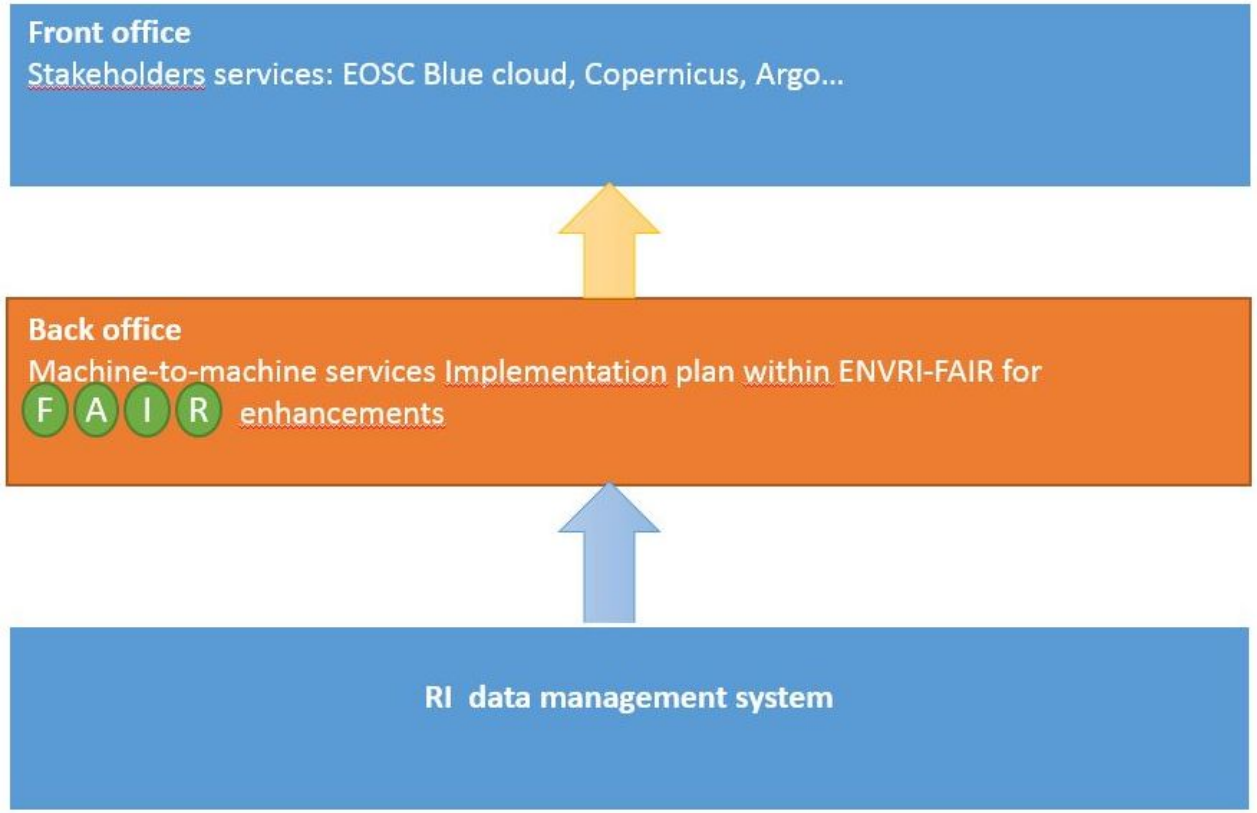
Filling the gaps



- Define priorities in filling gaps
  - Known RI User requirements
  - Marine domain demonstration requirements
- Propose an approach to solve the gaps: possible technologies, planning, involved partners

