eosc | Blue-Cloud2026

VRE Info day: Blue-Cloud VRE Open Science services for building, hosting and operating Virtual Labs

Massimiliano Assante, PhD

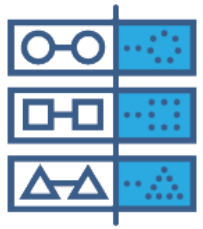
ISTI - CNR

https://orcid.org/0000-0002-3761-1492

Follow @maxassante

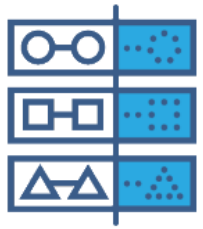
Funded by the European Union

Key products and services



- **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
- **Blue-Cloud Virtual Research Environment platform** 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
- **Blue-Cloud Virtual Labs**, configured with specific analytical workflows to serve as **Demonstrators,** which can be adopted and adapted for other inputs and analyses

Key products and services

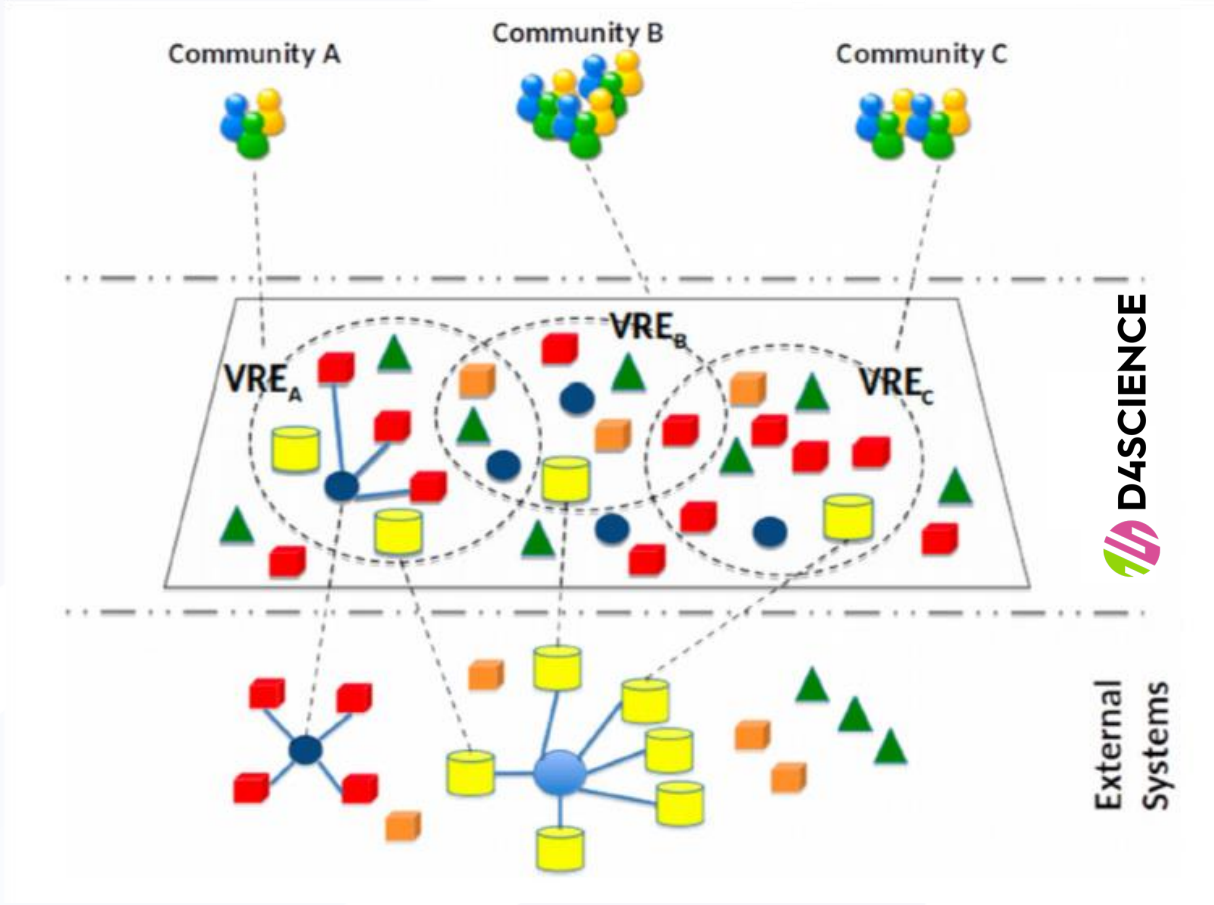


- **Blue-Cloud Data Discovery & Access service**, federating key European data management infrastructures, to facilitate users in finding and retrieving multi-disciplinary datasets from multiple repositories
- **Blue-Cloud Virtual Research Environment platform** to provide a range of services and to facilitate orchestration of computing and analytical services for constructing, hosting and operating Virtual Labs for specific applications
- **Blue-Cloud Virtual Labs**, configured with specific analytical workflows to serve as **Demonstrators,** which can be adopted and adapted for other inputs and analyses

Blue-Cloud VRE platform based on D4Science Infrastructure

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

- leverage **external systems** (e.g. data, storage services, computational resources, cloud computing 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 Labs**



Assante, M., Candela, L., Castelli, D., Cirillo, R., Coro, G., Frosini, L., Lelii, L., Mangiacrapa,, Pagano, P., Panichi, C., Sinibaldi, F. **Enacting open science by D4Science**. Future Generation Computer Systems (Vol. 101) https://doi.org/10.1016/j.future.2019.05.063

D4Science as a System of Systems

System of Systems

D4Science is built with dedicated services leveraging on existing e-infrastructures and other domain-specific infrastructures, EOSC resources and services

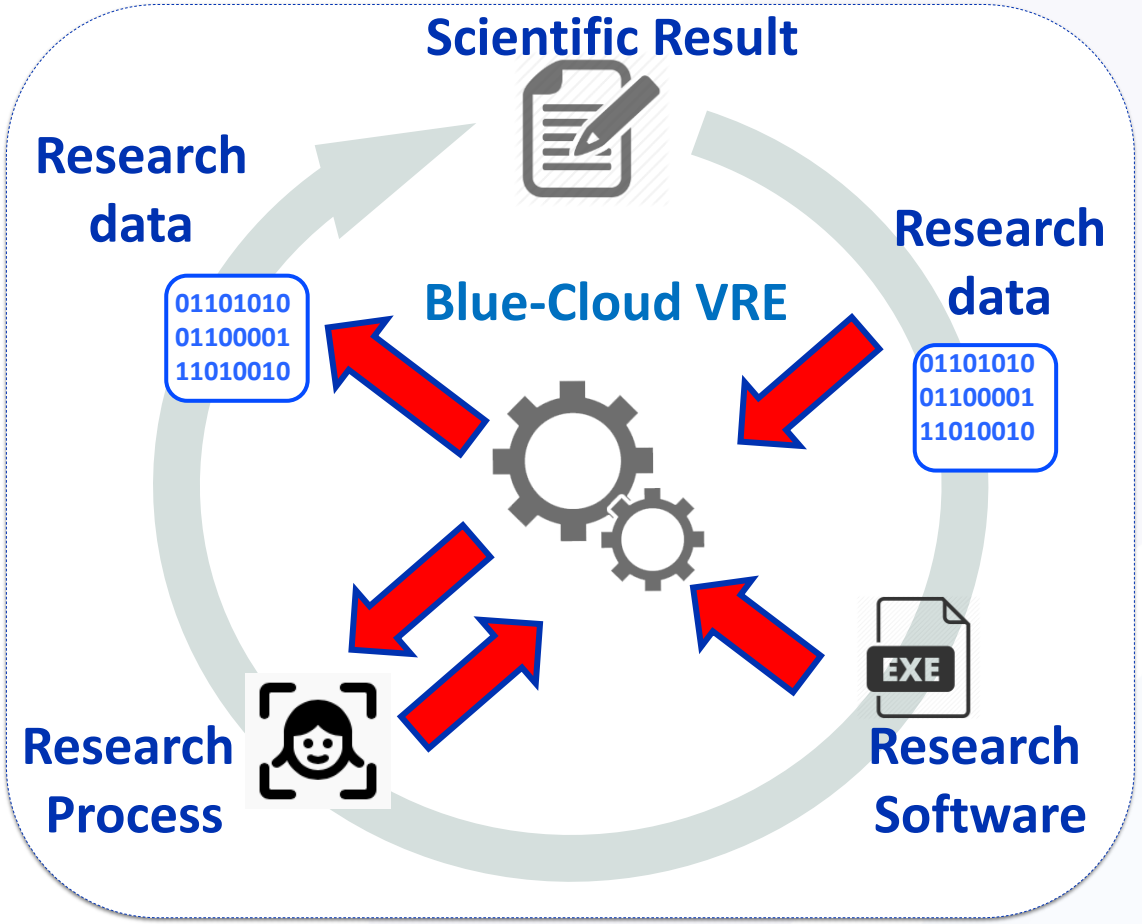
Extensible

integrates services and resources resulting from existing initiatives

Open

promotes OpenSscience and OS practices

A SoS to support and promote Open Science practices for Data Driven Science



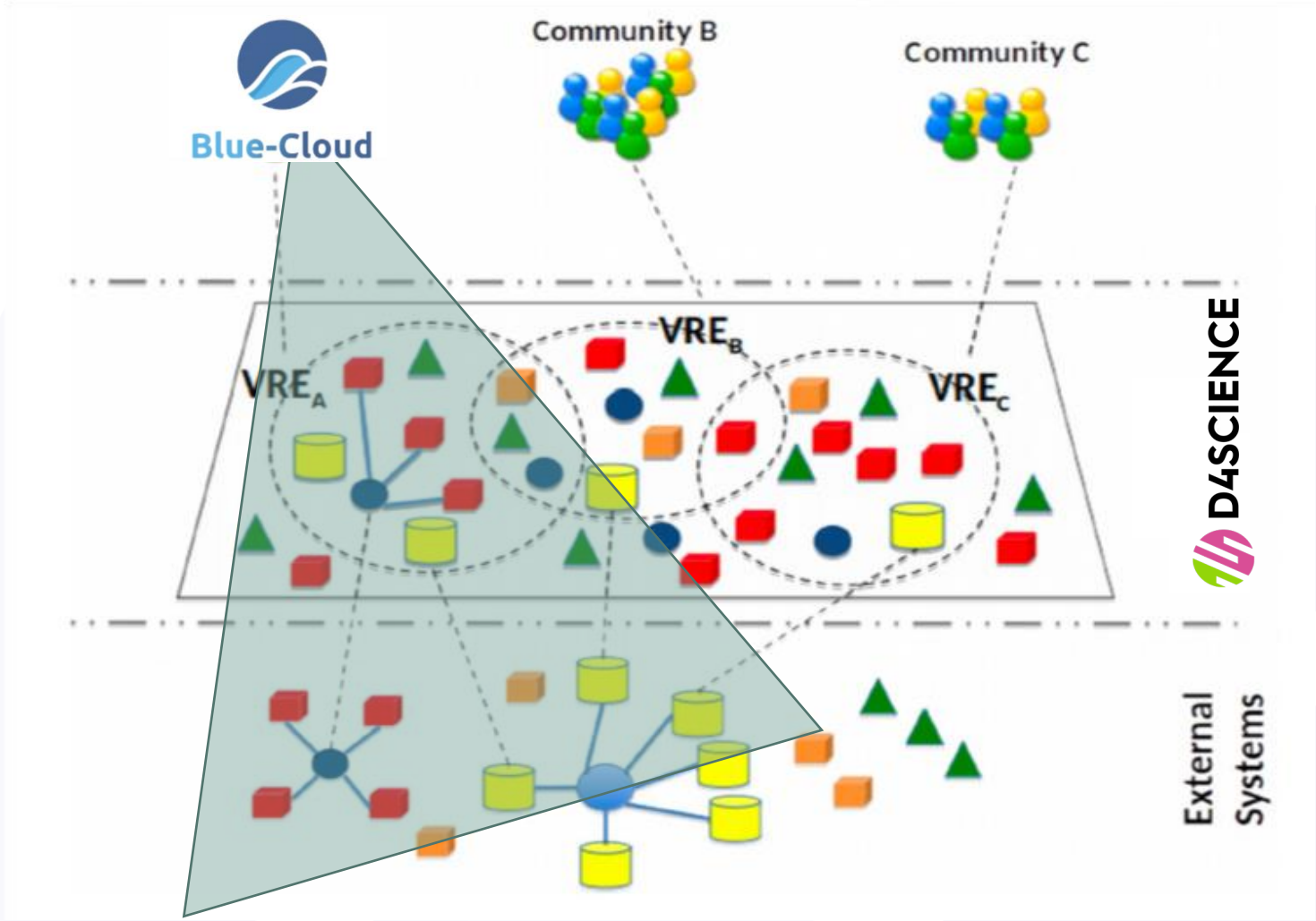
Enable

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

Adopt

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

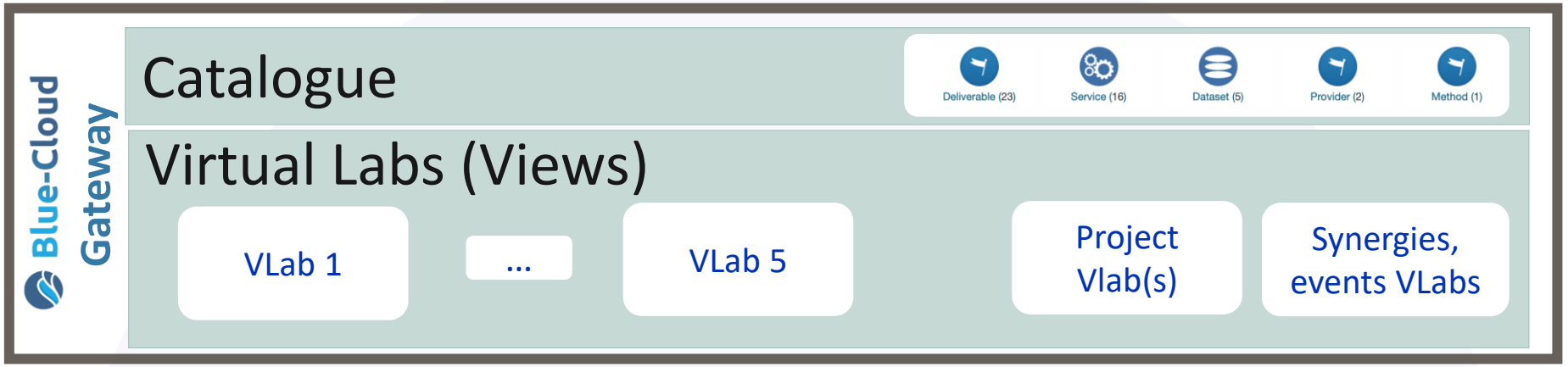
Blue-Cloud VRE on D4Science

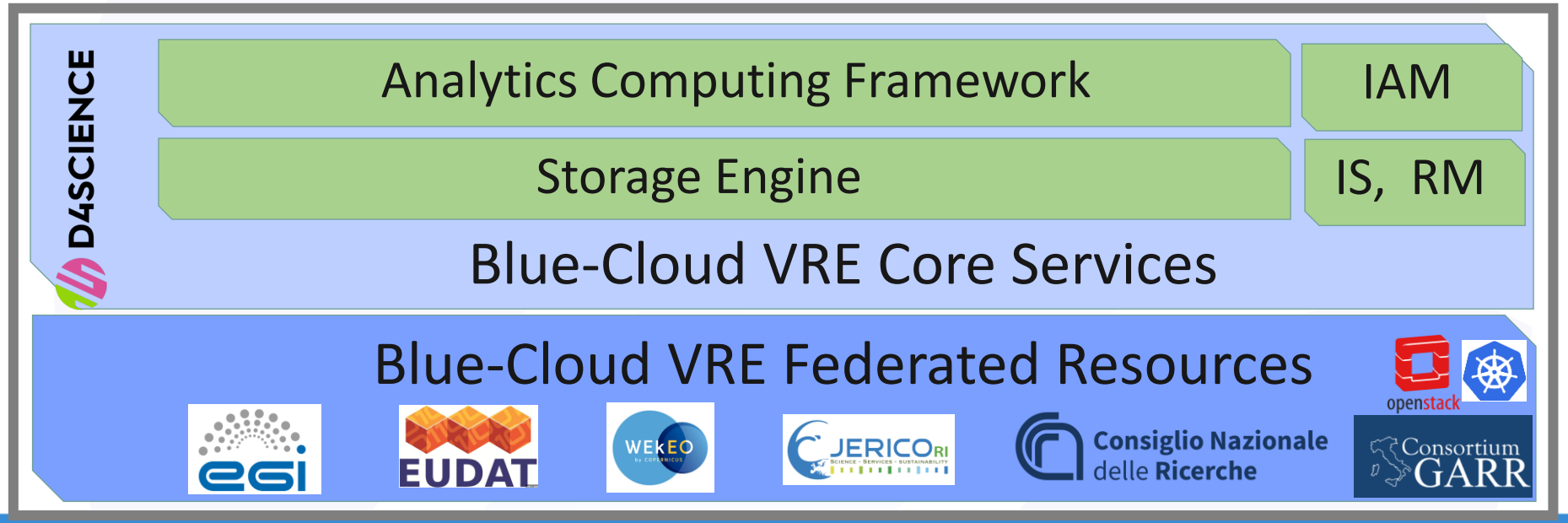


Assante, M., Candela, L., Castelli, D., Cirillo, R., Coro, G., Frosini, L., Lelii, L., Mangiacrapa,, Pagano, P., Panichi, C., Sinibaldi, F.
Enacting open science by D4Science. Future Generation Computer Systems (Vol. 101) https://doi.org/10.1016/j.future.2019.05.063

