



eosc | FAIR-EASE

Virtual Research Environment : process data as FAIR as possible

What's Galaxy ? How do we use it ?

Marie Jossé - CNRS

GALAXY & FAIR

INTERNET PLATFORM FOR SHARING AND
PROCESSING RESEARCH DATA

4 KEY POINTS :

- Accessibility
- Reproducibility
- Transparency
- Community work





Peer review

- Each tool must be pull request on a github repo
- It goes through some automatic checks
- Then, it's verified, tested and improved in collaboration with the galaxy community

A screenshot of a GitHub pull request page for 'Odv #166'. The page shows the repository 'usegalaxy-eu / galaxy' and indicates that the pull request is 'Merged'. A comment from 'Marie59' dated 'Jun 7' reads: 'Hi @bgruening ! I just had the approval of ODV's author to put ODV interactive tool on galaxy. So, I initiate this PR to add the tool to the successful ones from the Lille training. Tell me if there is anything to change.' Below the comment is a list of 23 commits, including 'EU: LOCALONLY Fix FTP URL in UI', 'EU: Favico', 'EU: Expand velvet kmer values', 'EU: LOCALONLY fix loading of tools', 'EU: Add local phinch site in lieu of GIE', 'EU: add SEEK data source', and 'EU: Adding support for juicer hic and test data for it (#49)'. On the right side, there are sections for 'Reviewers' (listing 'bgruening'), 'Assignees' (listing 'No one assigned'), and 'Labels' (listing 'None yet').

FAIR

Tool name, id,
description, ...



Open source
platform for
everyone



One tool for
multiple kind of
datasets



To be chained in
different
workflows



Add tools on Galaxy



Atomisation

Divided in elementary
bricks

Generalisation

Standardised tools

Rigorous

Peer review

```
<tool id="te_productivity" name="Trends earth productivity" version="0.0.0" profile="tool"
<description>calculation for the 1.5.1 SDG indicator.</description>
<requirements>
  <requirement type="package" version="1.24.3">numpy</requirement>
  <requirement type="package" version="2.1.14">trends_earth_schemas</requirement>
  <requirement type="package" version="2.1.14">trends_earth_algorithms</requirement>
</requirements>
<required_files>
  <include type="literal" path="productivity.py"/>
</required_files>
<command detect_errors="exit_code">
python
  '$__tool_directory__/te_productivity.py'
  '$__tool_directory__/_init_.py'
  '$__tool_directory__/productivity.py'
  '$input_data'
  '$output_geojson'
]]&gt;
&lt;/command&gt;
&lt;inputs&gt;
  &lt;param name="input_data" type="data" format="json" label="Input a json file" /&gt;
&lt;/inputs&gt;
&lt;outputs&gt;
  &lt;data name="output_geojson" from_work_dir="productivity.json" format="json" /&gt;
&lt;/outputs&gt;
&lt;tests&gt;
  &lt;test expect_num_outputs="1"&gt;
    &lt;param name="input_data" value="sample_raw_job.json" /&gt;
    &lt;output name="output_geojson"&gt;
      &lt;assert_contents&gt;
        &lt;has_text text="E6P" /&gt;
      &lt;/assert_contents&gt;
    &lt;/output&gt;
  &lt;/test&gt;
&lt;/tests&gt;
&lt;help&gt;<![CDATA[
=====
Productivity indicator
=====

**What it does**

**Input description**

One input file in json format.</pre></div><div data-bbox="530 130 870 200" data-label="Section-Header"><h1>Standardisation</h1></div><div data-bbox="530 260 810 300" data-label="Section-Header"><h2>AN XML FILE FOR EACH TOOL</h2></div><div data-bbox="540 360 930 590" data-label="List-Group"><ul><li>• Each script (python, R, java ...) is linked to a “wrapper the xml file where some necessary info must be written.</li><li>• Everything is available on a github repository</li></ul></div><div data-bbox="495 675 855 985" data-label="Complex-Block"><img alt="Screenshot of a Galaxy tool interface for 'Calculate community metrics'."/><p>The screenshot shows the Galaxy tool interface for "Calculate community metrics". The tool title is "Calculate community metrics calculate community metrics from abundance data (Galaxy Version 0.0.2)". There are buttons for "Versions" and "Options". Under "Input file", there is a file selection area with a message "No tabular dataset available." and a description: "Observation data file, with location, year, species and abundance." Below this is a section "Choose the community metrics you want to compute" with a "Select/Unselect all" checkbox and a list containing "x All". A note states: "Presence/absence, Species richness, Simpson and Shannon index are systematically computed." At the bottom is an "Execute" button.</p></div><div data-bbox="865 800 1000 1000" data-label="Image"><img alt="A 3D illustration of a gift box with a pink ribbon."/>A 3D illustration of a gift box with a pink ribbon. The box is dark blue with white outlines, and the ribbon is a vibrant pink color, tied in a bow on top.</div>
```


