

27-29 May 2024 



imdis

International conference on Marine Data and Information Systems

D4Science: Advancing Ocean Science Through Collaborative Data Analysis

Massimiliano Assante, PhD – CNR, Italy



MARIS



National
Oceanography
Centre



iImagine



eosc
Blue-Cloud2026

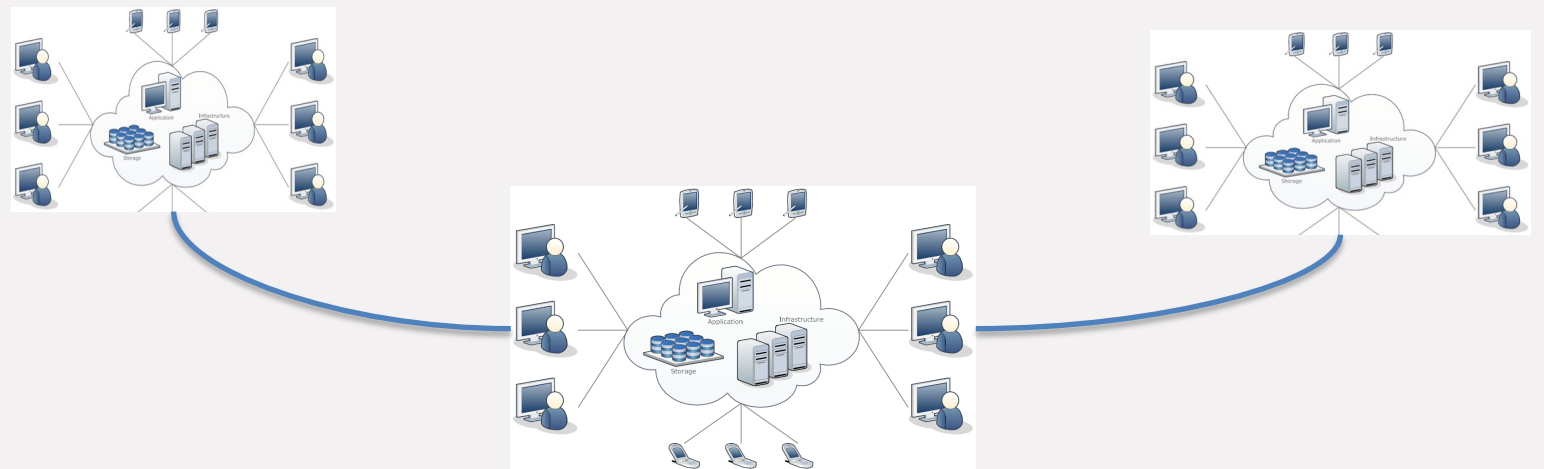
D4Science: Advancing Ocean Science Through Collaborative Data Analysis