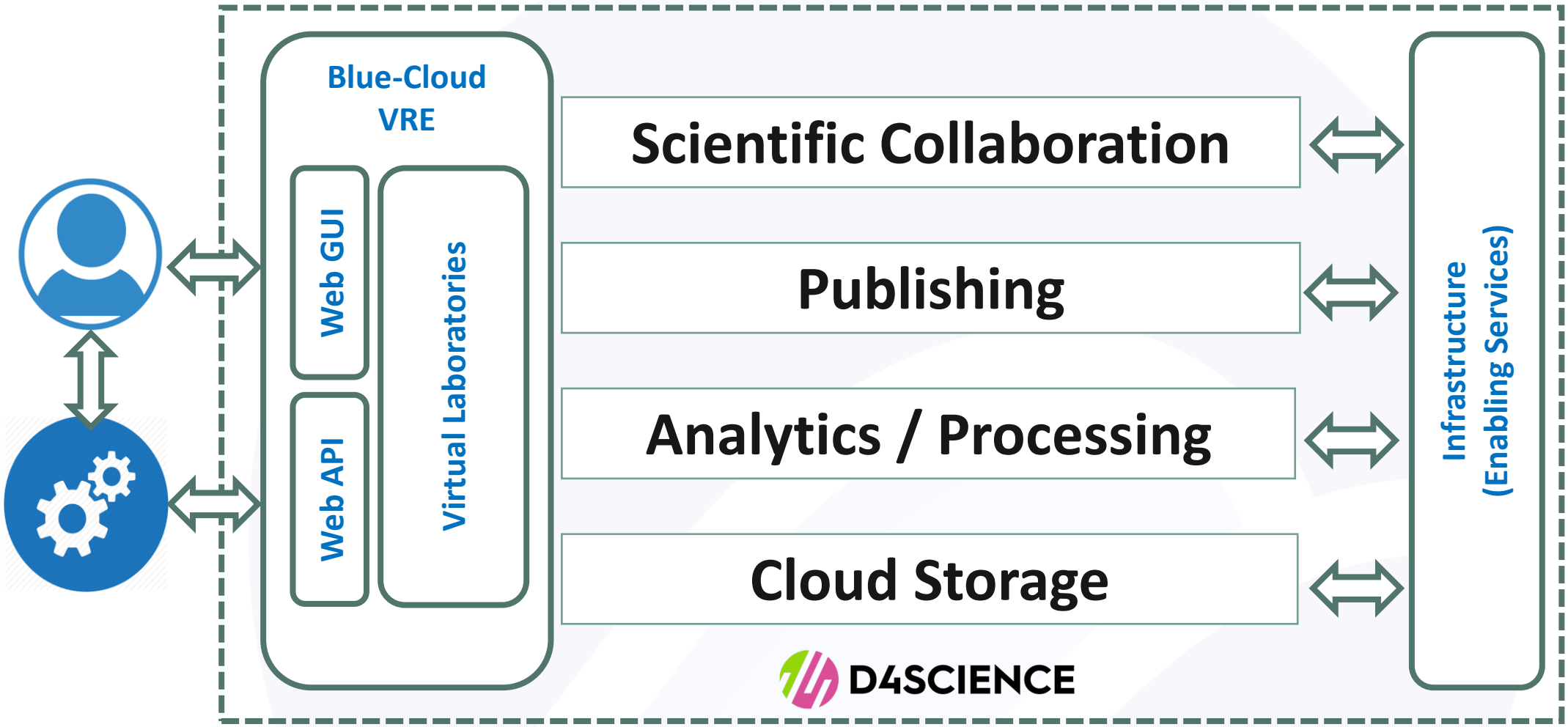
Blue-Cloud VRE – Logical View Today

https://blue-cloud.d4science.org





Blue-Cloud VRE Open Science Services for VLab tools

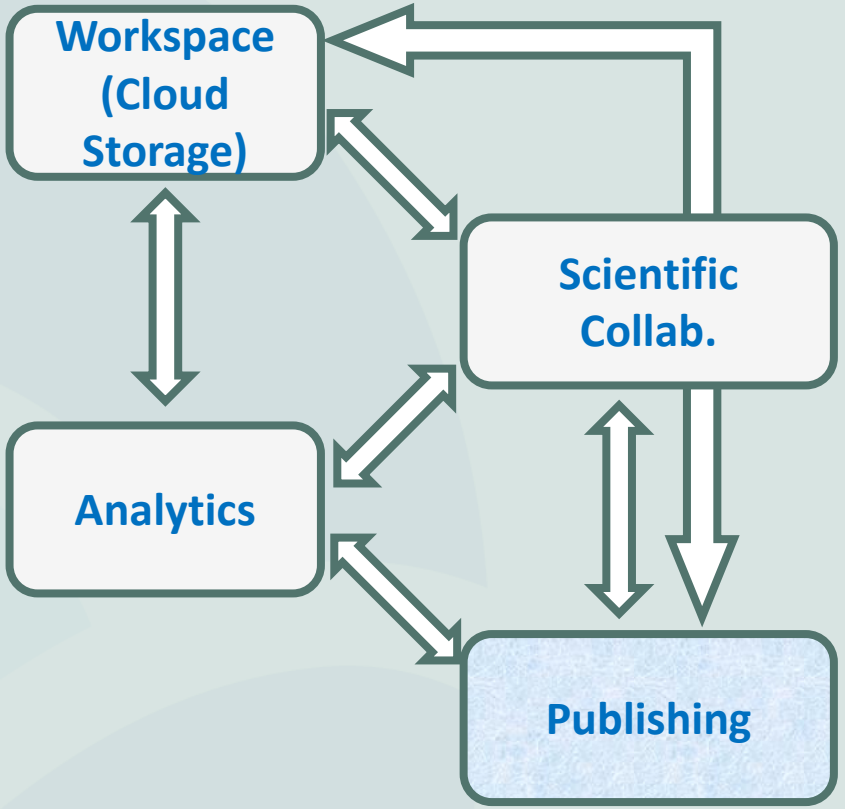


VLab tools

VLab tools

- Highly Configurable Jupyter Notebook env.
- Rstudio Version 4+ env.
- (R) Shiny Apps
- Analytics Engine
- Custom Service/Apps

VRE Core Services for Open Science

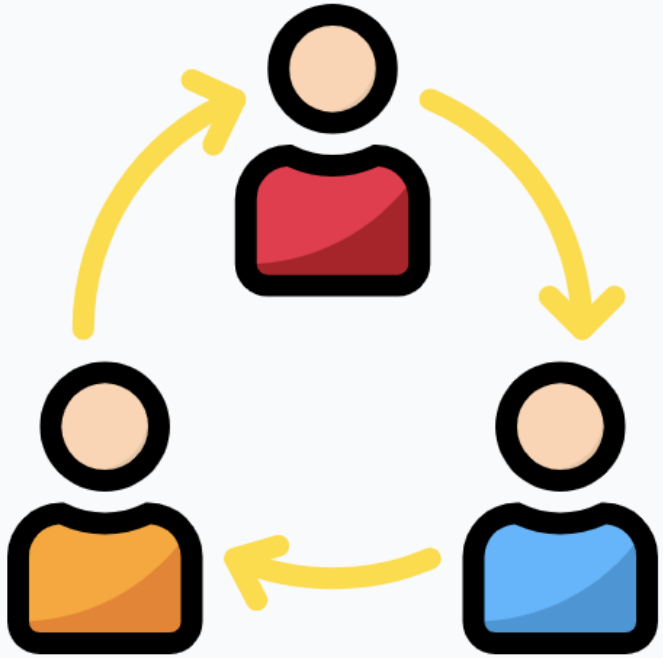


Scientific collaborations can be defined as interactions taking place within a social context among two or more scientists that ***facilitates the sharing of meaning and completion of tasks*** with respect to a mutually shared, superordinate goal.

Blue-Cloud2026 promotes technologies enabling the **sharing of datasets and software methods/algorithms**

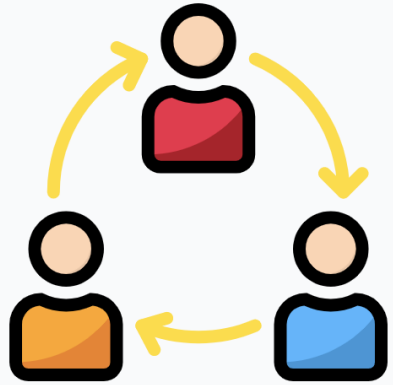
Sharing Patterns

Selective Sharing



Controlled Sharing

Publishing



Blue-Cloud promotes technologies enabling the

sharing of datasets and methods

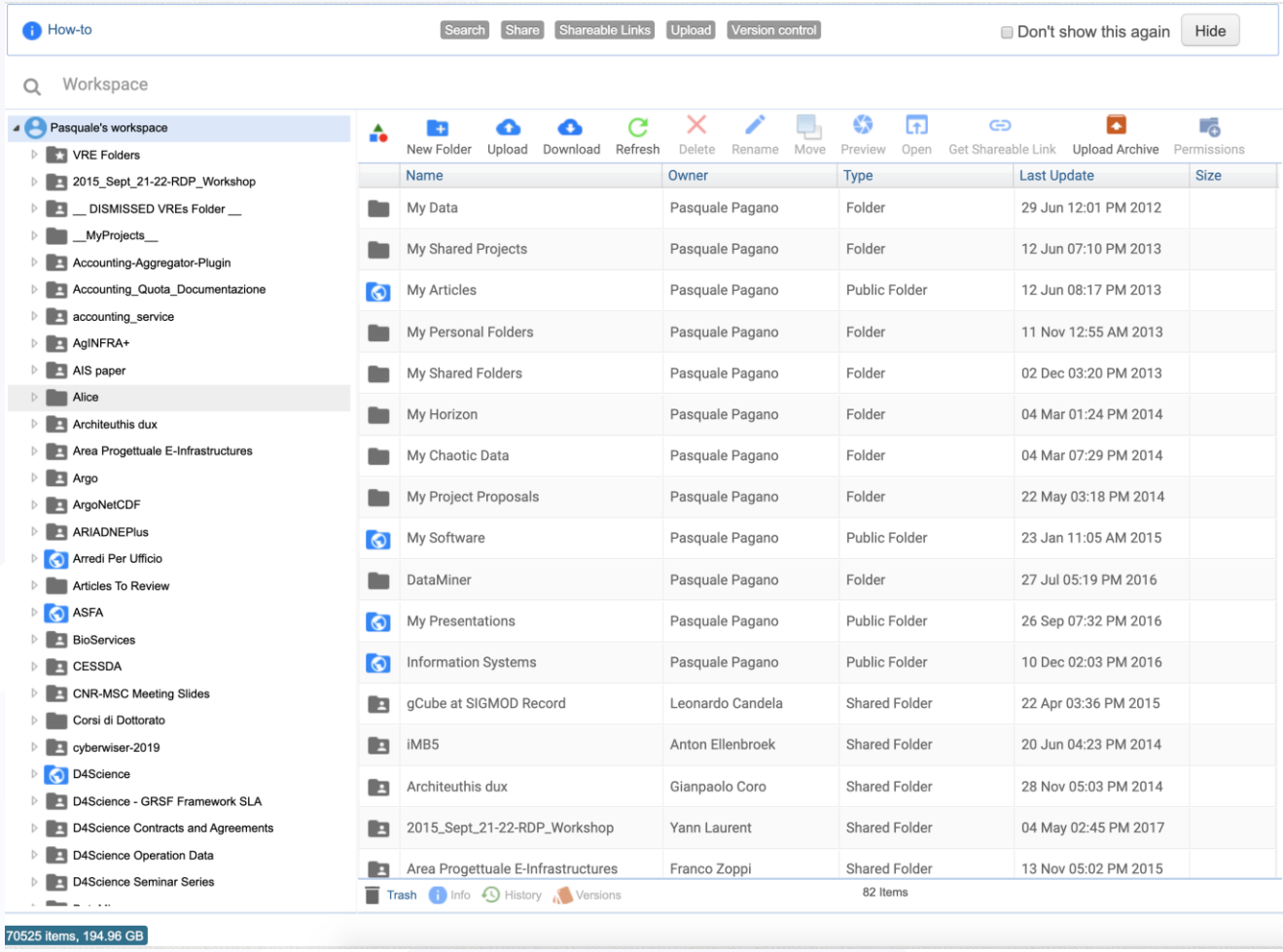
WORKSPACE

Resembles a **typical file system**

with **files organised in folders**,

yet it supports

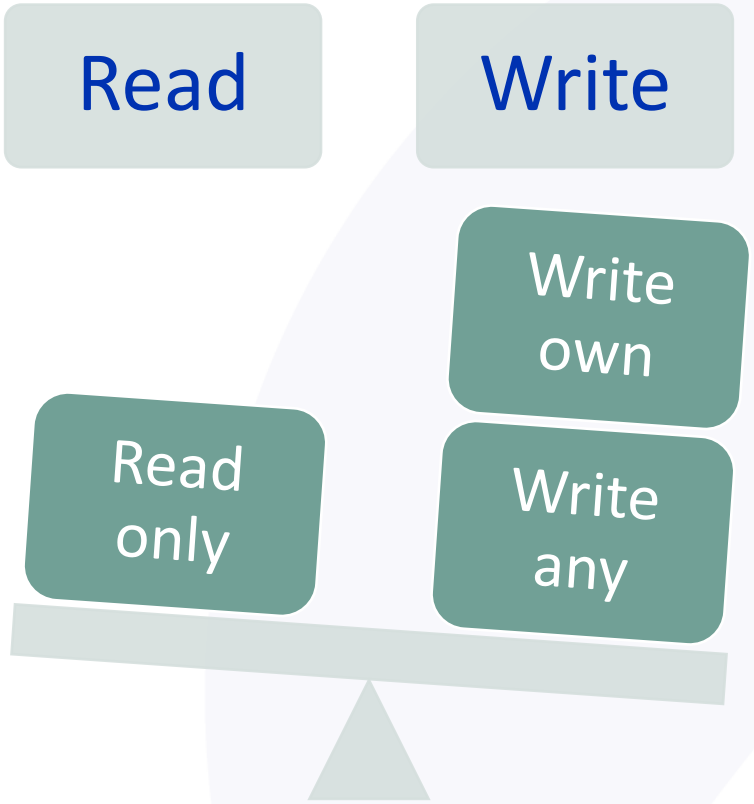
files, datasets, software methods, workflows, maps, …



Workspace Features

Folder Types

Shared folder permissions



Shared folder policies:

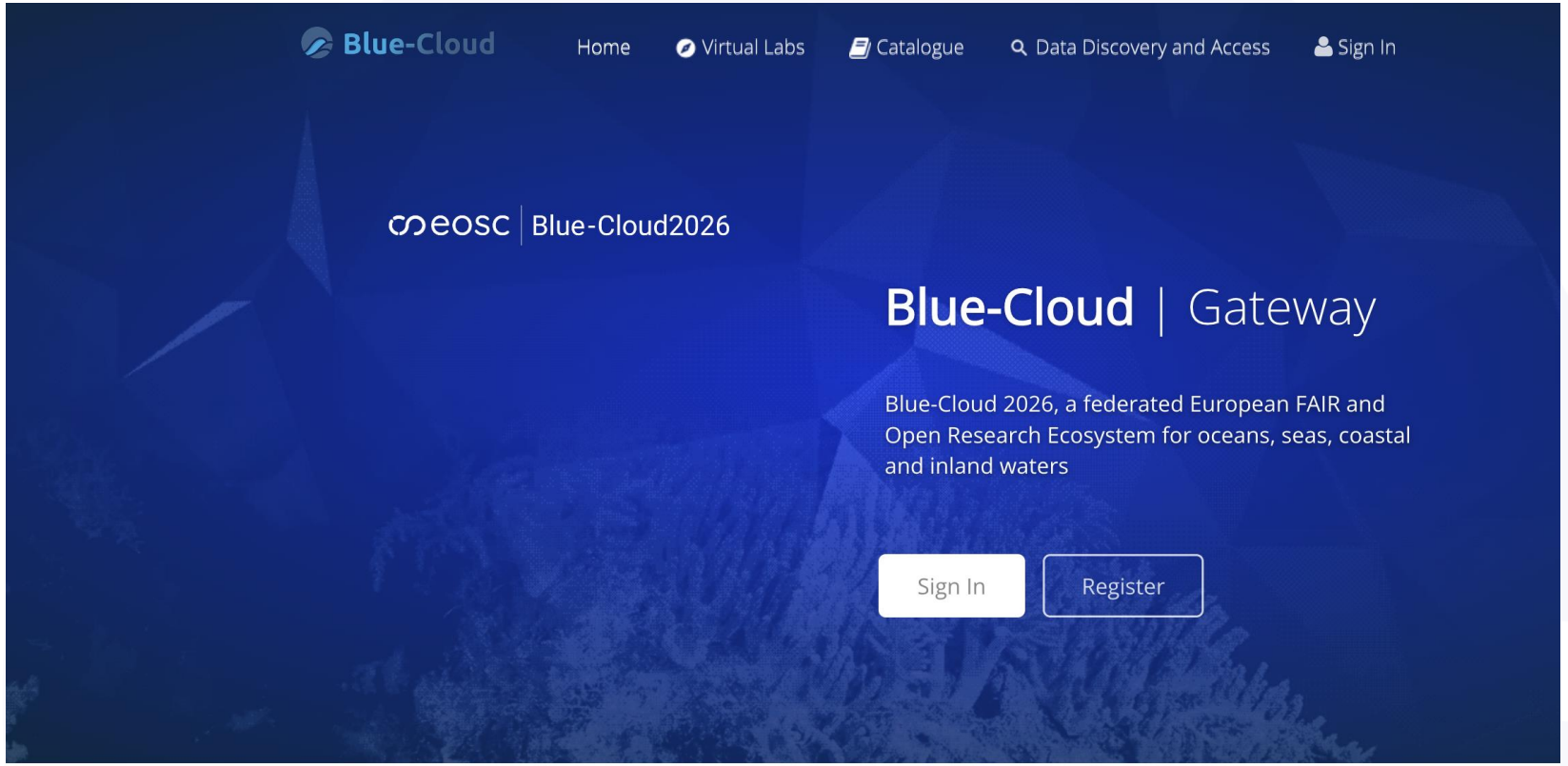
- ***Write Own***: users can only update/delete their own files
- ***Write Any***: any user can update/delete any file
- ***Read Only***: users can read any file but cannot update/delete