Atomisation

CREATE WORKFLOWS FROM ELEMENTARY BRICKS

The screenshot displays the Galaxy workflow editor interface. The top navigation bar includes 'Galaxy Europe' and various utility icons. A left sidebar lists biological categories such as 'Immunoproteins and antigens', 'Model organisms', and 'Genetics'. The main workspace shows a workflow titled 'Coastal Water Dynamics' with five steps: 1. Model (output: netcdf), 2. Observations (output: netcdf), 3. Interactive Source Notebooks (action: Include data into the environment, input: Netcdf file source), 4. ODV (input: Netcdf file, output: outputs), and 5. Interactive DIVAnd Notebooks (action: Include data into the environment, input: Netcdf file diva). A right-hand metadata panel provides details for the workflow, including its name, version (7: Sep 11th 2023, 5 steps), annotation ('Explore, analyse, visualise coastal data.'), license (Creative Commons Attribution 4.0 International), creator (Marie Josse and Fair-Ease), and tags.



Allows better understanding of an analysis

Smaller scripts easier to read

Allows researchers to combine tools

Can combine tools to have a full operational workflow

Allows to create different kind of analysis

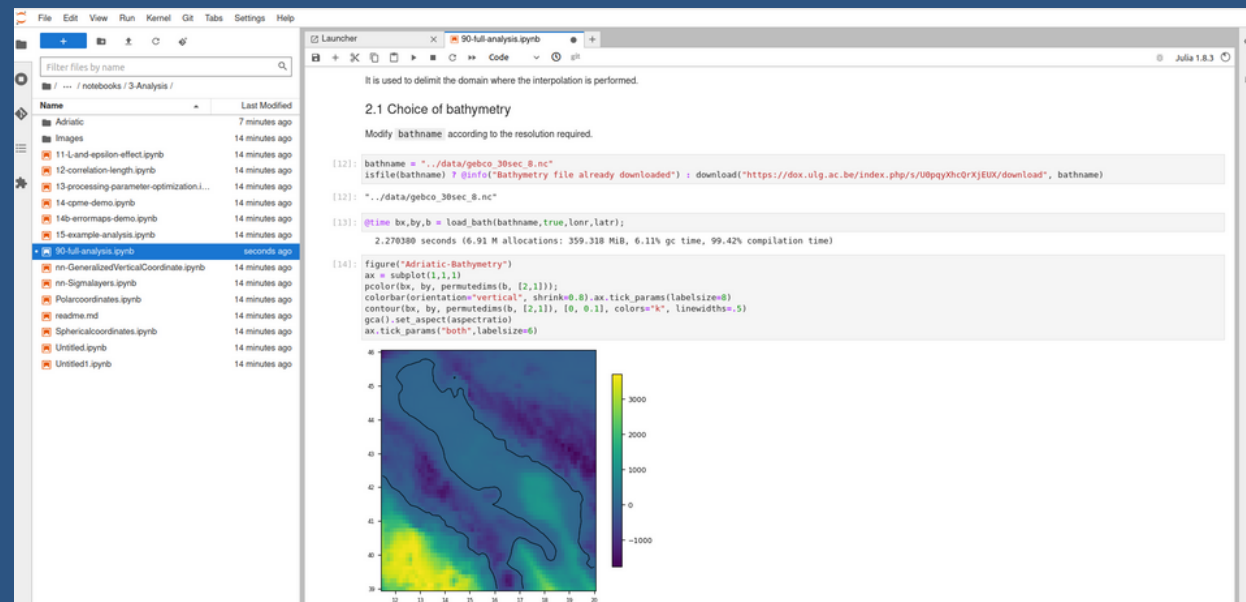
Can use the tool with different ones and have multiple workflows

Atomisation

WHAT ARE THE
ELEMENTARY BRICKS

Workflows

Can be shared, published, runned, ...