M. Assante, L. Candela, L. Frosini, F. Mangiacrapa, E. Molinaro, P. Pagano

CNR - Istituto Scienza e Tecnologia dell'Informazione "A. Faedo", Pisa, Italy

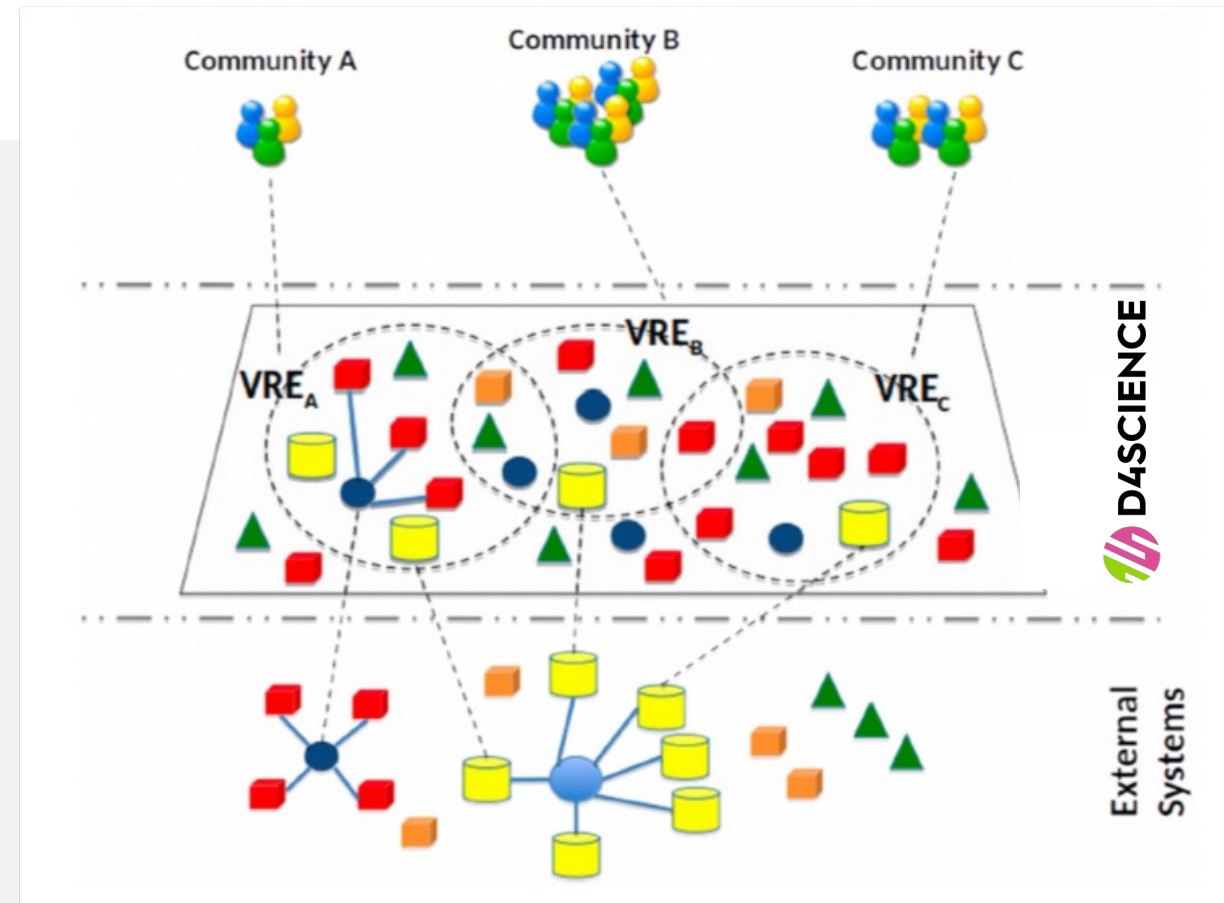
What is D4Science

- A digital **infrastructure** for **collaborative scientific research**, combining over 500 software components into a coherent and centrally managed system of **hardware, software, and data resources**