- Shared folder owner can
  - do anything;
  - nominate administrators to delete mngt. rights

- Subfolders
  - inherit parent permissions;
  - permissions can be restricted, not the opposite

How to access it

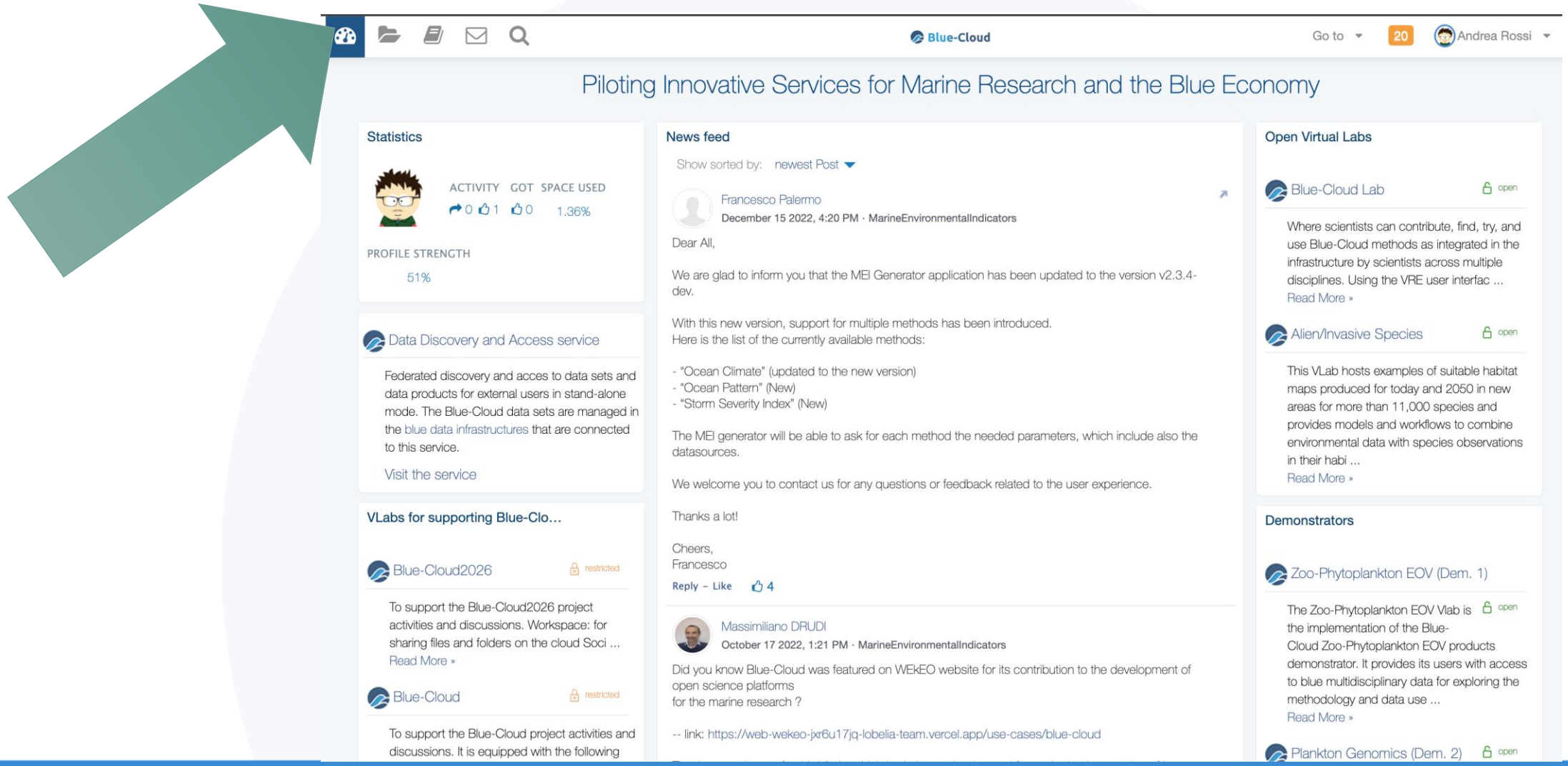
https://blue-cloud.d4science.org



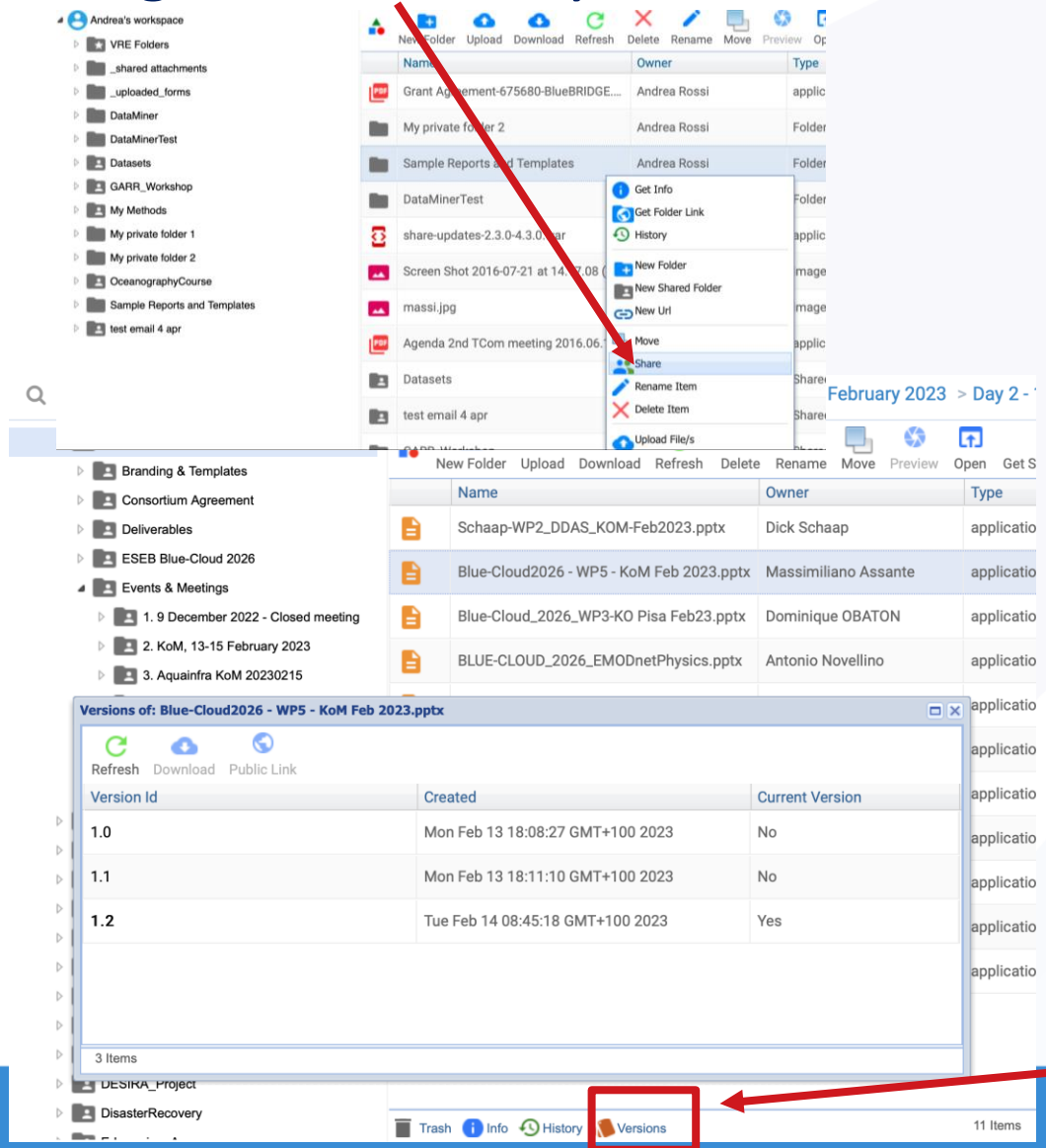
Blue-Cloud 2026 has received funding from the European Union's Horizon Europe programme call HORIZON-INFRA-2022-EOSC-01, Project ID 101094227. Blue-Cloud has received funding from the European Union's Horizon programme call BG-07-2019-2020, topic: [A] 2019 -

How to access it

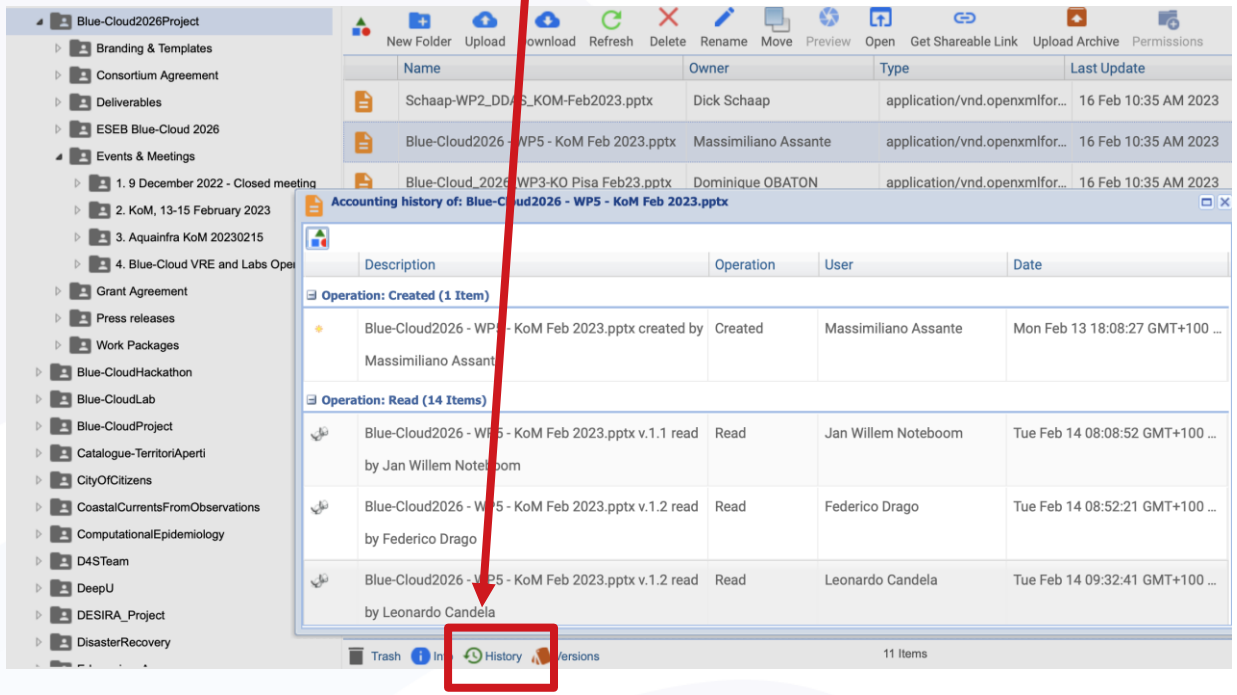
https://blue-cloud.d4science.org



Right click on any folder to share

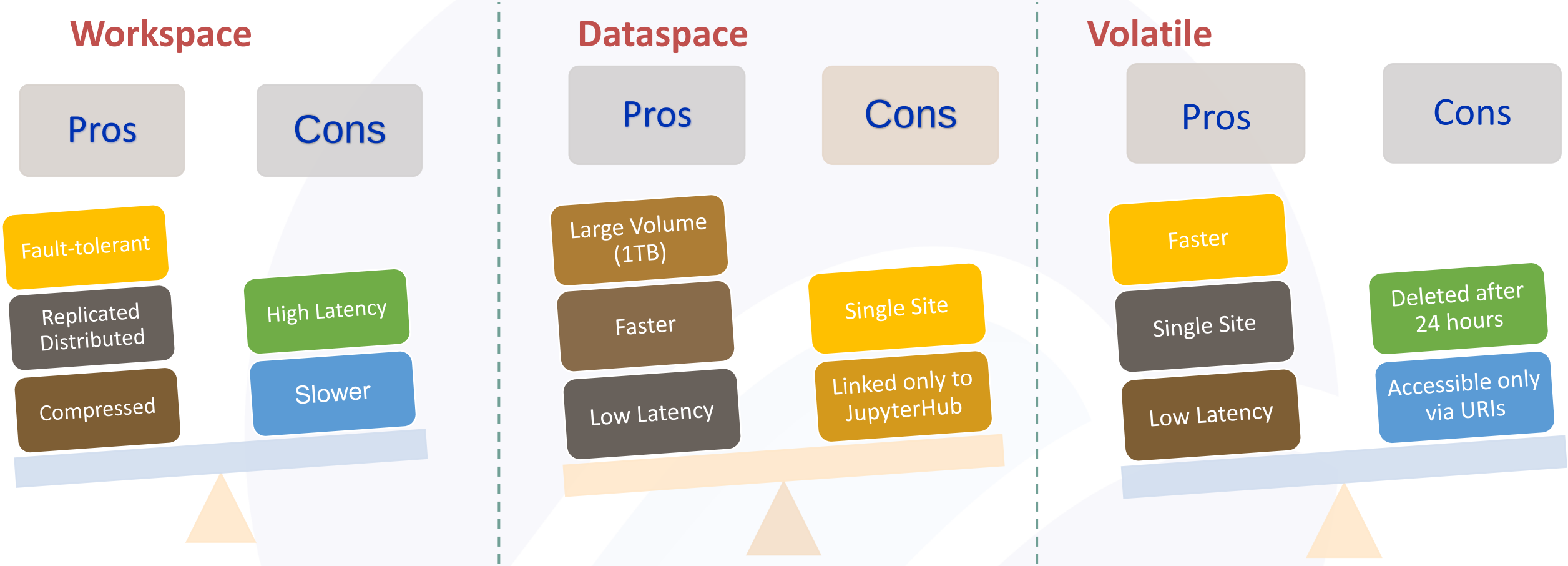


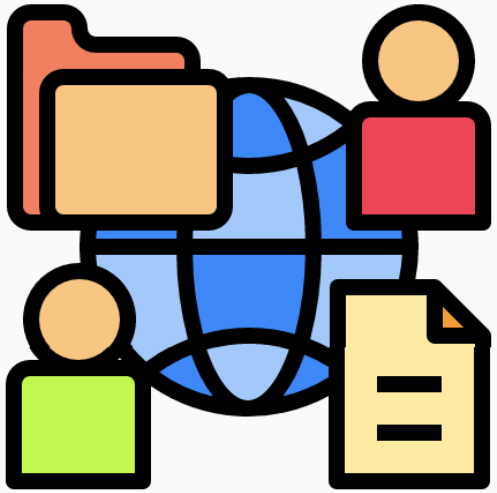
Click on History to see the activities



Click on Versions to see and access the different versions

Storage Layer





Blue-Cloud promotes technologies enabling the

sharing of datasets and methods

PUBLISHING

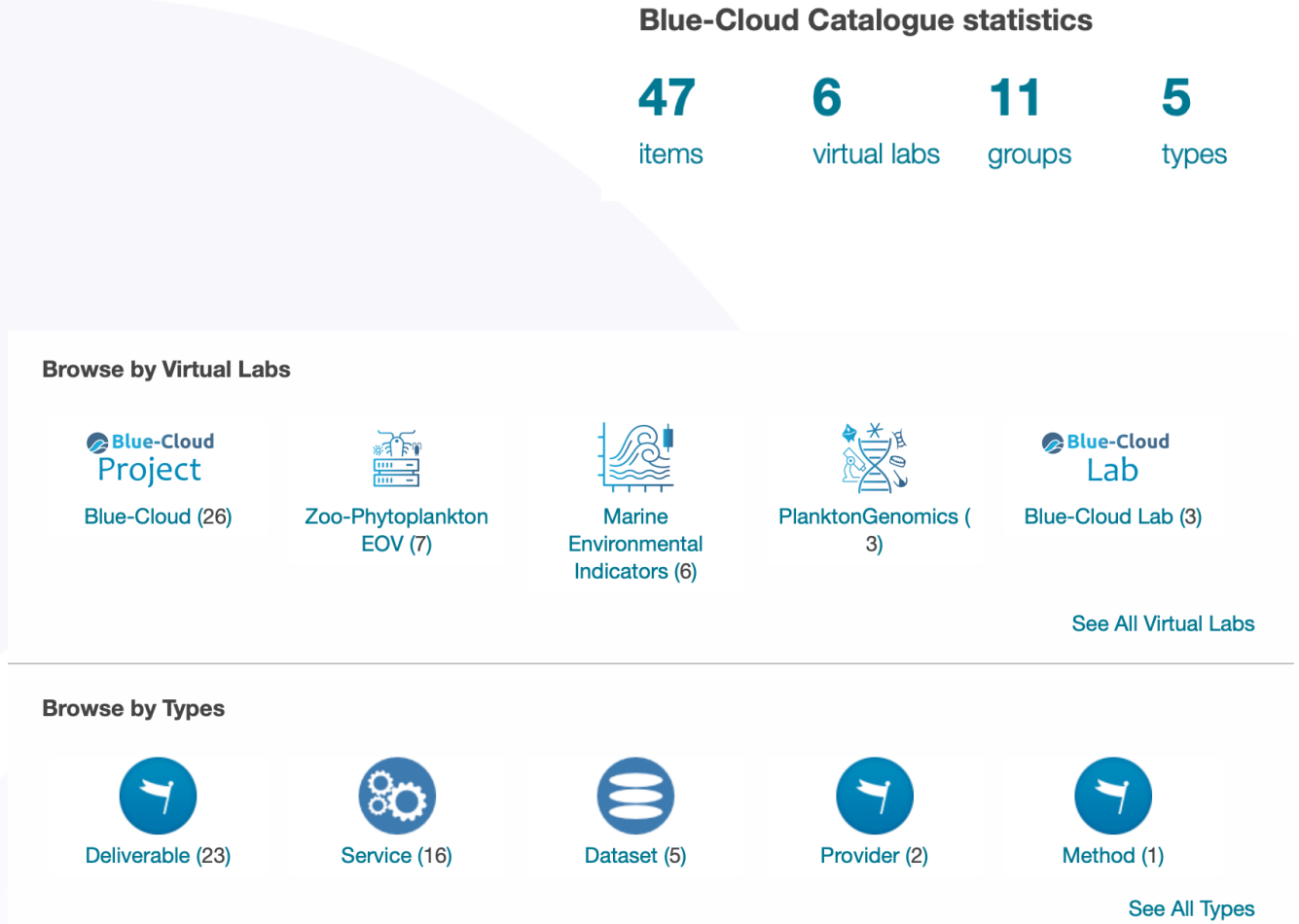
Resembles a catalogue of artefacts with **search and browse**