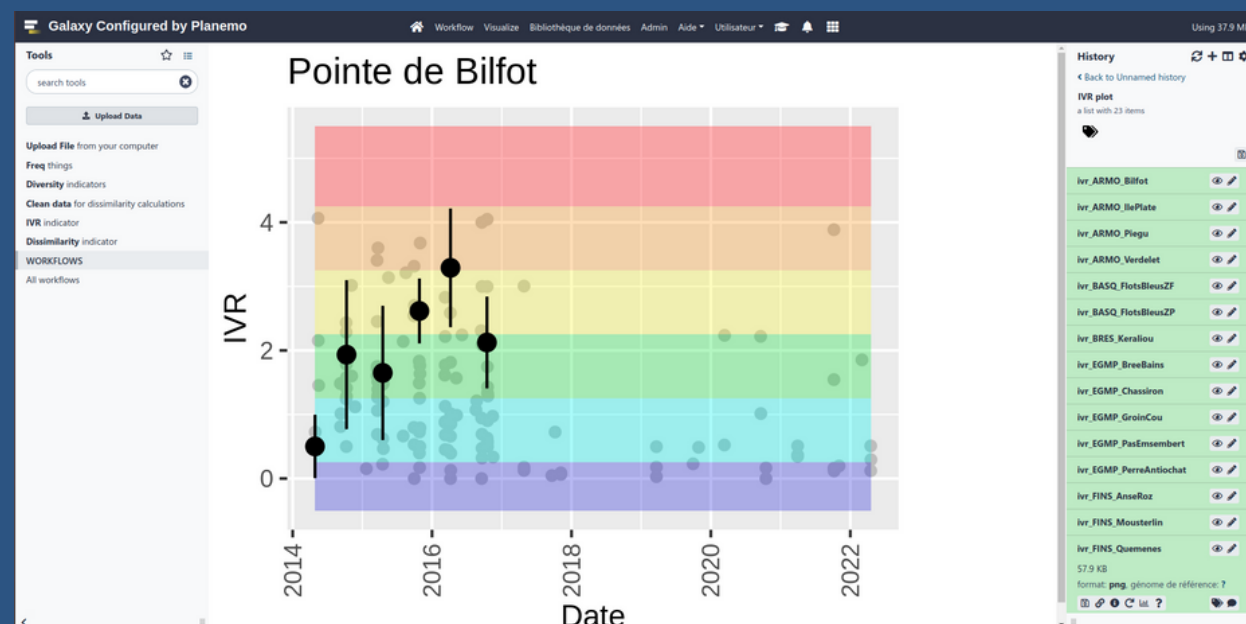


Interactive tools

Workflows with human in the loop : jupyterlab, desktop applications, ...

Non-interactive tools

Fully automatic workflows : click button tools without any knowledge in programmation