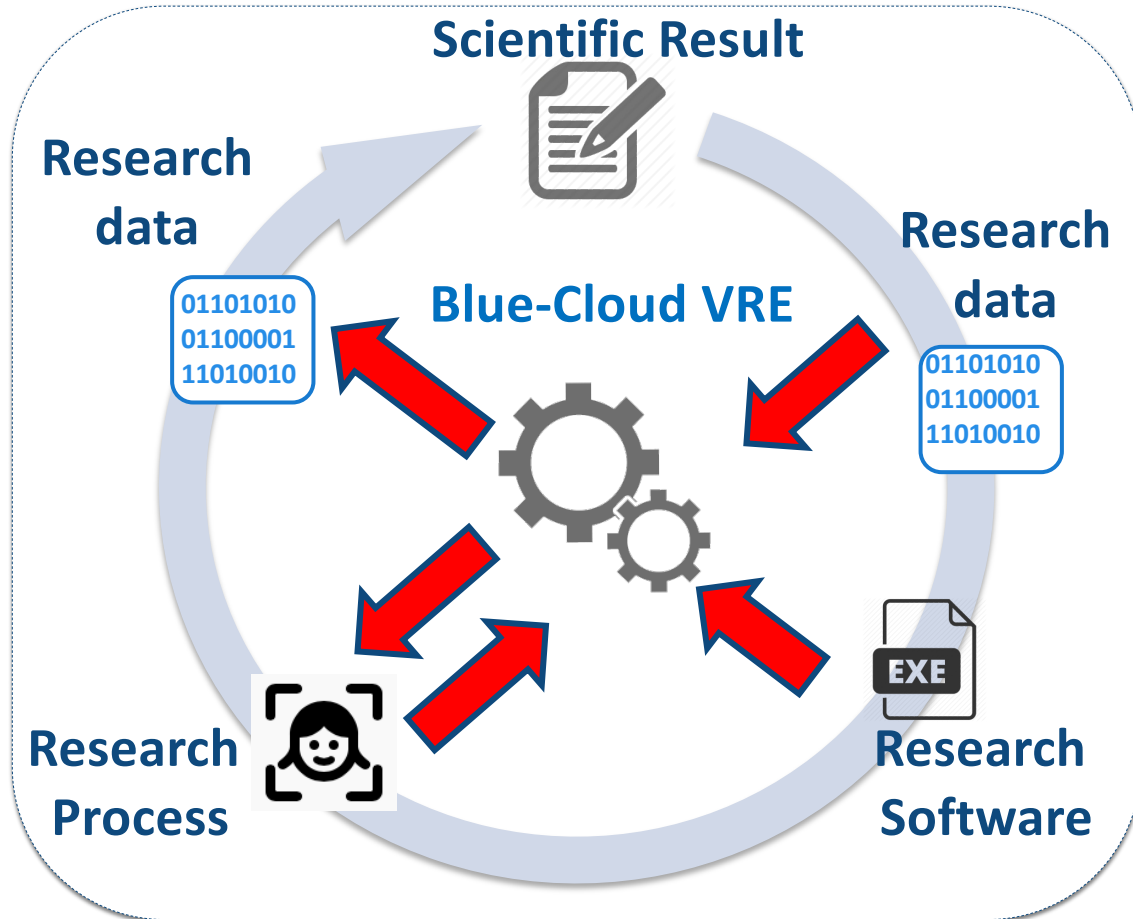
 **D4SCIENCE** promotes Open Science practices through the operation of a Data Infrastructure service

- leverage **external systems** (e.g. data, storage, cloud-based infrastructures)
- by exposing them as a **common unified space of resources**
- to serve **diverse community of researchers**
- via the **provision of tailored services and sharing tools**
- made accessible through a flexible, web-based and on-demand environments called **Virtual Research Environments (VRE)**



M. Assante et al. (2023) **Virtual research environments co-creation: The D4Science experience.** Concurrency Computat Pract Exper. 2023; 35(18):e6925. doi:[10.1002/cpe.6925](https://doi.org/10.1002/cpe.6925)

M. Assante, et al. (2019) **Enacting open science by D4Science.** Future Generation Computer Systems (Vol. 101) doi: [10.1016/j.future.2019.05.063](https://doi.org/10.1016/j.future.2019.05.063)



Enable

- Repeat, Reproduce, Reuse, Evaluate
- Active collaboration
- Effective sharing
- Provenance and attribution

Adopt

- As-a-service approach
- Standards
- Economy-of-scale to reduce operational costs