Every published item in the catalogue is characterised by:

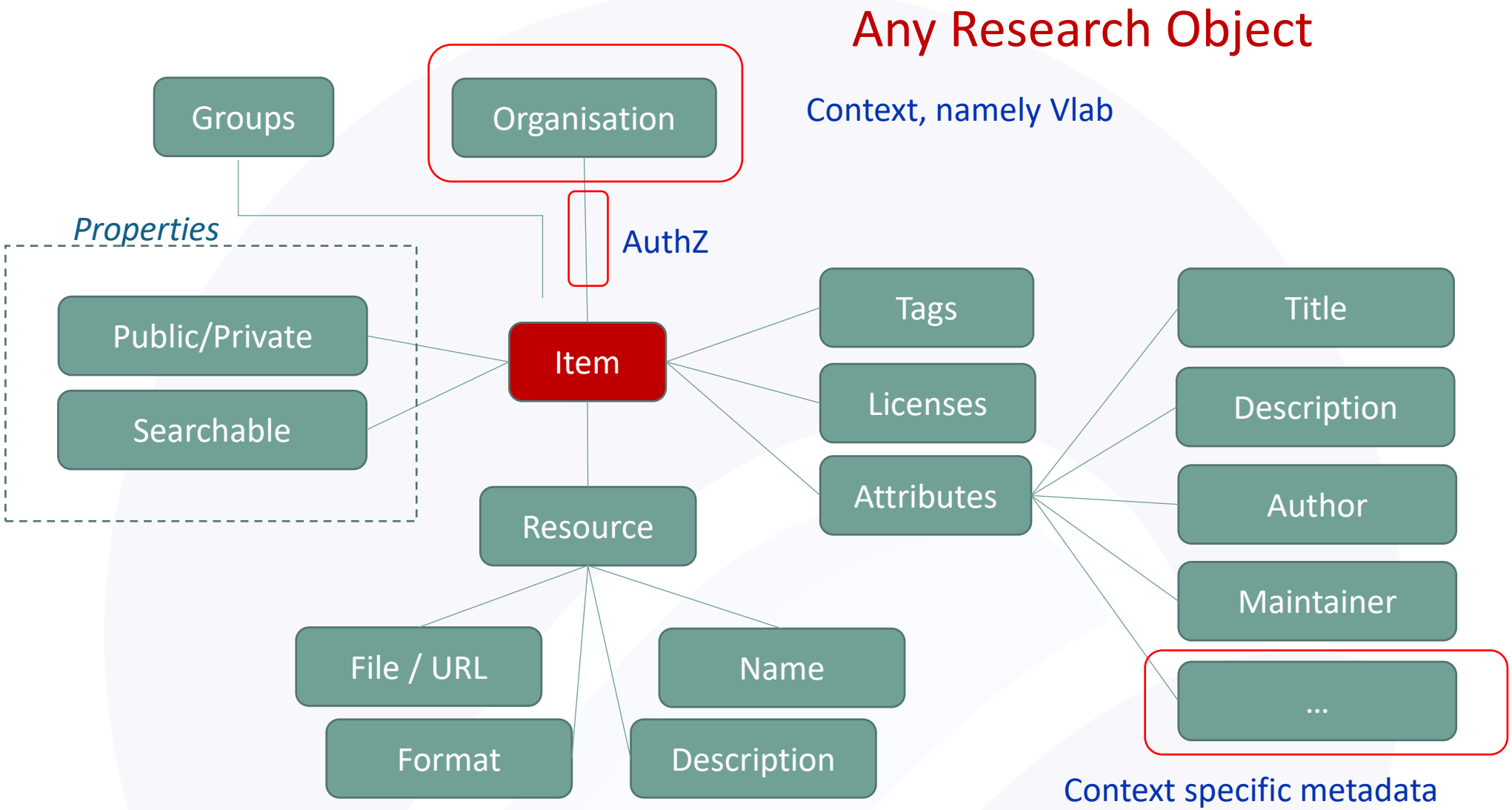
a **type**, which highlights its features

an **open ended set of metadata**

optional **resource**(s) representing the actual payload of the item.



What is an Item?



An item example

Organization

Title & Description

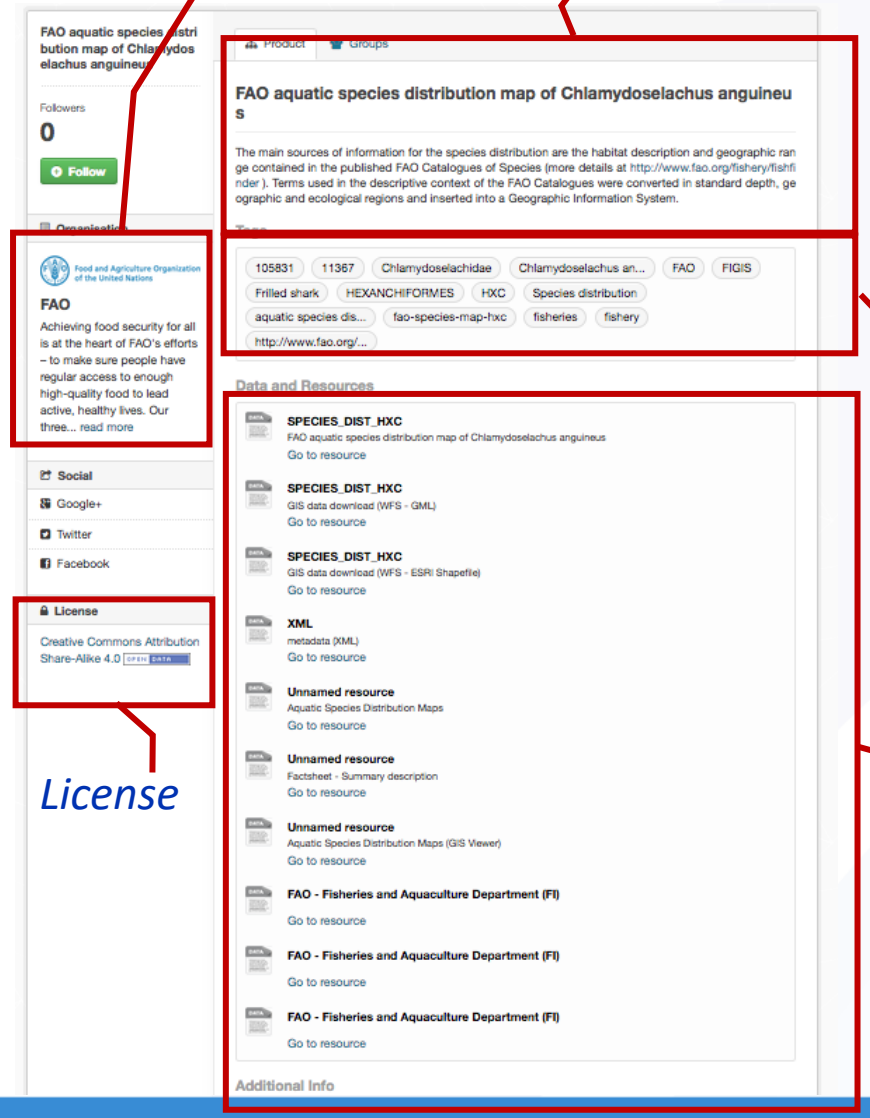
PURL
QR-Code

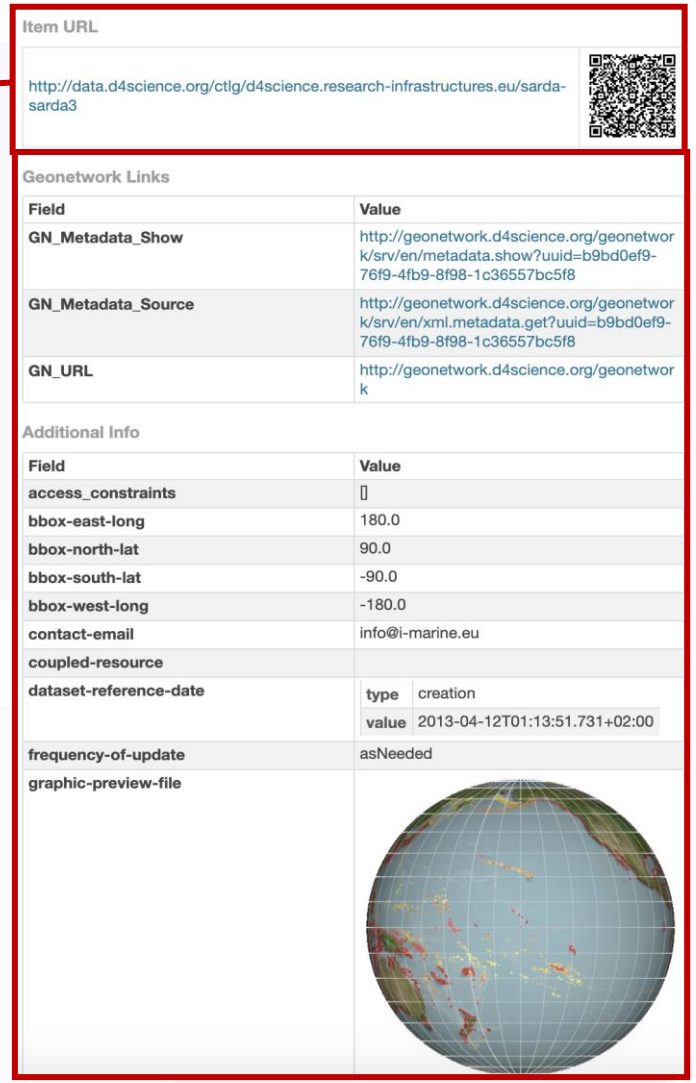
Tags

License

Resources
(item payloads)

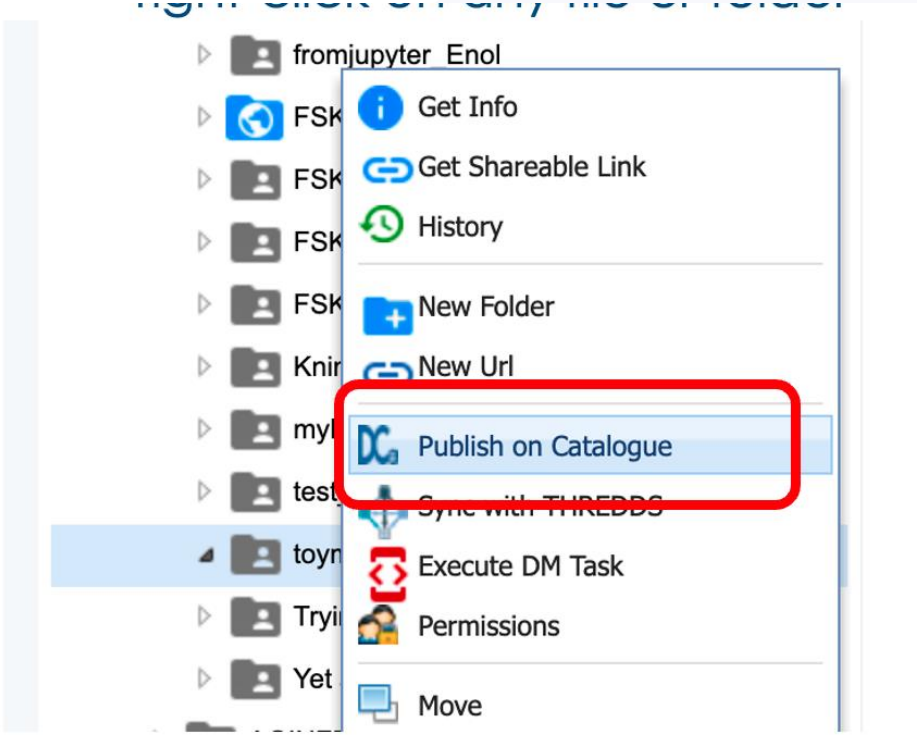
Metadata



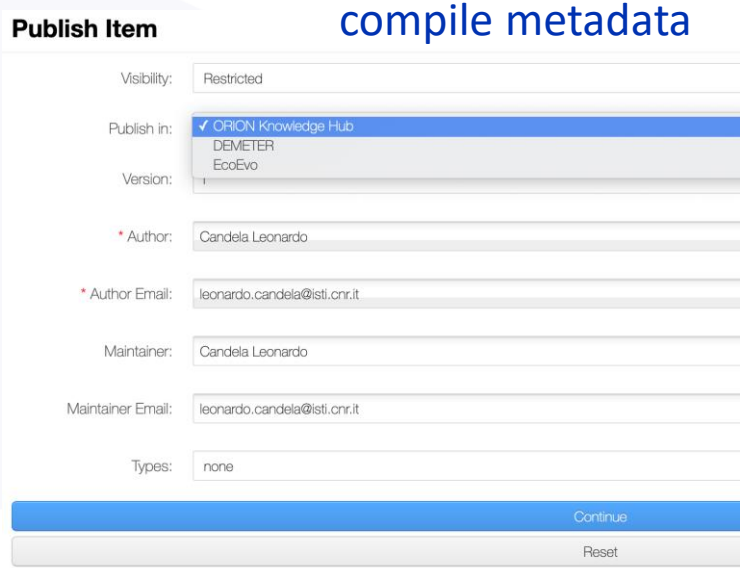


Workspace Integration

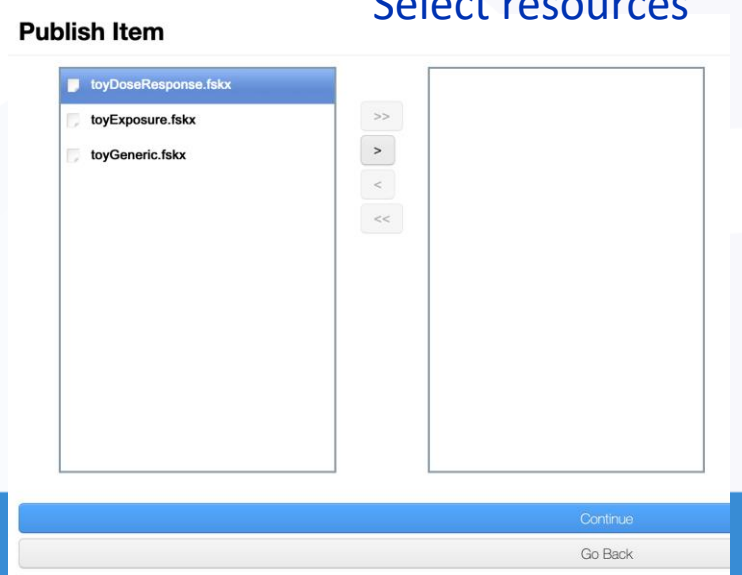
right-click on any file or folder



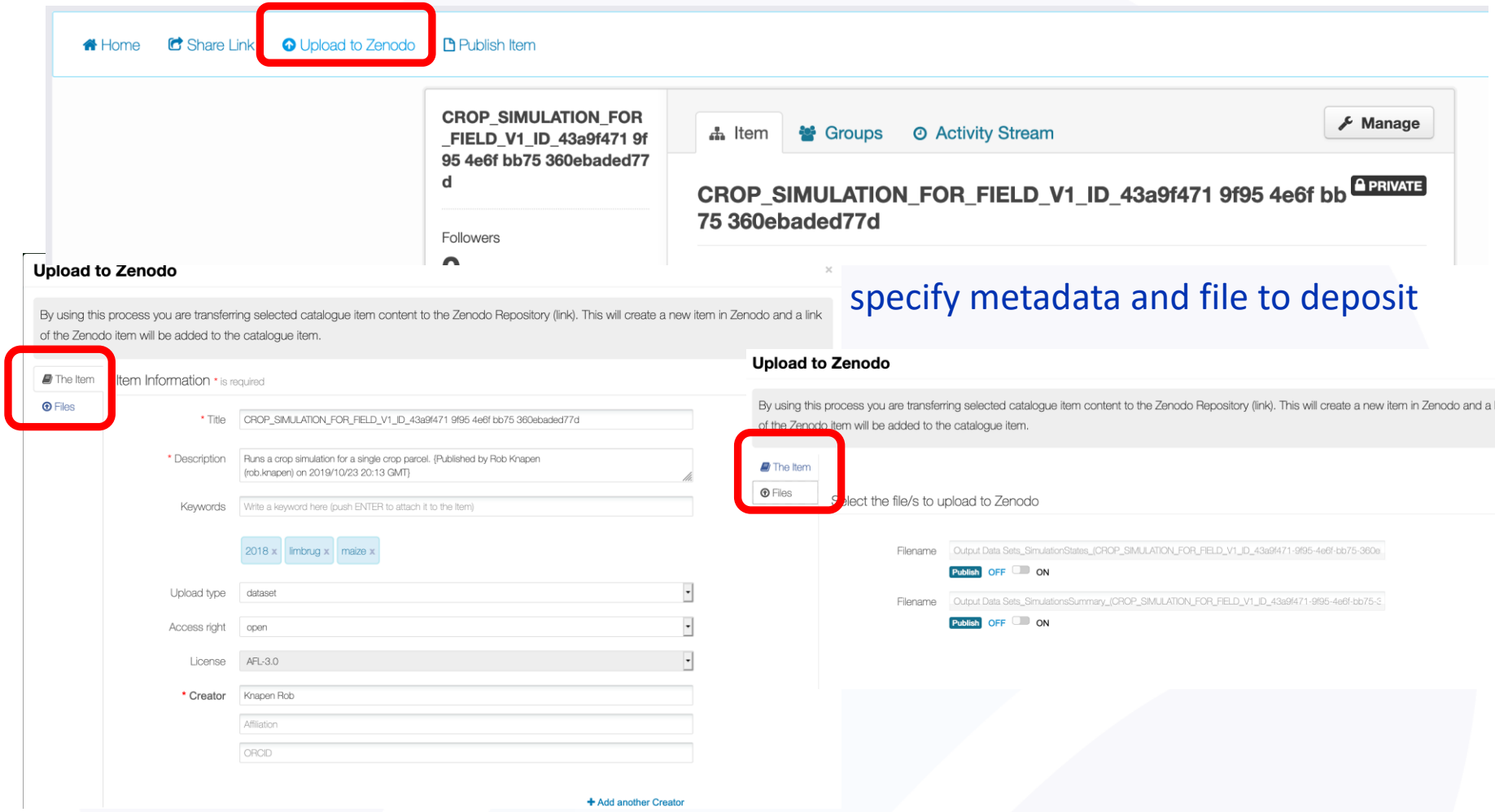
Select organization, typology and compile metadata