Common Strategy for implementation in Ocean domain

Work at the interfaces

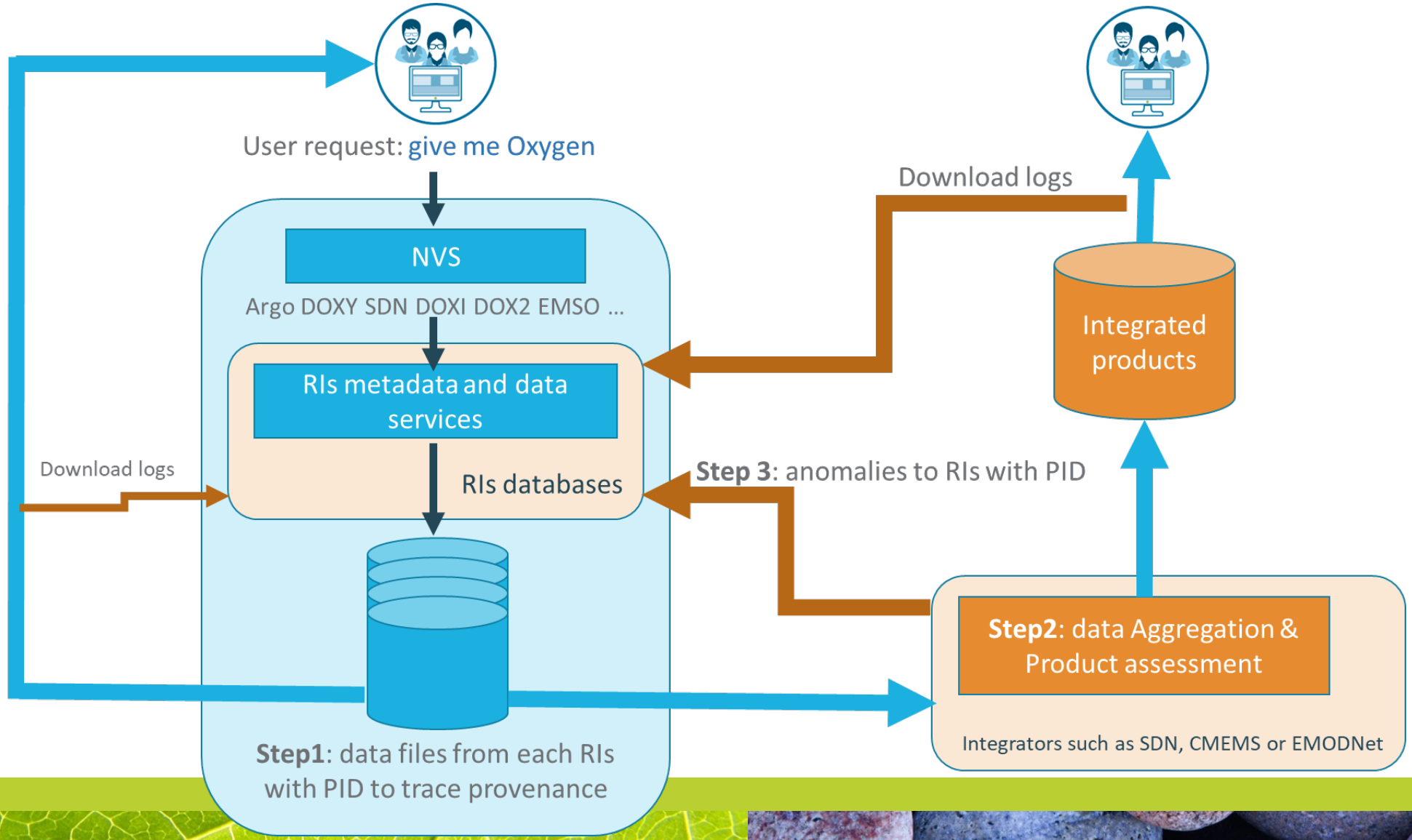


Including cross-domain services within ENVRI-FAIR project

In the scope of ENVRI-FAIR project

Out of the scope of ENVRI-FAIR project

- ENVRI-FAIR activities are at the level of back-office services and relies on RI data management system
- Enhancement of FAIRness of the service will impact Front-end services that can be managed
  - at the level of the RI
  - by end users such as Copernicus
  - EOSC is one of the users of the enhanced services


User request: give me Oxygen
NVS
Argo DOXY SDN DOXI DOX2 EMSO …
RIs metadata and data services
RIs databases
Download logs
Step1: data files from each RIs with PID to trace provenance
Download logs
Integrated products
Step 3: anomalies to RIs with PID
Step2: data Aggregation & Product assessment
Integrators such as SDN, CMEMS or EMODNet

The Clusters View on EOSC

European **O**pen **S**cience **C**loud =

Enable researchers to access data, storage and compute ("cloud") via an Europe wide federation of IT services ("e-Infrastructure")

E-Infrastructure consolidation

\+

Drive the transition to Open Science (Open Data, Open Standards, Open Literature) - bring research benefits to European societies at large

Open Science

\+

Populate EOSC with the scientific data resources and computational tools from research infrastructures – drive usage by to Europe's 1.7 M researchers