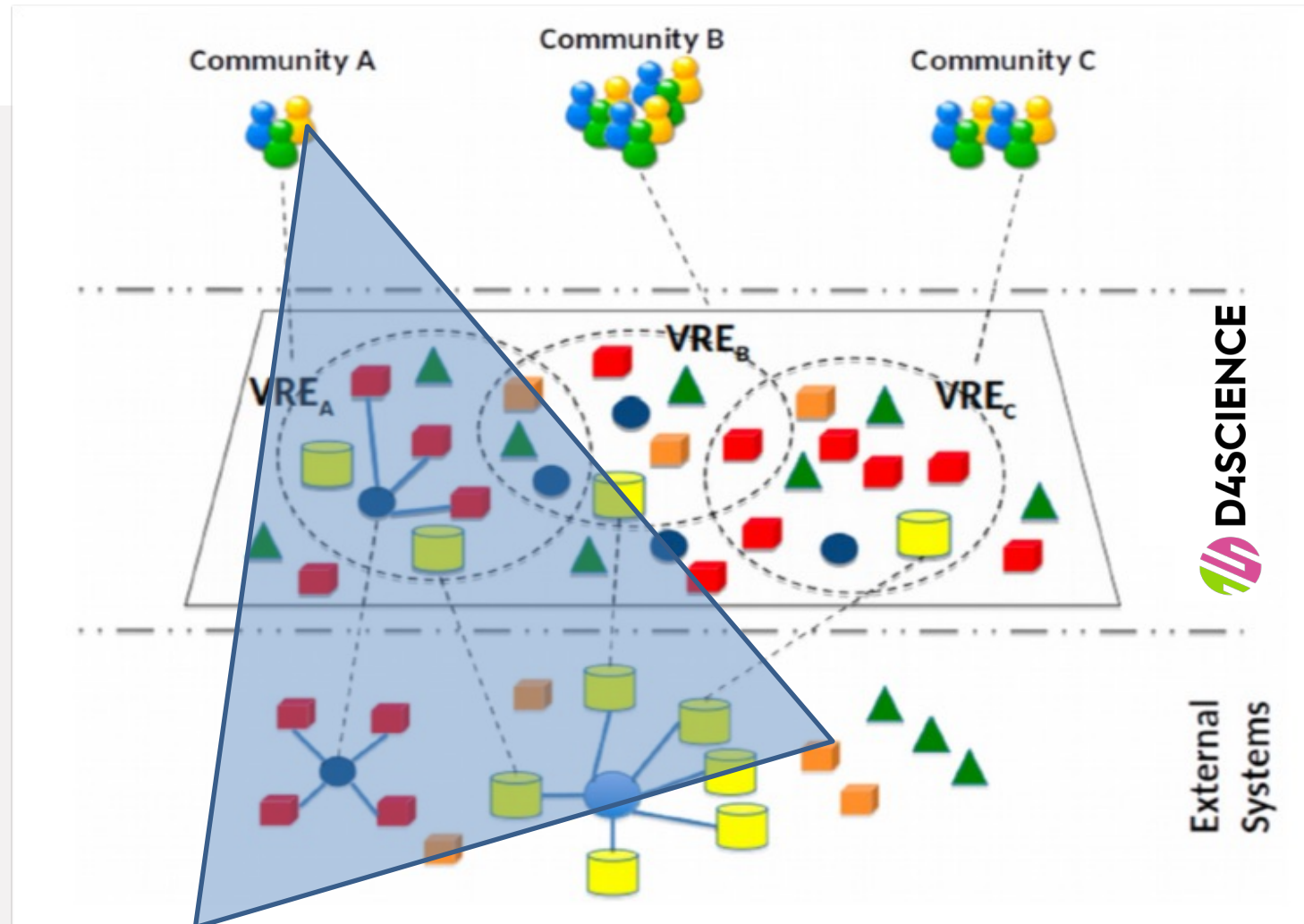
D4Science is a federated digital infrastructure promoting Open Science