Select resources

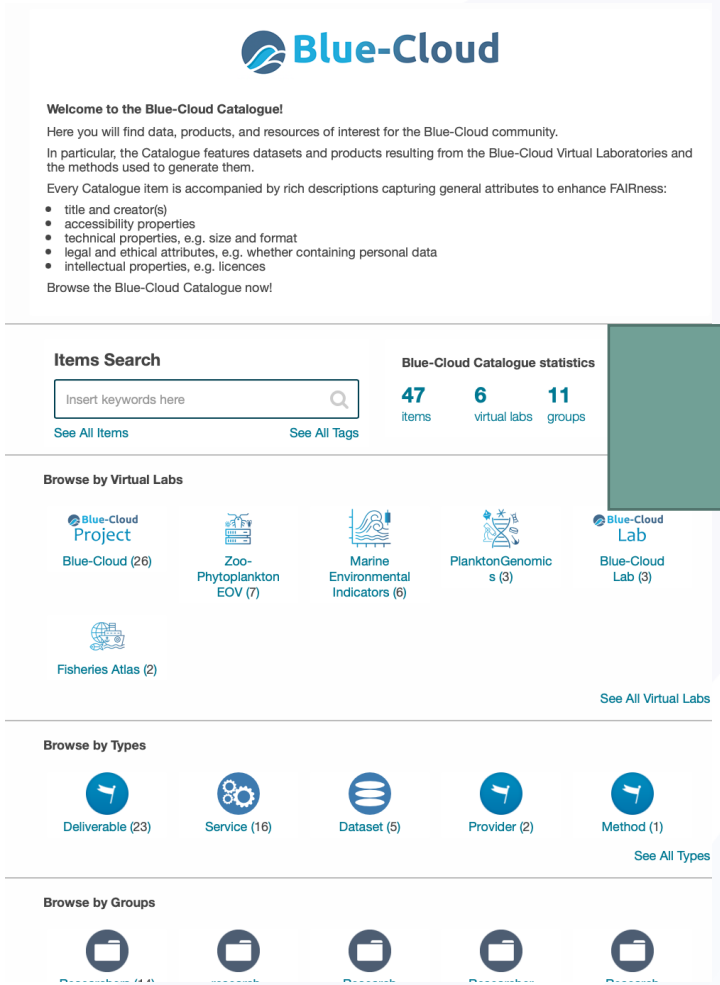


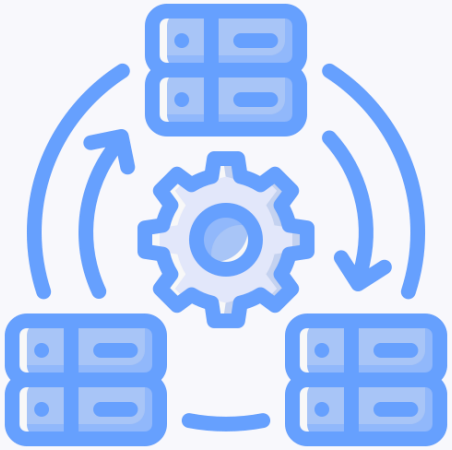
Zenodo Integration



Medium term: Blue-Cloud Catalogue to EOSC Market place

Blue-Cloud services, regularly registered in the Blue-Cloud Catalogue, will also be made accessible through the EOSC marketplace.

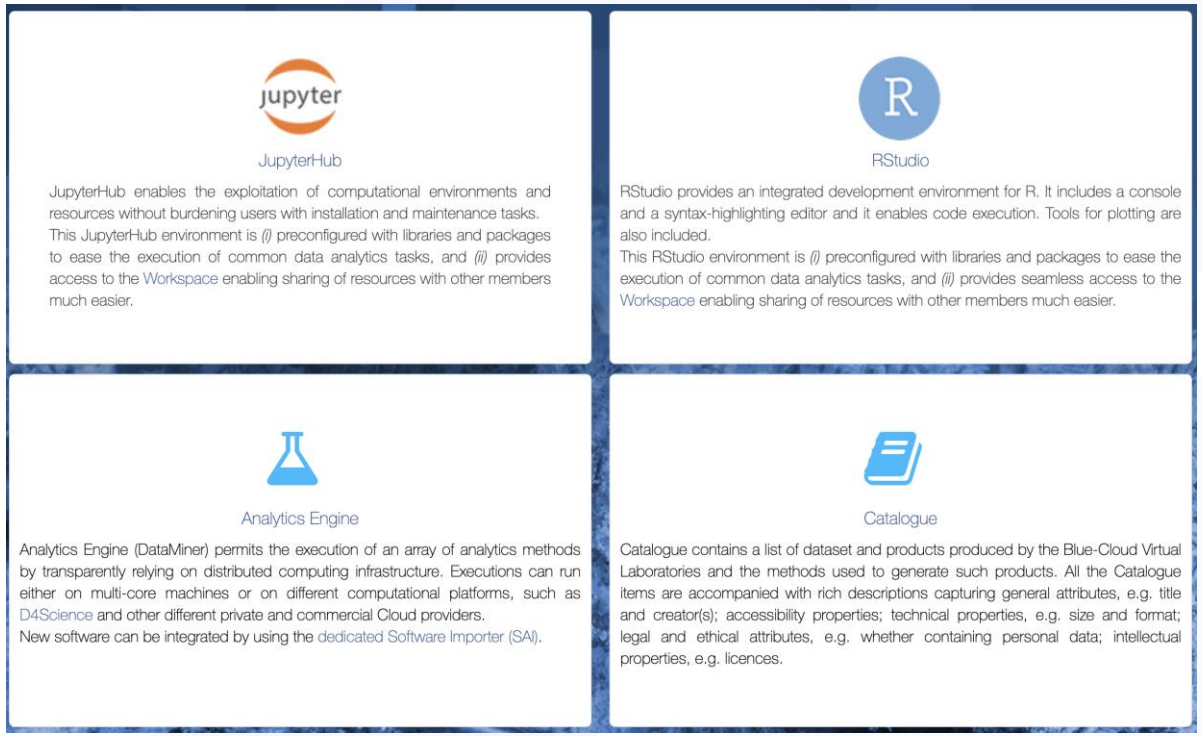


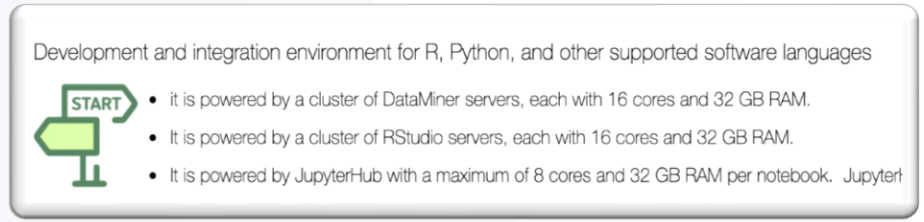


ANALYTICS COMPUTING FRAMEWORK

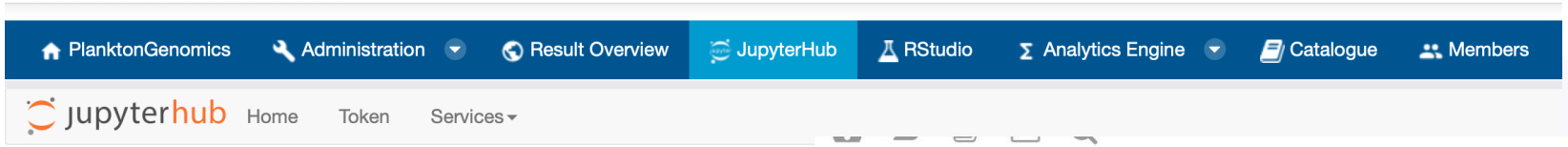
One place for executing analysis and processes

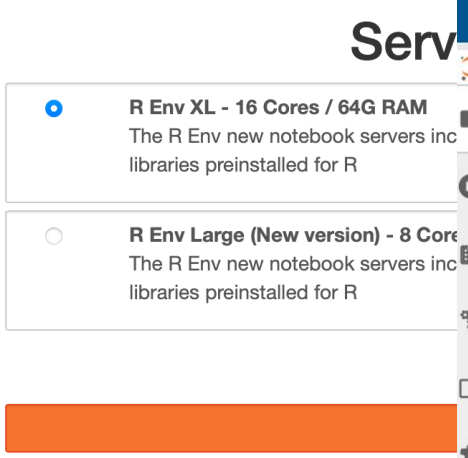
- interactive notebooks via JupyterHub and community-specific applications delivered as a Docker container extend the Analytics framework