Scientific **C**ommunities' content and users

Contributes to Open Science by

- FAIRness assement methodologies
- FAIR policies
- FAIRness training

Contributes to Scientific content by

- Designing the ENVRI catalogue of services
- Designing ENVRI-HUB architecture

ENVRI

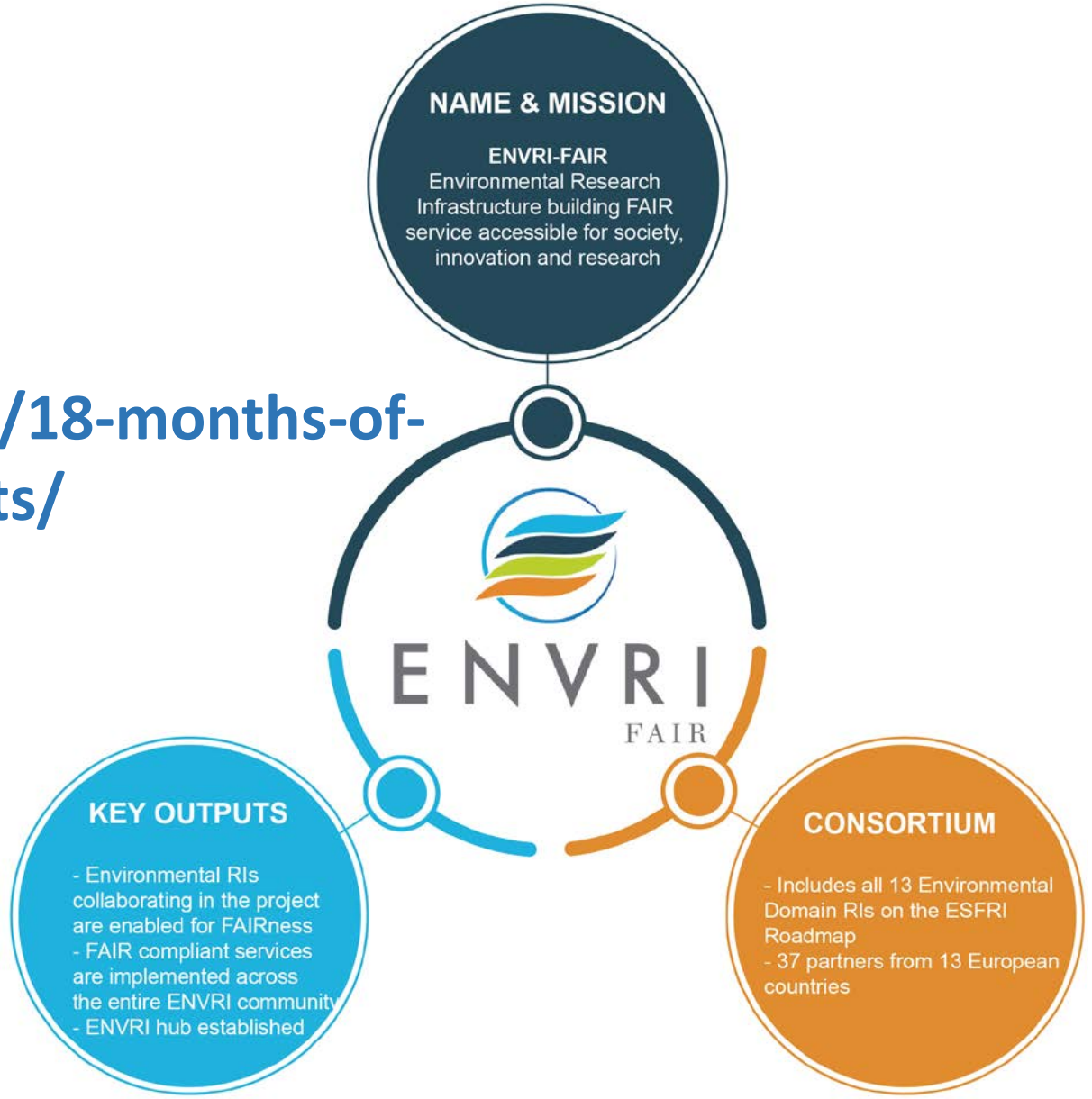
ENVRI-FAIR Key Messages

- Making data **FAIR** requires resources and expertise which are provided by the **ENVRI scientific communities.**
- Ensuring **coherence of methodologies and technologies** important for FAIRification process across subdomains.
- Putting highest priority on the **provision of** relevant high-quality **open data** using open licenses and open links is an ENVRI-FAIR priority.

ENVRI-FAIR is the connectio
off the Environmental Research Infrastructure community to the European Open Science Cloud

https://envri.eu/18-months-of-accomplishments/

Questions ?



ENVRI-FAIR brings together all Environmental Domain RIs from the ESFRI Roadmap -

but collaborates with the entire community of Environmental Research Infrastructures

ENVRI-FAIR has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824068

Find us

- www.envri-fair.eu
- @ENVRIcomm
- ENVRI community
- ENVRIcomm

annex

ENVRI-FAIR has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824068