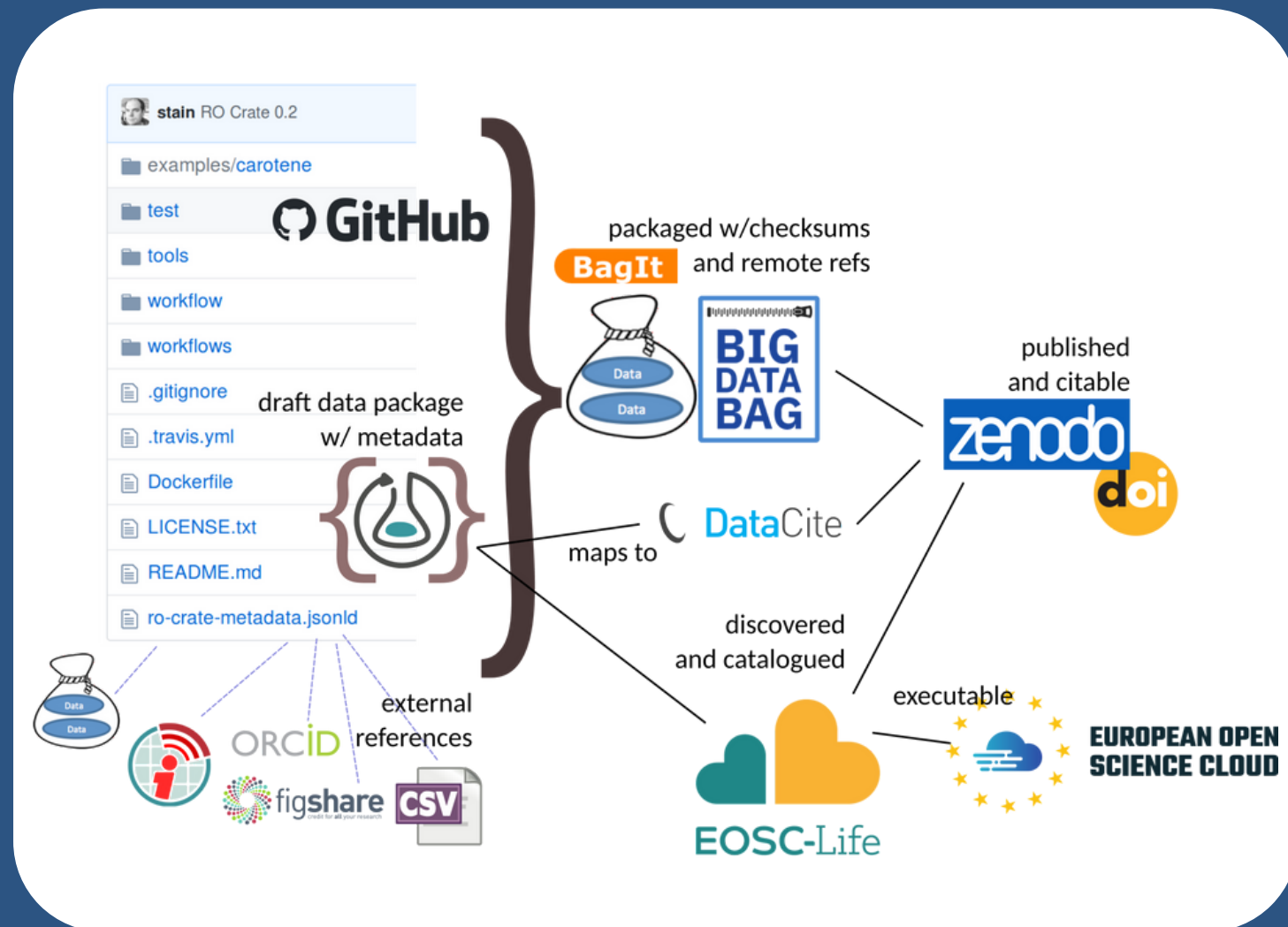


RO-Crate

ENABLE FAIR AND
REPRODUCIBLE DATA ANALYSIS

Galaxy as an excellent start point

Stores the complete analytical workflow, including its metadata. The model involves the history entity, including all steps performed in a specific analysis, and the workflow entity, defining the structure of an analytical pipeline.



Integration of the FAIR Digital Object principles

Integrating RO-Crate, Galaxy enables a standardised exchange of FDOs (workflow definition, single datafile, datasets collection or entire history analysis) with other platforms such as WorkflowHub improved Research Data Management (RDM)