- 4 sites (1 Pisa, 3 @ GARR)
- 3590 CPUs core
- 13 TB RAM
- 600 TB Storage

D4Science is owned and managed by CNR

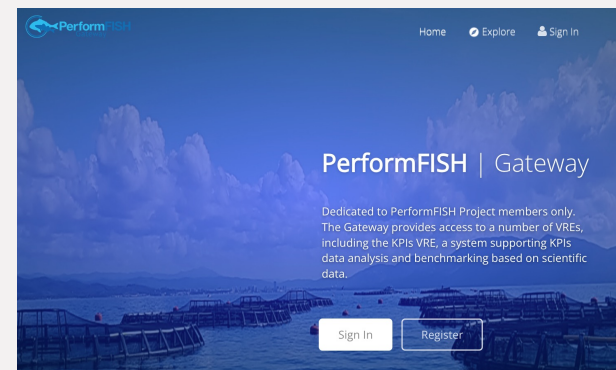
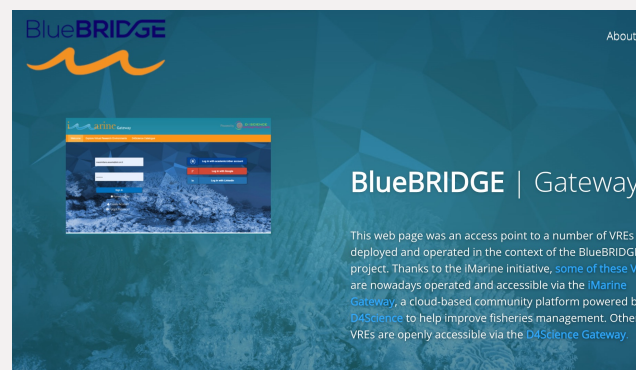
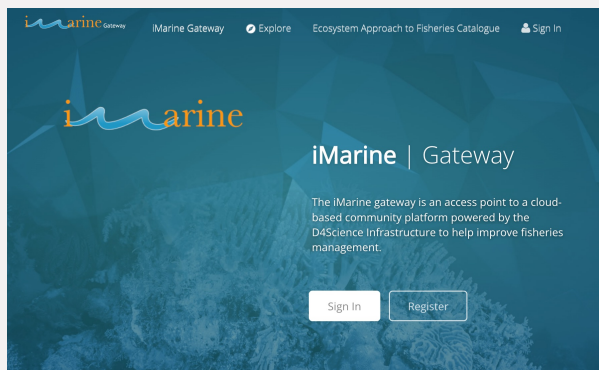
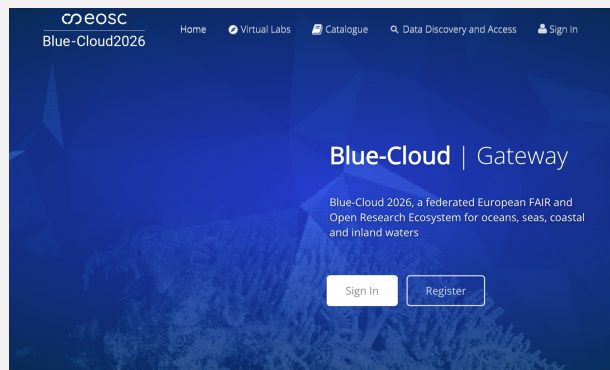
It supports ESFRI RIs, national and European projects, national and international initiatives

Community VREs on D4Science



D4Science in Marine and Ocean Science

- Actively supports diverse marine and ocean science projects and initiatives.
- Contributions to European Open Science Cloud (EOSC) through projects such as **Blue-Cloud** and **EOSC Blue-Cloud2026**.
- Supported H2020 **BlueBRIDGE** and **iMarine** EU FP7 Project.