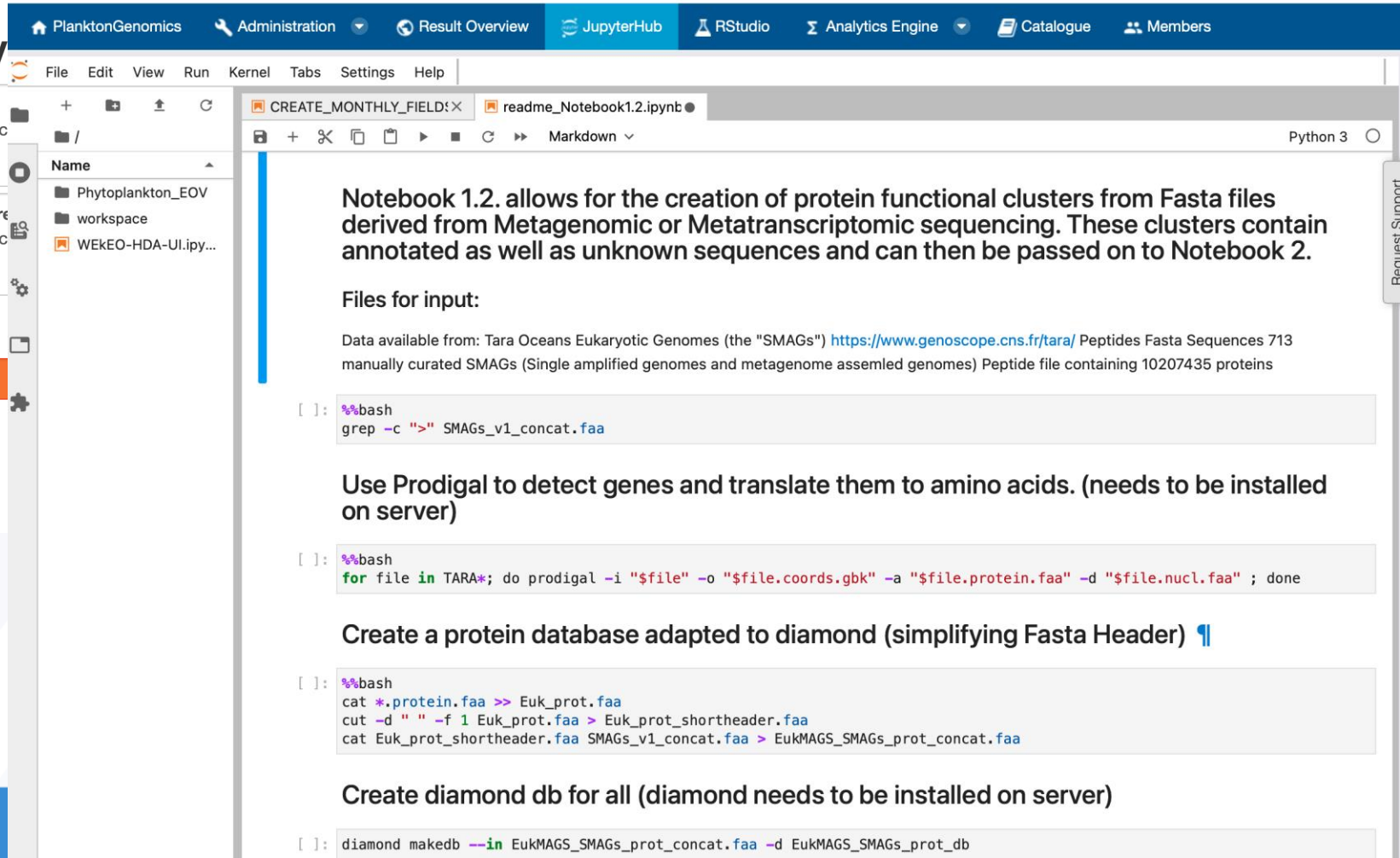




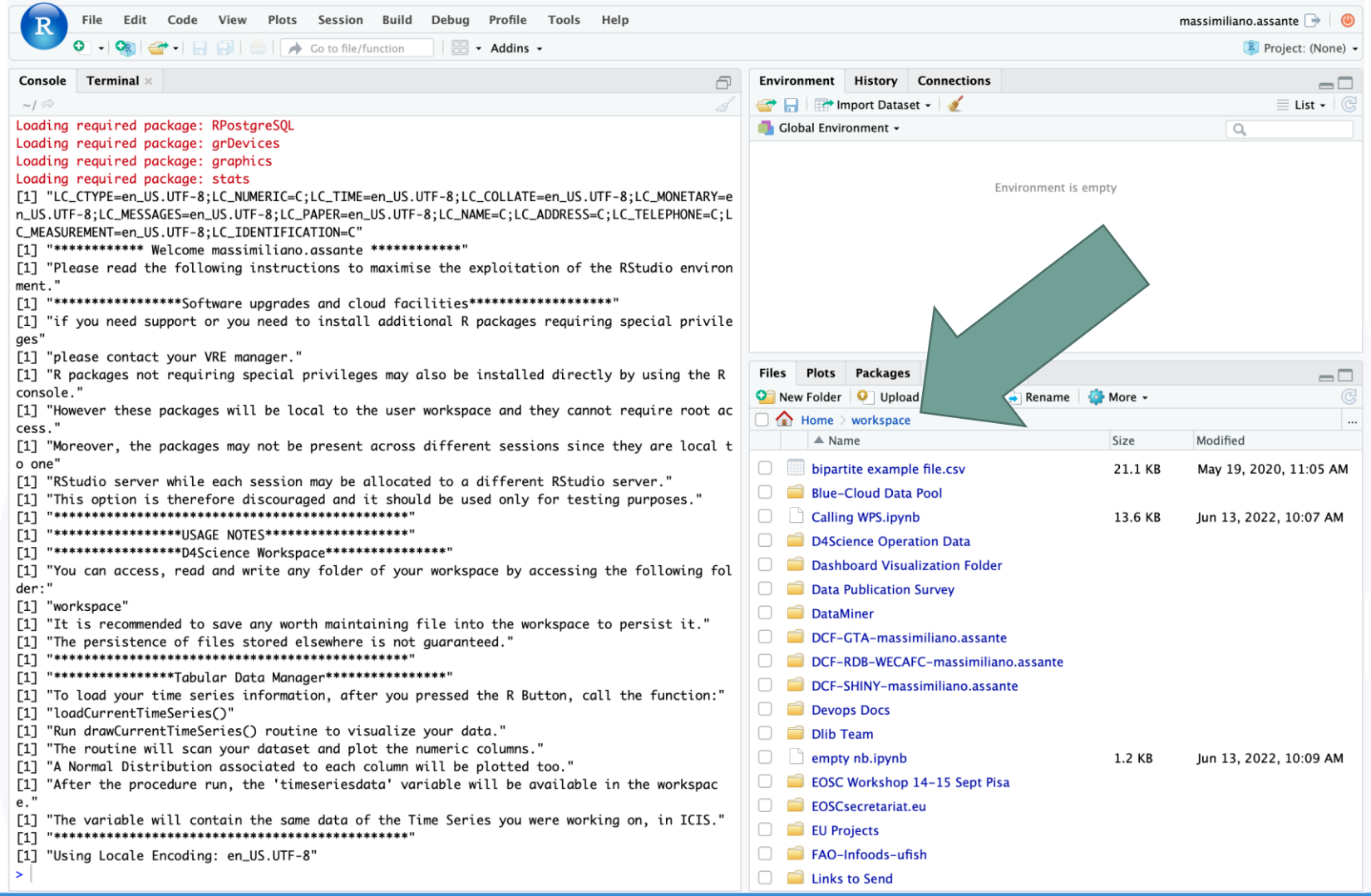
JupyterHub Example







Rstudio Example



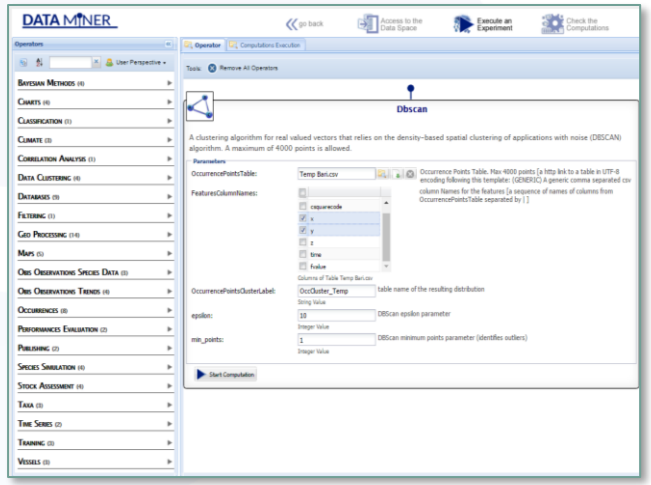
ANALYTICS ENGINE

Script/SW

Automatic Integration Process

SOFTWARE IMPORTER

Resulting Web UI and Web Service



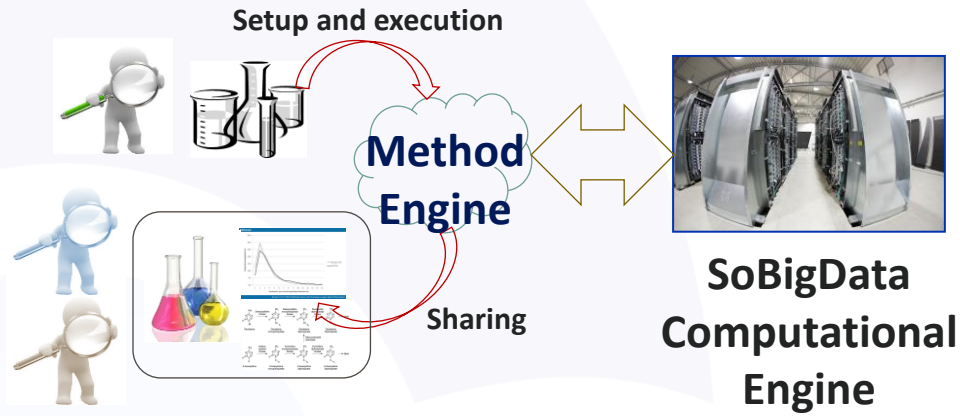
ANALYTICS ENGINE

Computational / Method engine in a Nutshell

- Helps scientists in performing **in-silico experiments**
- Supplies *"precooked"* methods

  as-a-Service
- Performs calculations in a

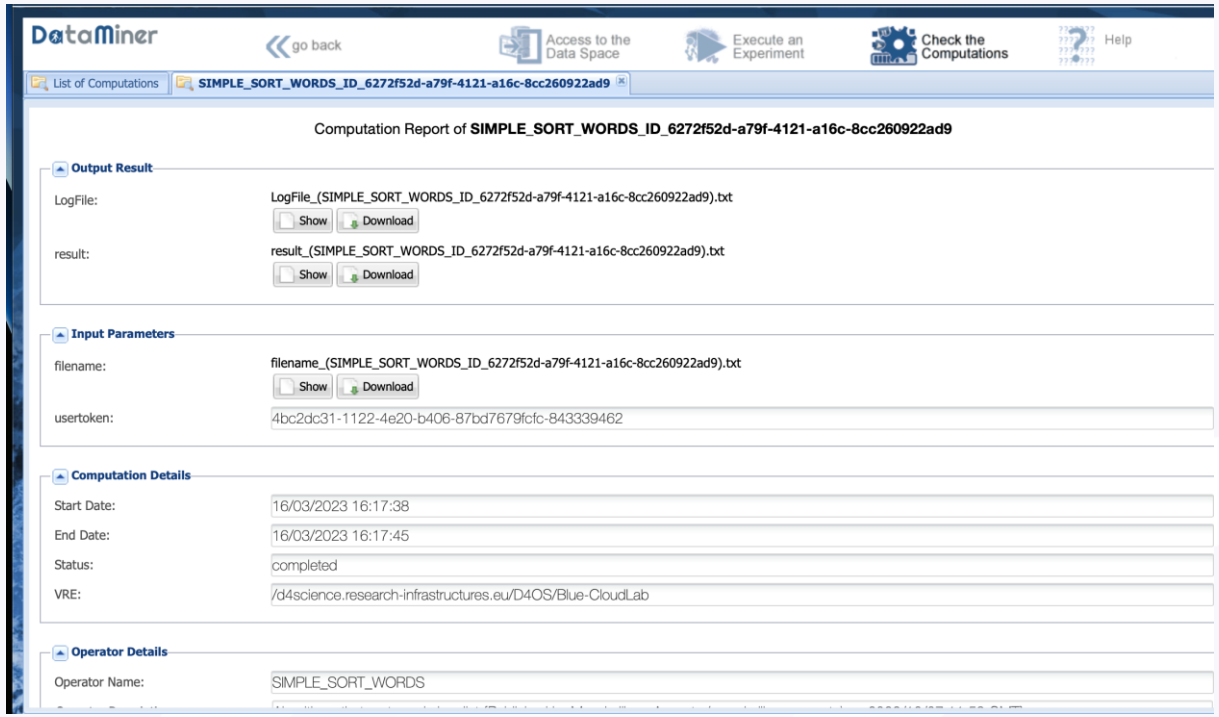
  seamless way to the users



- **Share** input, **results**, and **parameters** with colleagues by means of VREs

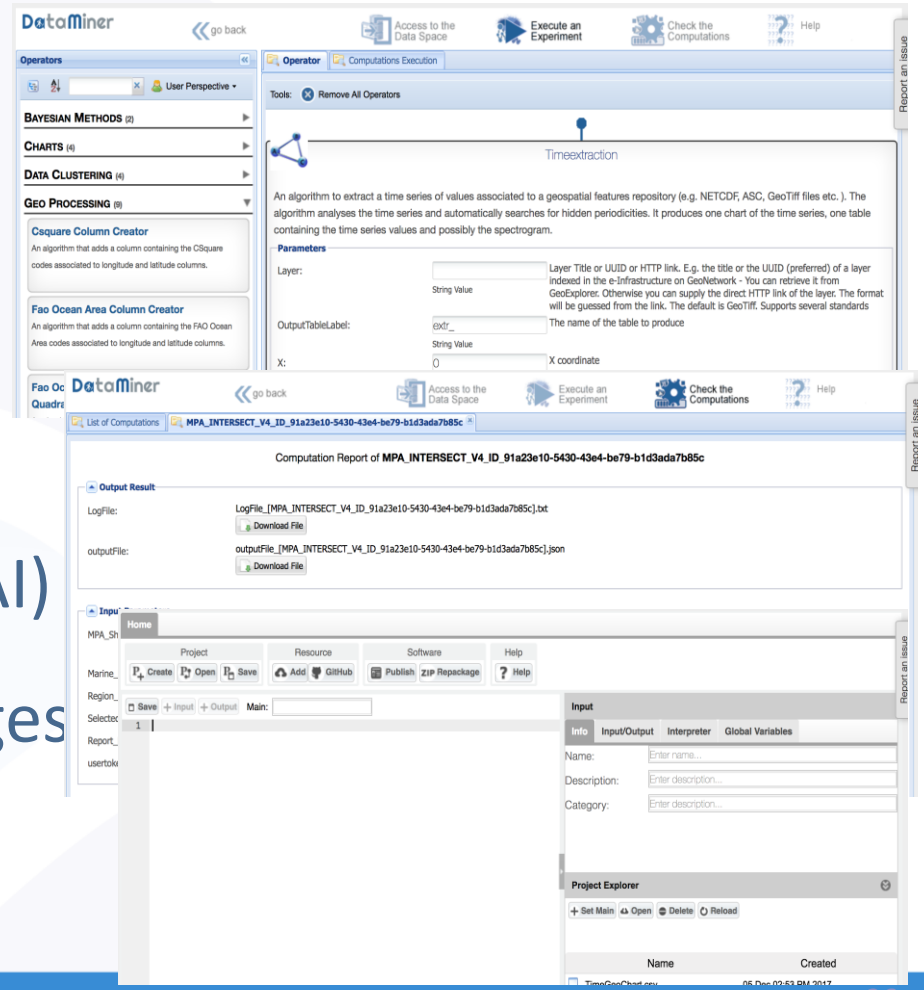
Process Execution Details

The execution can be repeated with or without the same input parameters



Method Engine Features

- User-friendly data analytics platform
- Per-method GUI and HTTPs access by WPS Standard
- Provenance management (PROV-O)
- Method integration support via dedicated tool (SAI) with extensibility WRT supported method languages



A note on the Provenance, Prov-O

(https://www.w3.org/TR/prov-o/)

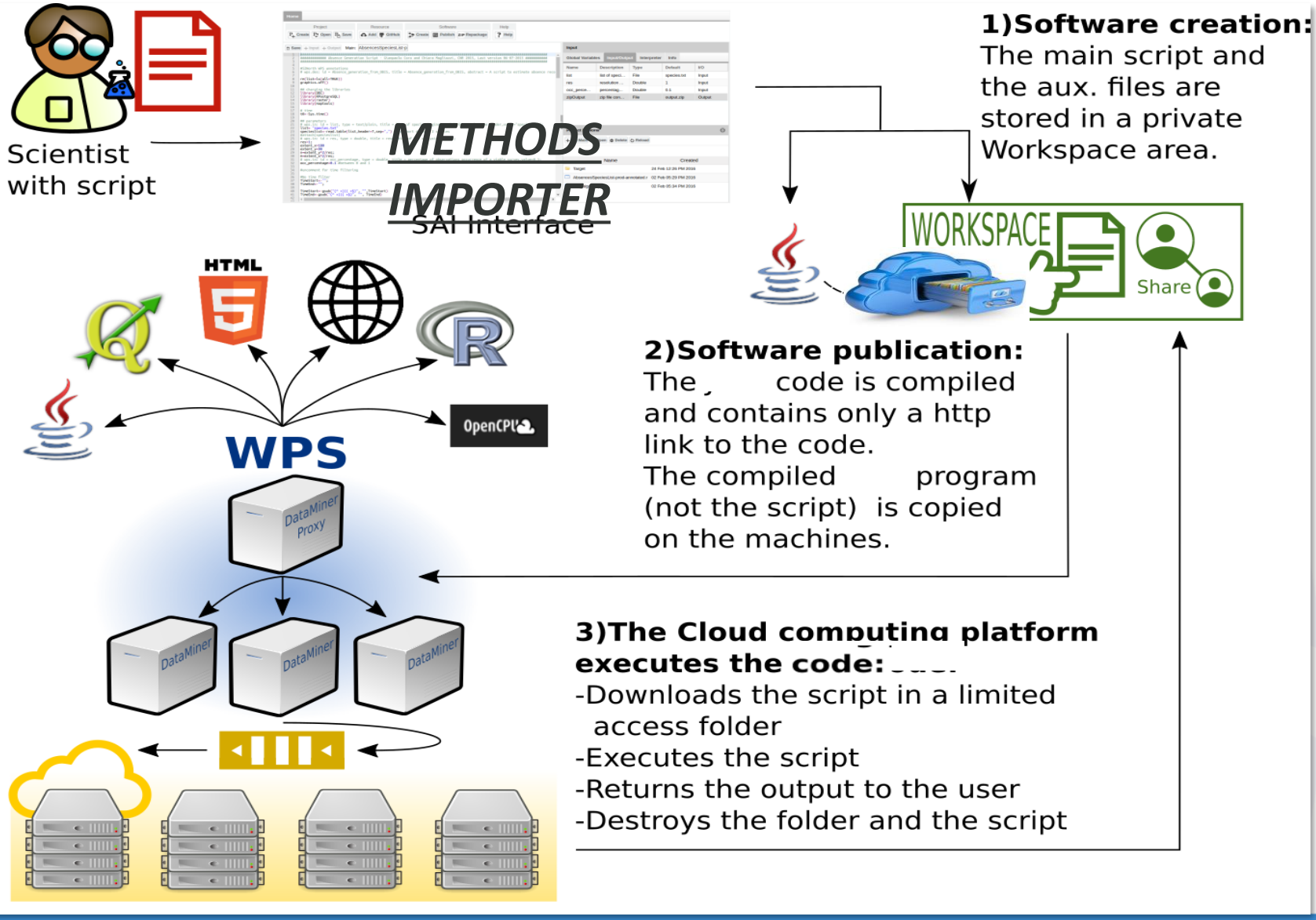
“Provenance is information about entities, activities, and people involved in producing a piece of data or thing, which can be used to form assessments about its quality, reliability or trustworthiness.”

The PROV Ontology (PROV-O) expresses the PROV Data Model using the OWL2 Web Ontology Language (OWL2).

It provides a set of classes, properties, and restrictions that can be used to represent and interchange provenance information generated in different systems and under different contexts.

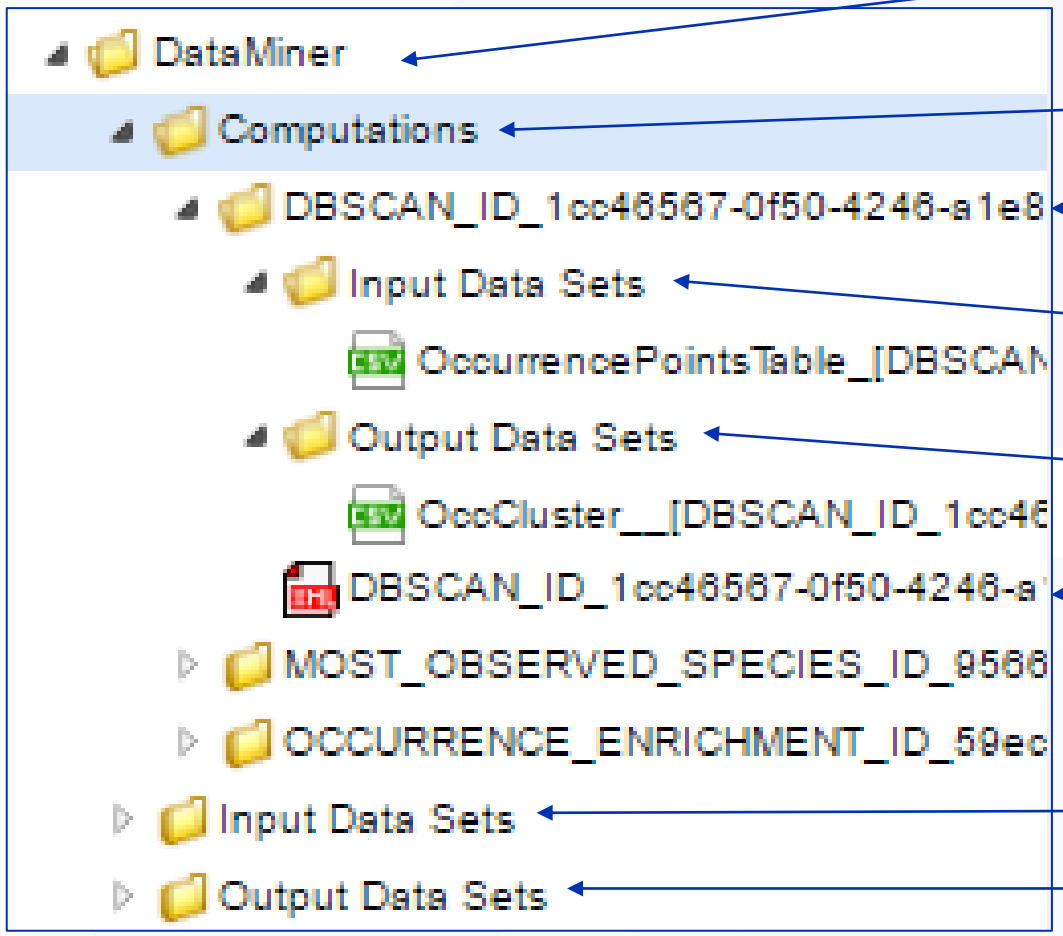

```
<?xml version="1.0" encoding="UTF-8"?><prov:document xmlns:d4s="http://d4science.org/#"
    xmlns:prov=http://www.w3.org/ns/prov#
xmlns:xsd="http://www.w3.org/2001/XMLSchema" xmlns:xsi="http://www.w3.org/2001/XMLSchema-
    instance">
  <prov:activity prov:id="d4s:FEED_FORWARD_NEURAL_NETWORK_TRAINER_ID_ba09841f-6656-4f6d-b6f1-
      b8fd532d32e9">
    <prov:startTime>11/11/2016 15:05:53</prov:startTime>
    <prov:endTime>11/11/2016 15:05:55</prov:endTime>
    <prov:type xsi:type="xsd:QName">d4s:computation</prov:type>
    <prov:softwareAgent prov:id="d4s:dataminer.d4science.org"/>
    <prov:person prov:id="d4s:gianpaolo.coro"/>
    <prov:entity prov:id="d4s:operator_name">
      <prov:value xsi:type="xsd:string">FEED_FORWARD_NEURAL_NETWORK_TRAINER</prov:value>
    </prov:entity>
    <prov:entity prov:id="d4s:operator_id">
      <prov:value xsi:type="xsd:string">org.gcube.dataanalysis.wps.statisticalmanager.
          synchserver.mappedclasses.
    transducerers.FEED_FORWARD_NEURAL_NETWORK_TRAINER</prov:value>
    </prov:entity>
    <prov:entity prov:id="d4s:operator_description">
      <prov:value xsi:type="xsd:string">The algorithm trains a Feed Forward Artificial
          Neural Network using an online Back-Propagation
   procedure and returns the training error and a binary file containing the trained network<
      /prov:value>
    </prov:entity>
    <prov:entity prov:id="d4s:VRE">
      <prov:value xsi:type="xsd:string">/d4science.research-infrastructures.eu/gCubeApps/
          RPrototypingLab</prov:value>
    </prov:entity>
    <prov:entity prov:id="d4s:status">
      <prov:value xsi:type="xsd:string">100</prov:value>
    </prov:entity>
    <prov:entity prov:id="d4s:inputTable_[FEED_FORWARD_NEURAL_NETWORK_TRAINER_ID_ba09841f
        -6656-4f6d-b6f1-b8fd532d32e9].csv">
      <prov:value xsi:type="xsd:string">http://data.d4science.org/
          Tm5UMkwyel1GY0tEa2RCMXpYd3JUZmdyWXYyb2NubjdHbWJQNStIS0N6Yz0</prov:value>
      <prov:activity prov:ref="d4s:FEED_FORWARD_NEURAL_NETWORK_TRAINER_ID_ba09841f-6656-4
          f6d-b6f1-b8fd532d32e9"/>
      <prov:entity prov:ref="d4s:FEED_FORWARD_NEURAL_NETWORK_TRAINER"/>
      <prov:entity prov:id="d4s:data_description">
        <prov:value xsi:type="xsd:string">inputTable</prov:value>
      </prov:entity>
      <prov:type xsi:type="xsd:QName">d4s:IMPORTED</prov:type>
      <prov:type xsi:type="xsd:QName">d4s:text/csv</prov:type>
    </prov:entity>
    <prov:entity prov:id="d4s:learningRate">
      <prov:value xsi:type="xsd:string">0.5</prov:value>
      <prov:activity prov:ref="d4s:FEED_FORWARD_NEURAL_NETWORK_TRAINER_ID_ba09841f-6656-4
          f6d-b6f1-b8fd532d32e9"/>
      <prov:entity prov:ref="d4s:FEED_FORWARD_NEURAL_NETWORK_TRAINER"/>
      <prov:entity prov:id="d4s:data_description">
        <prov:value xsi:type="xsd:string">learningRate</prov:value>
      </prov:entity>
    </prov:entity>
    <prov:entity prov:id="d4s:TrainedNeuralNetwork">
      <prov:value xsi:type="xsd:string">http://data.d4science.org/
          ST1YeDRwWDFIYn1Ea2RCMXpYd3JUY0JNWHZNTmN5VD1HbWJQNStIS0N6Yz0</prov:value>
      <prov:activity prov:ref="d4s:FEED_FORWARD_NEURAL_NETWORK_TRAINER_ID_ba09841f-6656-4
          f6d-b6f1-b8fd532d32e9"/>
      <prov:entity prov:ref="d4s:FEED_FORWARD_NEURAL_NETWORK_TRAINER"/>
      <prov:entity prov:id="d4s:data_description">
        <prov:value xsi:type="xsd:string">Trained Neural Network</prov:value>
      </prov:entity>
      <prov:type xsi:type="xsd:QName">d4s:COMPUTED</prov:type>
      <prov:type xsi:type="xsd:QName">d4s:application/d4science</prov:type>
    </prov:entity>
  </prov:activity>
</prov:document>
```