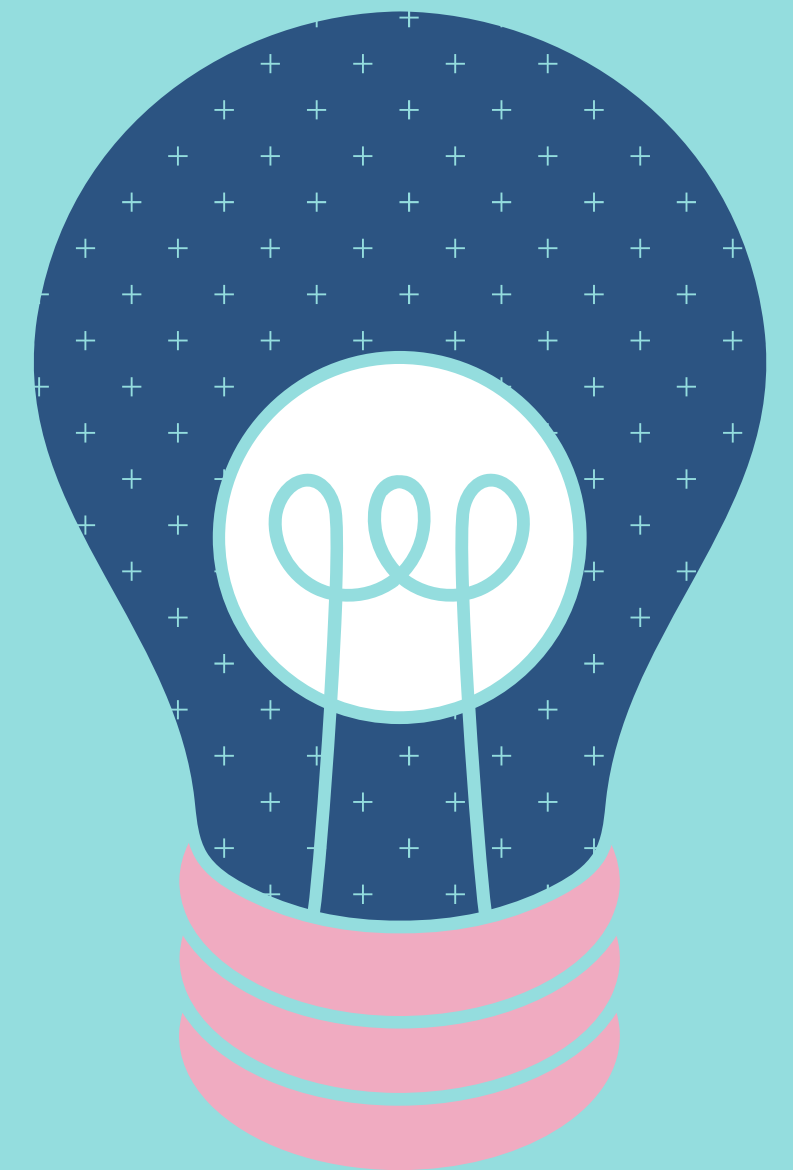
What you need to keep in mind !



Galaxy is an open-source platform for FAIR data analysis that enables users to:

- Use **tools** from various domains (that can be plugged into workflows) through its graphical web interface.
- Run code in **interactive environments** (RStudio, Jupyter...) along with other tools or workflows.
- Manage data by **sharing and publishing results, workflows, and visualizations**.
- Ensure **reproducibility** by capturing the necessary information to repeat and understand data analyses.

The Galaxy Community is actively involved in helping the ecosystem improve and sharing scientific discoveries.

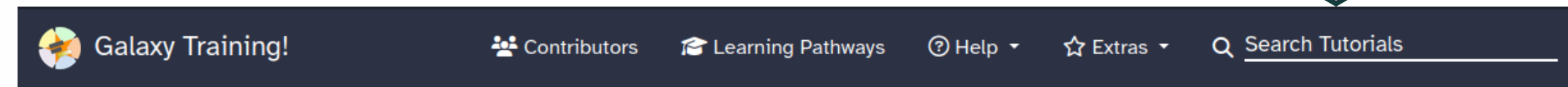
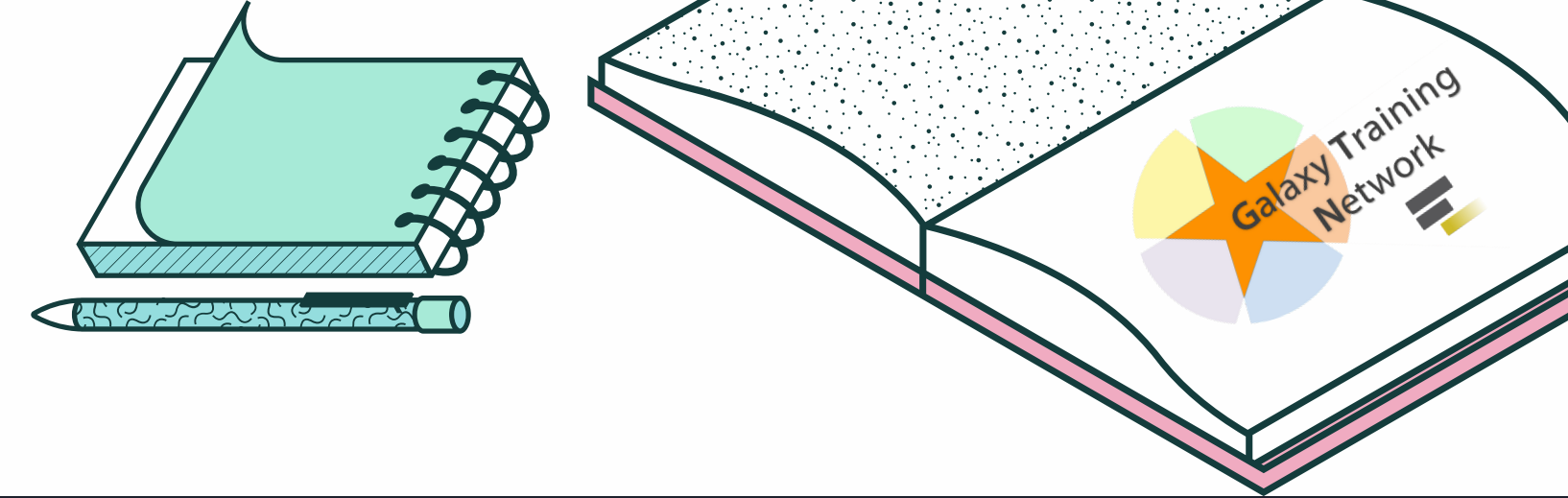