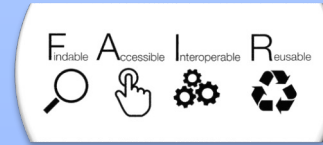
One common place



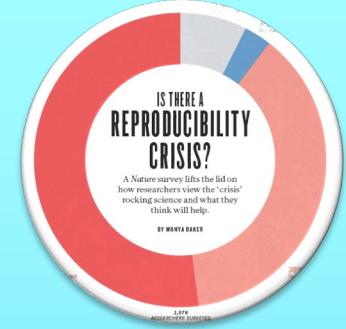
Collaboration



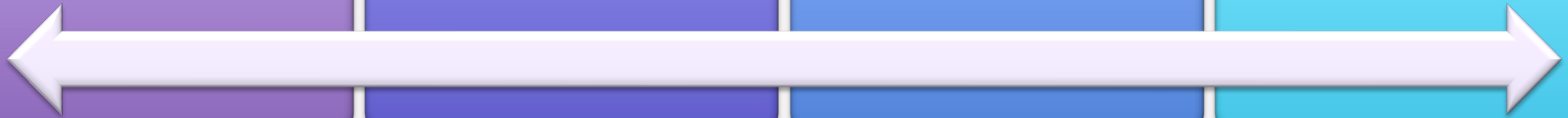
Sharing



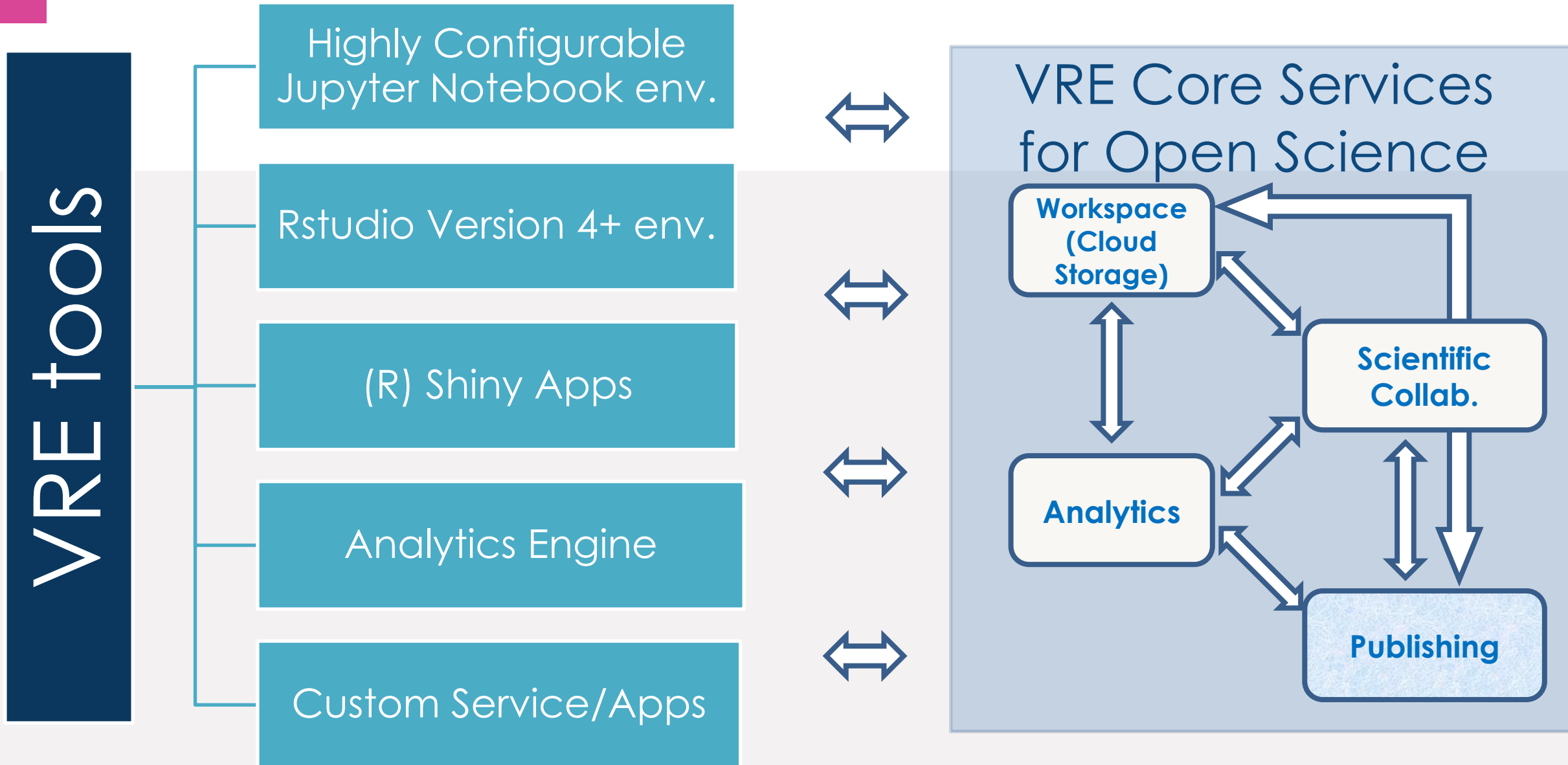
Reuse



Reproducibility



Collaborative Data Analysis



Publishing: an item example

Organization

Title & Description

PURL

QR-Code

Tags

License

Resources (item payloads)

Metadata

FAO aquatic species distribution map of *Chlamydoselachus anguineus*

The main sources of information for the species distribution are the habitat description and geographic range contained in the published FAO Catalogues of Species (more details at <http://www.fao.org/fishery/fishfinder>). Terms used in the descriptive context of the FAO Catalogues were converted in standard depth, geographic and ecological regions and inserted into a Geographic Information System.

Tags

105831 11367 Chlamydoselachidae Chlamydoselachus an... FAO FIGIS

Filled shark HEXANCHIFORMES HXC Species distribution

aquatic species dis... fao-species-map-hxc fisheries fishery

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

Data and Resources

- SPECIES_DIST_HXC**
FAO aquatic species distribution map of *Chlamydoselachus anguineus*
Go to resource
- SPECIES_DIST_HXC**
GIS data download (WFS - GML)
Go to resource
- SPECIES_DIST_HXC**
GIS data download (WFS - ESRI Shapefile)
Go to resource
- XML**
metadata (XML)
Go to resource