Integration Workflow, more details …



Method Engine and User's Workspace

User's Workspace



Dataminer folder

Overall computations folder

Computation

Comp. Input data

Comp. Output data

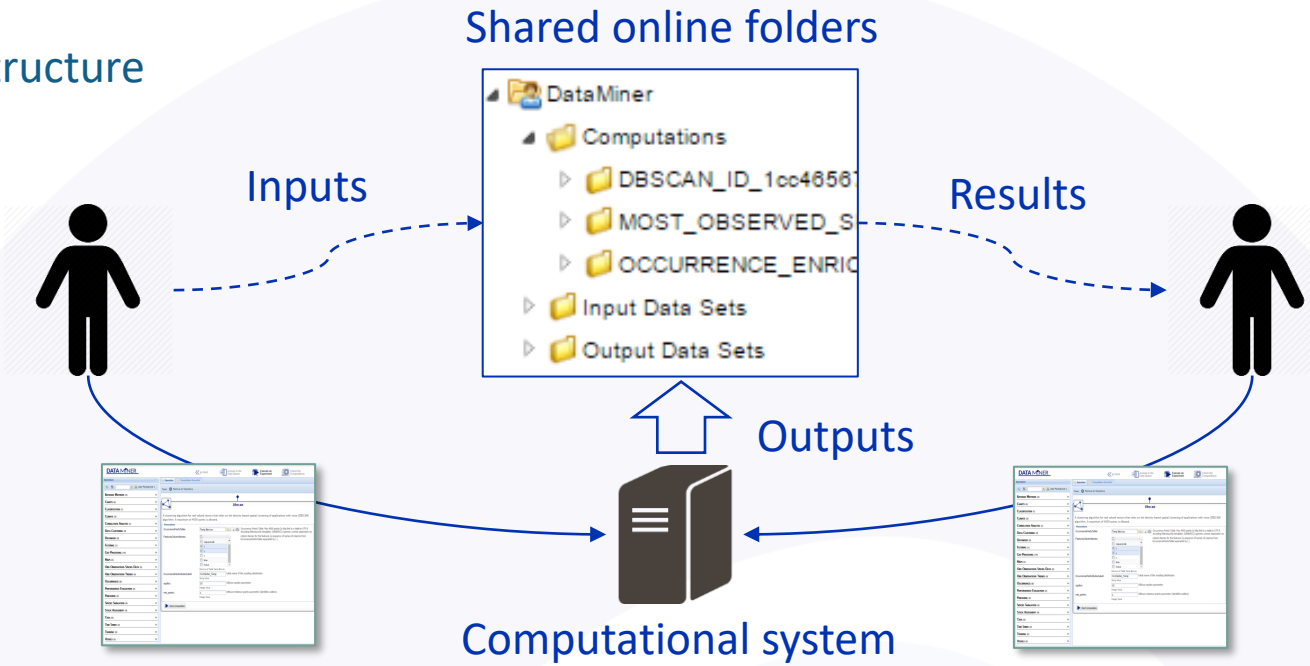
Provenance info

Overall Input Data

Overall Output Data

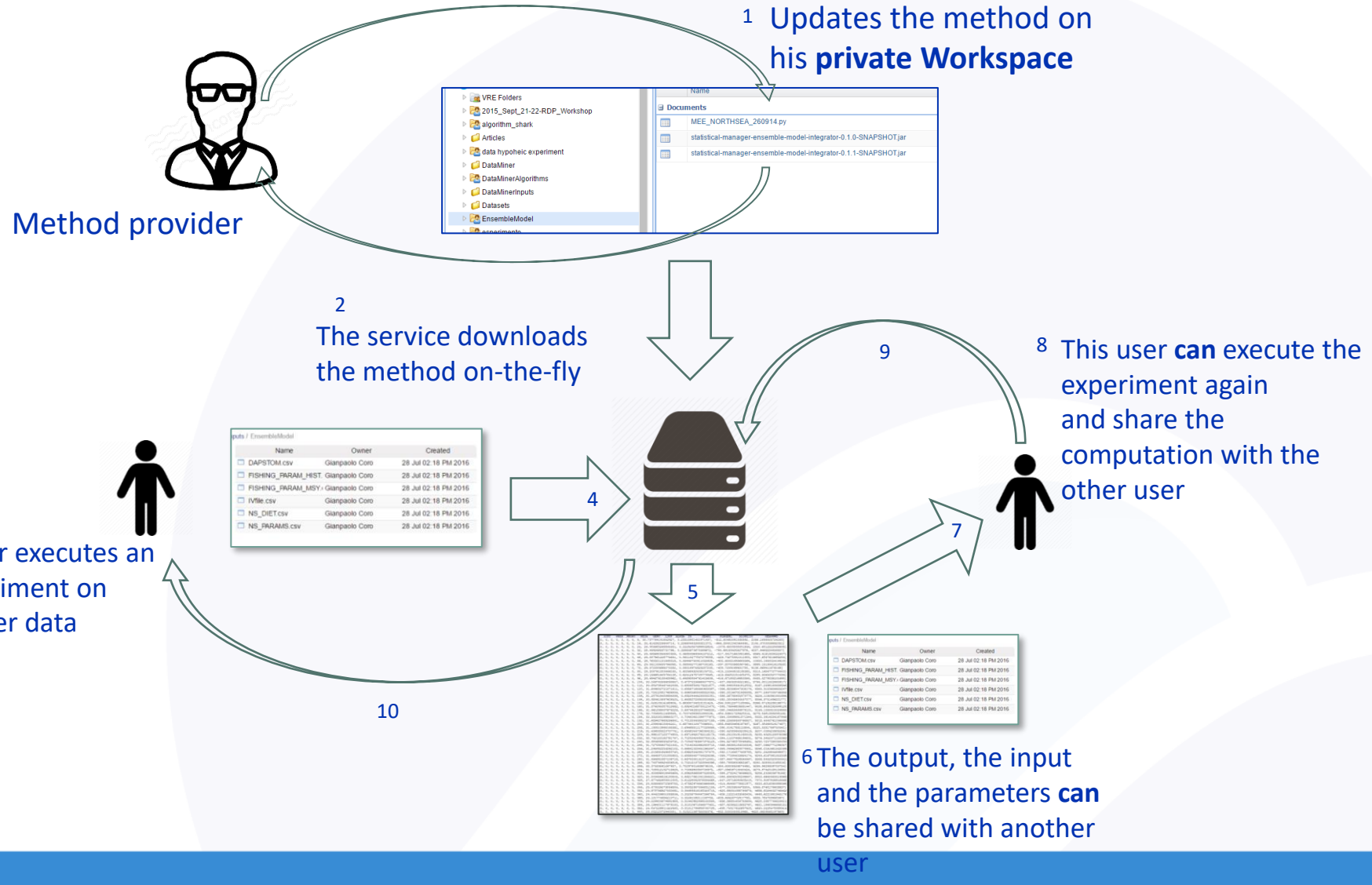
Collaborative experiments

In the e-Infrastructure

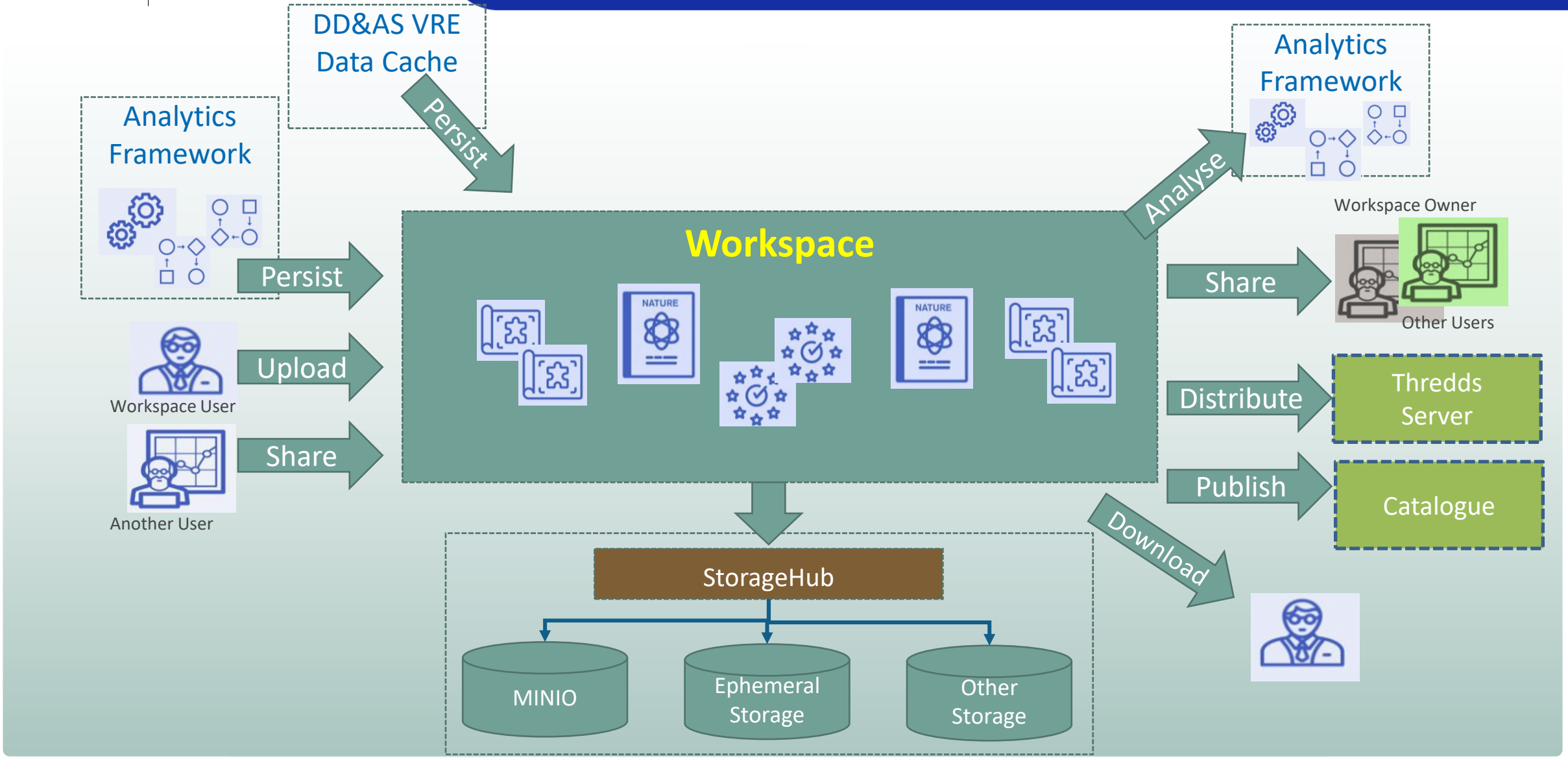


Through third party software

Code Privacy Guaranteed



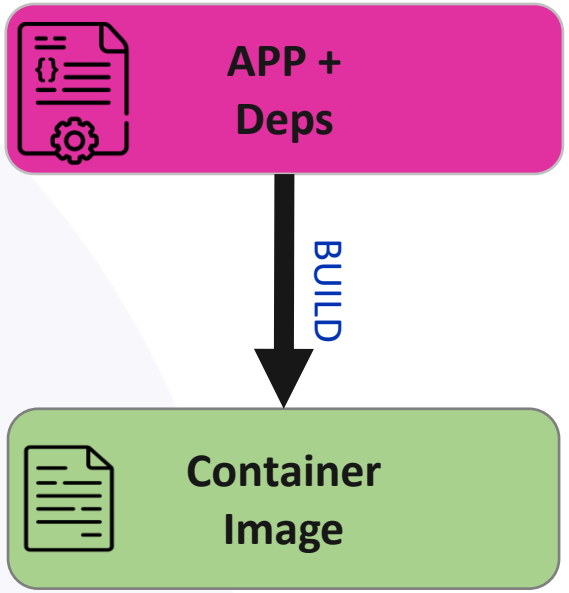
Workspace


DD&AS VRE
Data Cache
Persist
Analytics
Framework
Persist
Upload
Share
Workspace User
Another User
Workspace
NATURE
NATURE
Analyse
Analytics
Framework
Workspace Owner
Share
Other Users
Distribute
Thredds
Server
Publish
Catalogue
Download
StorageHub
MINIO
Ephemeral
Storage
Other
Storage

What if I have my own web application?