To go further : Galaxy Training Network

- Tutorials
- Classes and courses
- Pathways



An easy way to learn galaxy and improve your skills on various domains for instance a set of tuto are available on FAIR management



FAIR Data, Workflows, and Research

These lessons will teach you how to make your research objects more FAIR with practical, hands-on advice.

You can view the tutorial materials in different languages by clicking the dropdown icon next to the slides (📄) and tutorial (📖) buttons below.

Material

FAIR Data Management

The FAIR (Findable, Accessible, Interoperable, Reusable) data stewardship created the foundation for sharing and publishing digital assets. These lessons will apply to machine accessibility and emphasize that all digital assets should share data in a way that will enable maximum use and reuse.

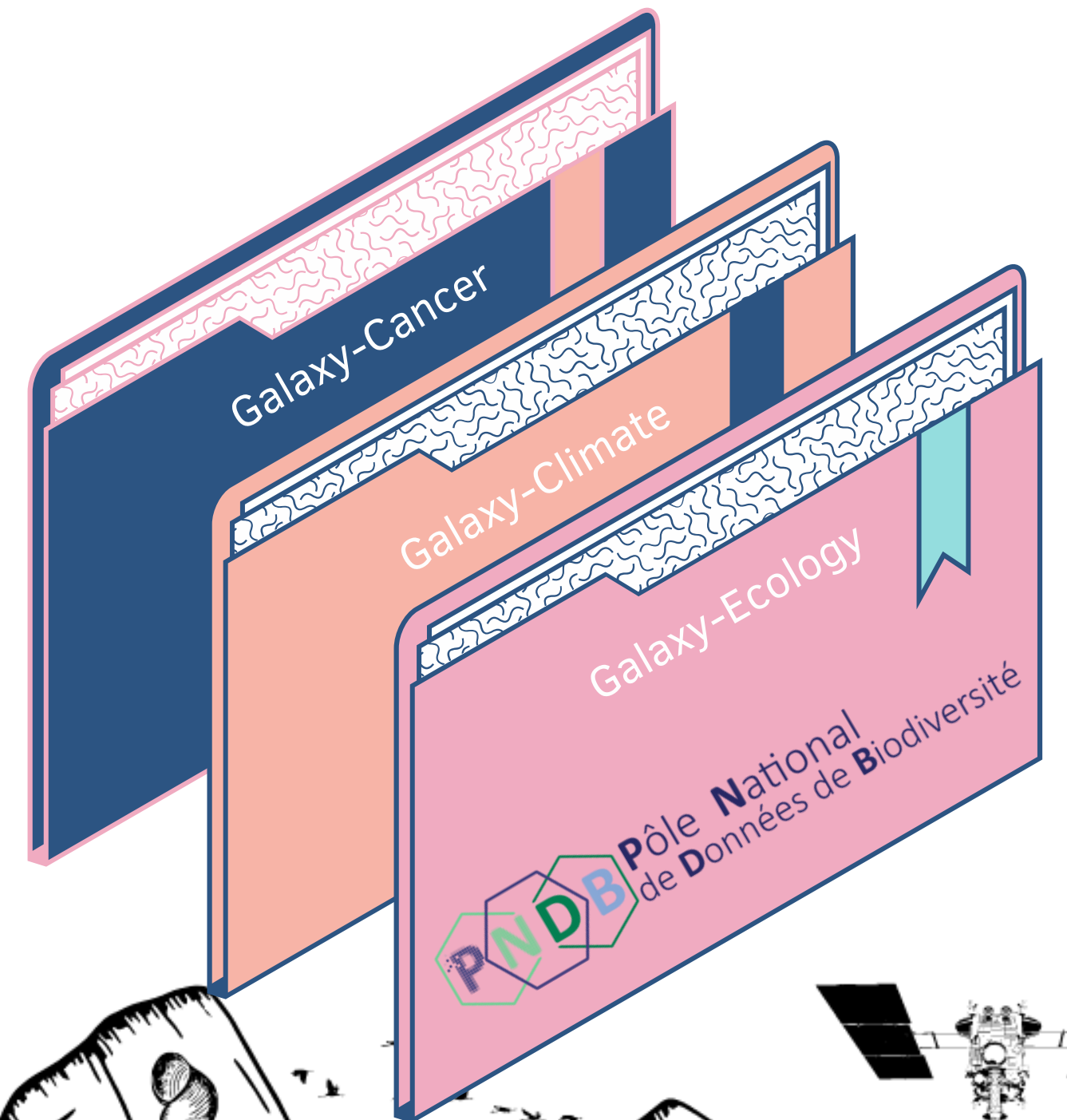
Lesson	Slides	Hands-on	Recordings
FAIR in a nutshell fair open data stewardship			📄 ▼
FAIR data management solutions fair dmp data stewardship			📄 ▼
FAIR Galaxy Training Material fair gtn training			📄 ▼