- Unnamed resource**
Aquatic Species Distribution Maps
Go to resource
- Unnamed resource**
Factsheet - Summary description
Go to resource
- Unnamed resource**
Aquatic Species Distribution Maps (GIS Viewer)
Go to resource
- FAO - Fisheries and Aquaculture Department (FI)**
Go to resource
- FAO - Fisheries and Aquaculture Department (FI)**
Go to resource
- FAO - Fisheries and Aquaculture Department (FI)**
Go to resource

Item URL

<http://data.d4science.org/ctlg/d4science.research-infrastructures.eu/sarda-sarda3>

Geonetwork Links

Field	Value
GN_Metadata_Show	http://geonetwork.d4science.org/geonetwork/srv/en/metadata.show?uuid=b9bd0ef9-76f9-4fb9-8f98-1c36557bc5f8
GN_Metadata_Source	http://geonetwork.d4science.org/geonetwork/srv/en/xml.metadata.get?uuid=b9bd0ef9-76f9-4fb9-8f98-1c36557bc5f8
GN_URL	http://geonetwork.d4science.org/geonetwork

Additional Info

Field	Value				
access_constraints					
bbox-east-long	180.0				
bbox-north-lat	90.0				
bbox-south-lat	-90.0				
bbox-west-long	-180.0				
contact-email	info@i-marine.eu				
coupled-resource					
dataset-reference-date	<table border="1"> <thead> <tr> <th>type</th> <th>creation</th> </tr> </thead> <tbody> <tr> <td>value</td> <td>2013-04-12T01:13:51.731+02:00</td> </tr> </tbody> </table>	type	creation	value	2013-04-12T01:13:51.731+02:00
type	creation				
value	2013-04-12T01:13:51.731+02:00				
frequency-of-update	asNeeded				
graphic-preview-file					