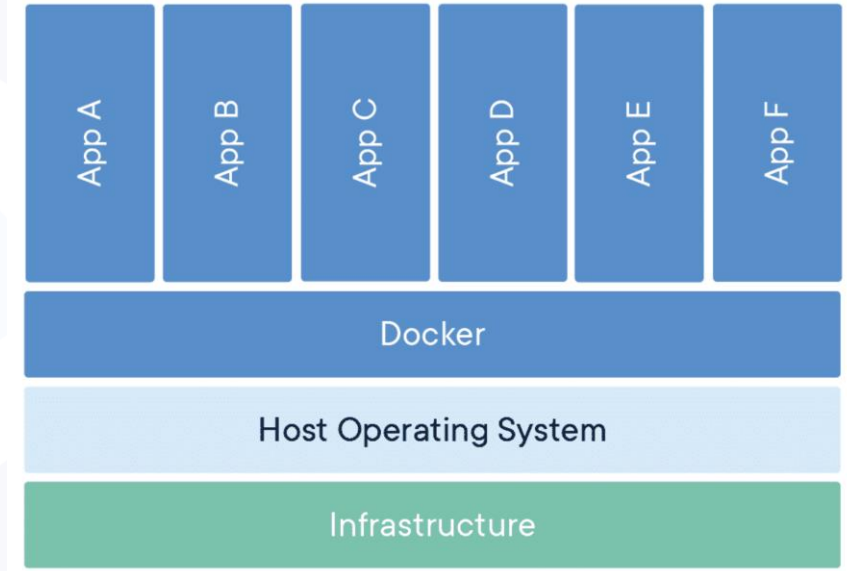
Custom Service as Containerized Application - Images

- Containerized applications are applications that run in isolated runtime environments called *containers*
- *containers* encapsulate an application with all its dependencies, including system libraries, binaries, and configuration files.
- Container images include (in a file) everything a container needs to run—the container engine such as Docker



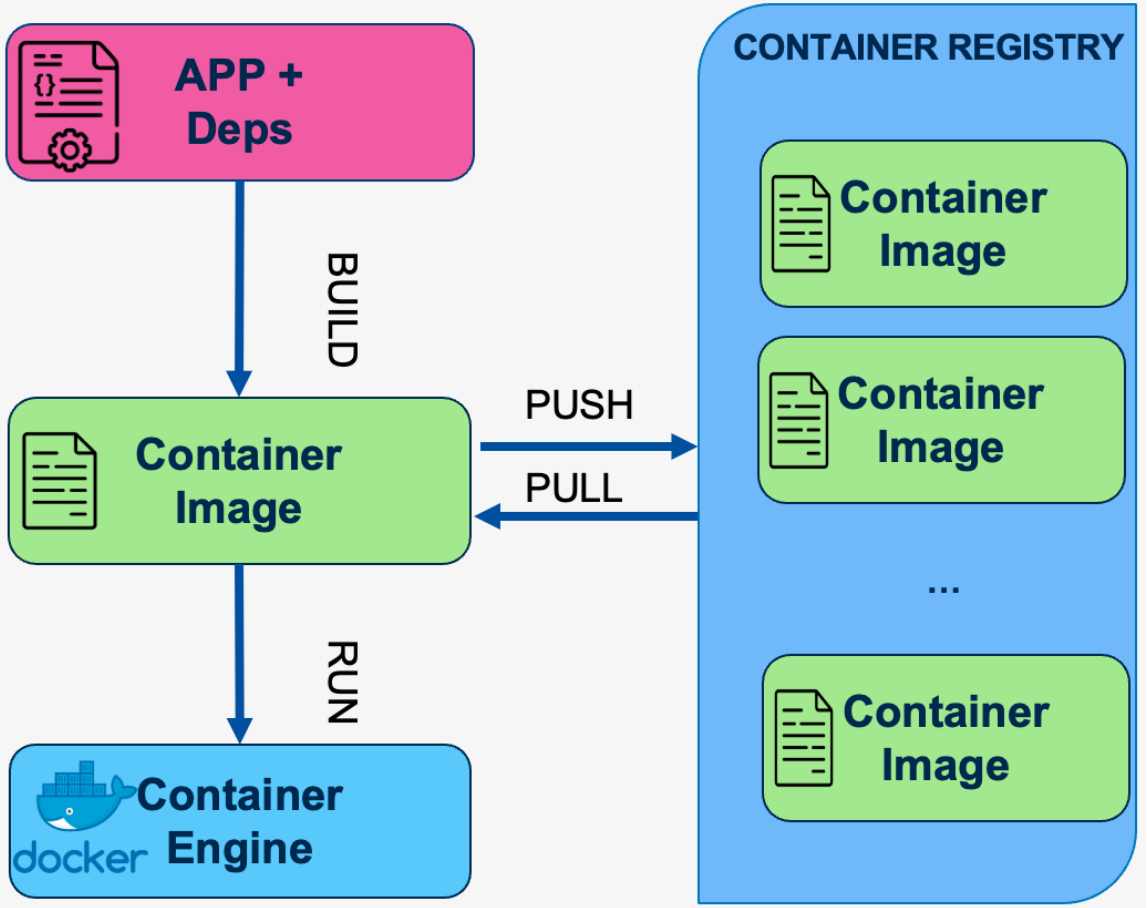
Containerized Application Context - Engines

- Container engines refer to the software components that enable the host OS to act as a container host
- A container engine accepts user commands to build, start, and manage containers through client tools (including CLI-based or graphical tools)
- *Blue-Cloud2026 uses Docker as Container Engine*



Containerized Application Context – Container Registry

- Container images can be shared with others via a public or private container registry
- Blue-Cloud uses Docker hub* as Registry



Integrating Applications in the Blue-Cloud VRE - How

To delegate to Blue-Cloud VRE not only the hosting

- Authentication & Authorization
- User Roles Mgmt.
- Auditing / Accounting
- Metrics / Monitoring

KEYCLOAK

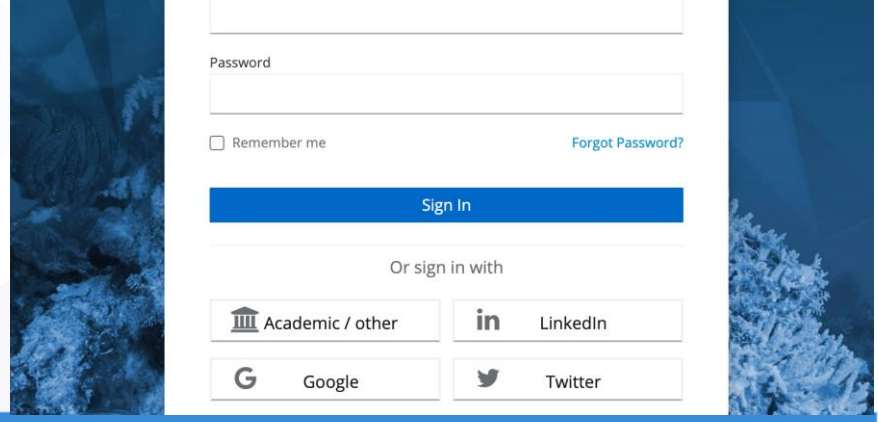
NGINX

Grafana

Prometheus

gCUBE FRAMEWORK

- Blue-Cloud VRE Identity and Access Management (IAM) Service
  - Single-Sign On & Identity Brokering and Social Login
  - *Integration via Standard Protocols*, support for OpenID Connect, OAuth 2.0, and SAML.

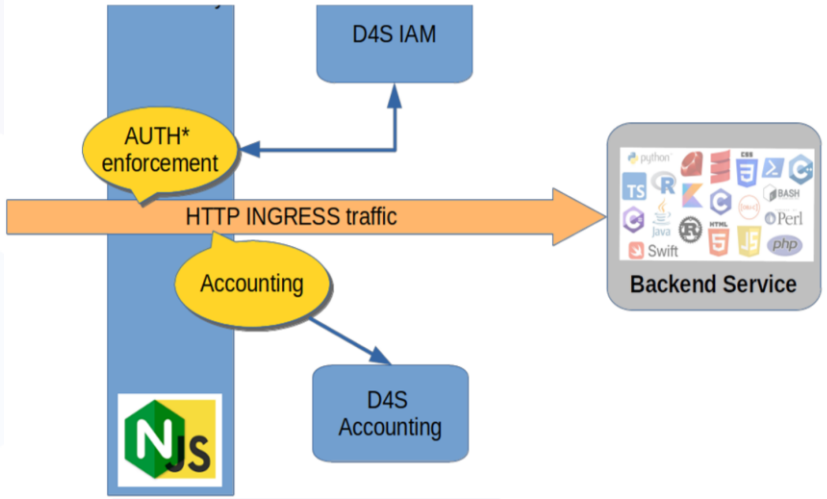


Integrating Applications – Auditing / Accounting

- Blue-Cloud uses D4Science Smart Proxy Tech.

NGINX based SmartProxy (nginx + js)

- Sending Auditing/Accounting information to the infrastructure Service
- Can be extended with few lines of JavaScript code to support specific needs (e.g. HTTP header extraction, parsing and adaption of custom tokens)
- Also employed in authentication and authorization enforcement



Integrating Applications – Monitoring / Metrics

- *Prometheus* is an open-source system monitoring and alerting toolkit
  - a multi-dimensional data model with time series data identified by metric name and key/value pairs
  - time series collection happens via a pull model over HTTP;
  - PromQL, a flexible query language to leverage this dimensionality;
- The multiple modes of graphing and dashboarding has been exploited by adopting *Grafana*, which allow us to query, visualise, alert on and understand Prometheus data on metrics.

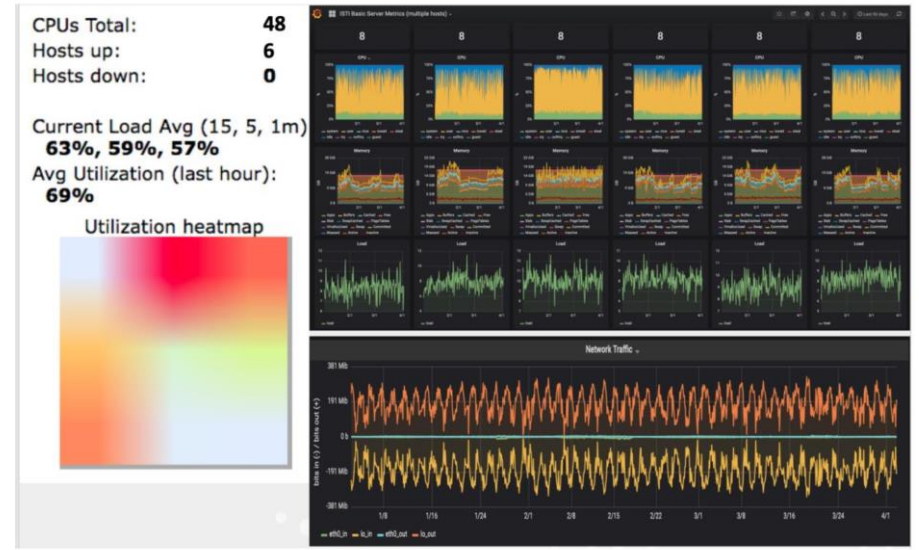
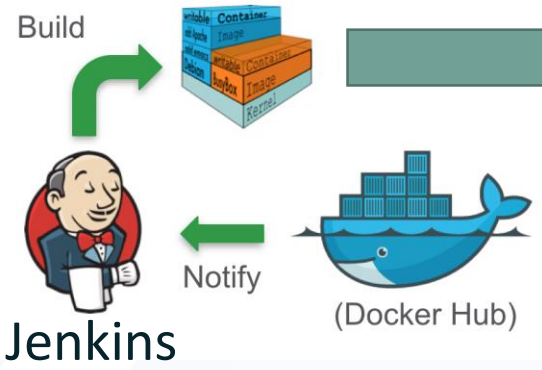
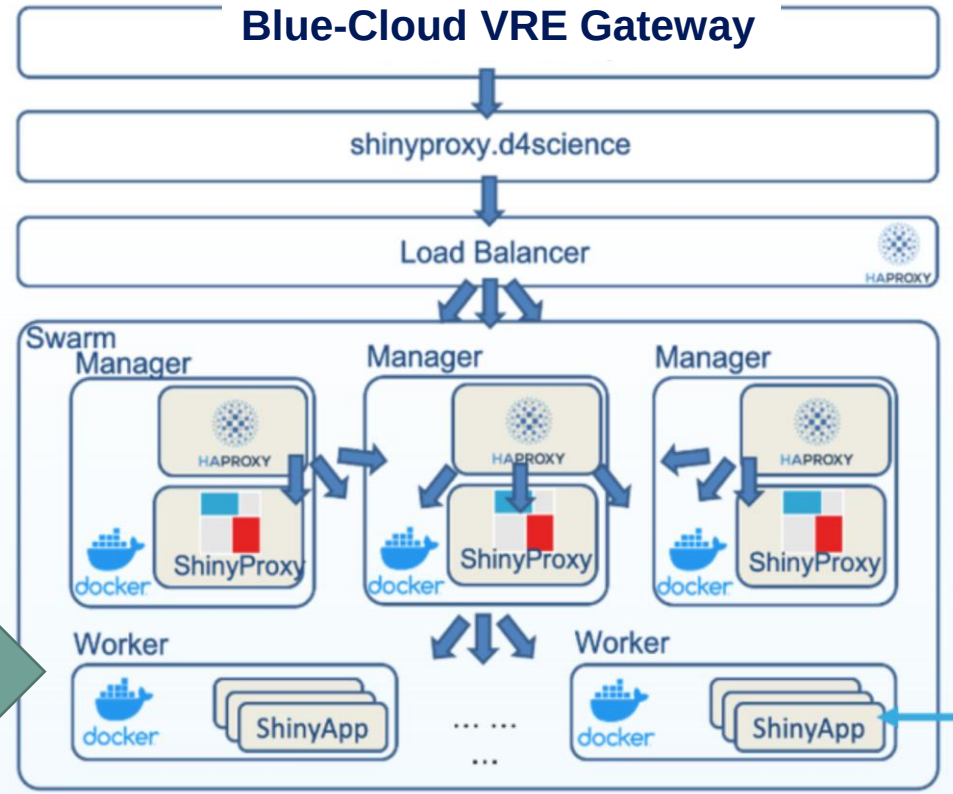


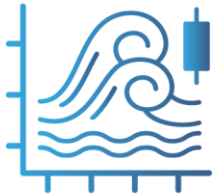
Figure 12 Prometheus Aggregated View via Grafana for the Auditing Cluster

- **Automated Deployment processes**
  - It can be a public app available in Docker Hub or any other public container registry.
  - The image name and the run command are the only requirements.



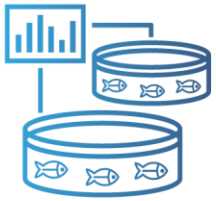


5 Virtual Labs at the VRE and more to come ...



Marine Environmental Indicators

Zoo & Phytoplankton EOV products

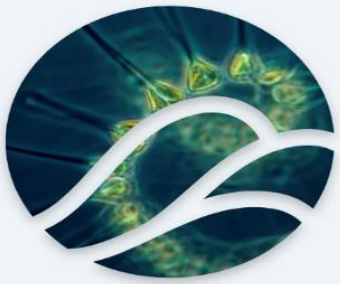


Aquaculture Monitor

Fish a matter of scales

Plankton Genomics

Vlabs for different domains



Biodiversity

Zoo and Phytoplankton EOV products

Genomics

Plankton Genomics

Environment

Marine Environmental Indicators

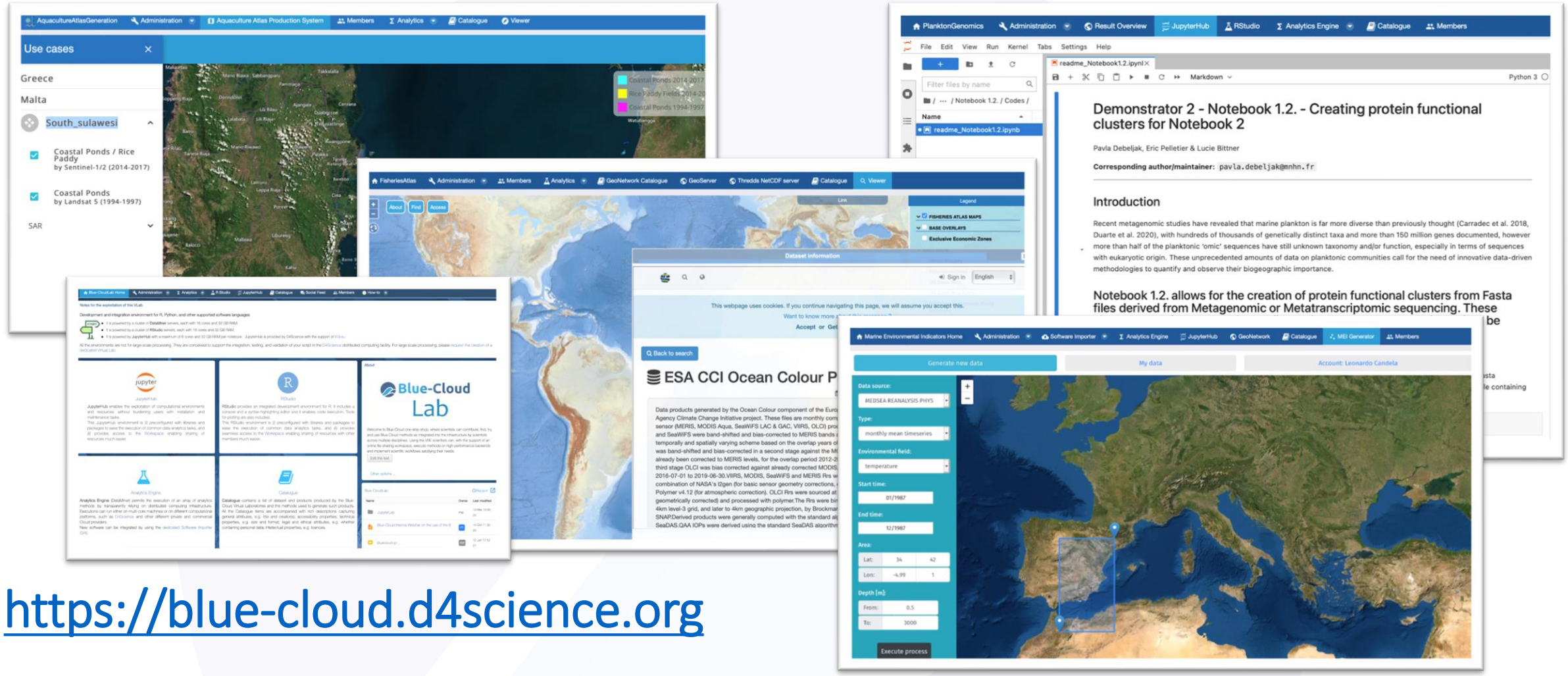
Fisheries

Global Record of Stocks and Fisheries

Aquaculture

Aquaculture Monitor

VLabs at-a-glance



https://blue-cloud.d4science.org

https://blue-cloud.d4science.org/web/marineenvironmentalindicators

Marine Environmental Indicators

The VLab provides support to analyse the quality of the marine environment, and inform decision makers about the good environmental status in a changing climate.

Partners:

cmcc Centro Euro-Mediterraneo sui Cambiamenti Climatici · Ifremer · KNMI · MERCATOR OCEAN INTERNATIONAL · UNIVERSITAS BERGENSIS

Data sources through Blue-Cloud:
Copernicus Marine Service, Copernicus Climate Service, EuroARGO, EMODnet

Main target users:
Environmental protection agencies and international stakeholders involved in environment management

Services introduction:
The VLab offers a web user interface and several scientific-based algorithms, that can be used to obtain environmental indicators and added-value data applying big data analysis and machine learning methods on multi-source data sets.

UN SDGs addressed

13 CLIMATE ACTION

14 LIFE BELOW WATER

SERVICES

THANK YOU

eosc | Blue-Cloud2026

blue-cloud.org @blueclouddeu blue-cloud org

Massimiliano Assante

Follow @maxassante

Funded by the European Union