License

Creative Commons Attribution-ShareAlike 4.0 [Open Data](#)

VLabs at-a-glance

AquacultureAtlasGeneration Administration Aquaculture Atlas Production System Members Analytics Catalogue Viewer

Use cases

greece

Malta

South_sulawesi

Coastal Ponds / Rice Paddy by Sentinel-1/2 (2014-2017)

Coastal Ponds by Landsat 5 (1994-1997)

SAR

PlanktonGenomics Administration Result Overview JupyterHub RStudio Analytics Engine Catalogue Members

File Edit View Run Kernel Tabs Settings Help

readme_Notebook1.2.ipynb

Filter files by name

Name

readme_Notebook1.2.ipynb

Demonstrator 2 - Notebook 1.2. - Creating protein functional clusters for Notebook 2

Pavla Debeljak, Eric Pelletier & Lucie Bittner

Corresponding author/maintainer: pavla.debeljak@mnhn.fr

Introduction

Recent metagenomic studies have revealed that marine plankton is far more diverse than previously thought (Carradec et al. 2018, Duarte et al. 2020), with hundreds of thousands of genetically distinct taxa and more than 150 million genes documented, however more than half of the planktonic 'omic' sequences have still unknown taxonomy and/or function, especially in terms of sequences with eukaryotic origin. These unprecedented amounts of data on planktonic communities call for the need of innovative data-driven methodologies to quantify and observe their biogeographic importance.

Notebook 1.2. allows for the creation of protein functional clusters from Fasta files derived from Metagenomic or Metatranscriptomic sequencing. These

FisheriesAtlas Administration Members Analytics GeoNetwork Catalogue GeoServer Thredds NetCDF server Catalogue Viewer

About Find Access

Legend

- FISHERIES ATLAS MAPS
- BASE OVERLAYS
- Exclusive Economic Zones

Dataset information

Blue-Cloud Lab

Development and integration environment for R, Python, and other supported software languages

- It is powered by a cluster of Dell servers, each with 16 cores and 32 GB RAM.
- It is powered by JupyterLab with a maximum of 32 cores and 32 GB RAM on request. JupyterLab is provided by Dell with the support of CERN.

All the environments are not for large-scale processing. They are conceived to support the integration, testing, and validation of your script in the Dell's distributed computing facility. For large-scale processing, please request the creation of a dedicated Virtual Lab.

ESA CCI Ocean Colour Product

Data products generated by the Ocean Colour component of the Euro Agency Climate Change Initiative project. These files are monthly composites from MERIS, MODIS, SeaWiFS, and OLCI products. The OLCI products were derived using the standard SeaDAS algorithm.

GlobalFisheriesAtlas Administration Members Analytics Spatial Data Services Catalogue Viewer

Tuna Atlas: Interactive Indicator

Example of displayed data

Map

Plot

Datatable

Show 10 entries

Search:

SQL Queries

```
SELECT geom_id, geom, species, gear_type, fishing_fleet, SUM(measurement_value) as measurer_measurement_unit, fishing_mode, ST_asText(geom) AS geom_wkt_year FROM public.chinycatch'
```

Marine Environmental Indicators Home Administration Software Importer Analytics Engine JupyterHub GeoNetwork Catalogue MEI Generator Members

Generate new data

My data

Account: Leonardo Candela

Data source: MEDSEA REANALYSIS PHYS

Type: monthly mean timeseries

Environmental field: temperature

Start time: 01/1987

End time: 12/1987

Area: Lat: 34, 42; Lon: -4,99, 1

Depth [m]: From: 0.5; To: 3000

Execute process

27-29 May 2024 



imdis

International conference on **Marine Data**
and **Information Systems**

Thanks for your attention



MARIS