FAIR Data Management

Galaxy-Ecology

Multiple sub domains for either a research theme or for region of the world (Africa, India, ...)

Galaxy for Ecology – a web platform to get, process, analyze and visualize ecological data



CC BY 4.0 Olivier Norvez



5 PILOTS FOR MOCKUP AN EAL EARTH SYSTEM

- Atmospheric and oceanic circulation and thermodynamics
- The biological and chemical processes that feedback on to the physics of climate
- Grid over the surface of the Earth and underneath the surface of the oceans

Earth and Environmental Dynamics



Coastal Water Dynamics : focuses on the coastal marine environment near river estuaries, where important processes take place.



Earth Critical Zone : monitors land and soil degradation.



Volcano Space Observatory : monitors global volcanic activity, allowing the focus on any major volcanic eruption worldwide

The Environmental BioGeochemical Asset

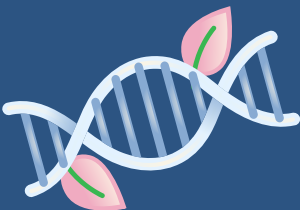
Ocean Bio-Geochemical Observations: addresses fundamental scientific questions regarding the health of marine ecosystems (e.g. ocean acidification, ...) and needs for ocean resource management.



Biodiversity Observation

Marine Omics Observation:

analyses of spatial- and time-comparable marine microbial metagenomics data sets for the exploration of biodiversity and its correlations with environmental quality



A new sub-domain : Galaxy - Earth System

NEW TOOLS, NEW WORKFLOWS, NEW TUTORIALS AND NEW DATA ACCESS

An environment for each subject to access and process their data

Galaxy | Europe

earth-system.usegalaxy.eu

Applications YouTube Music Zimbra: Réception gmail Mnhn Google Drive MyServices A generic remote s... An introduction to... pndb | Gather

Galaxy Earth System Europe

Workflow Visualize Shared data Help User

Tools

search tools

Upload Data

Get Data

Send Data

Collection Operations

GENERAL TEXT TOOLS

Text Manipulation

Filter and Sort

Join, Subtract and Group

Convert Formats

Coastal Water Dynamics

Earth Critical Zone

Volcano observations

BioGeoChemical

Marine Omics

The first interdomain digital architecture for integrated use of environmental data

eosoc | FAIR-EASE

The Environmental BioGeochemical Asset

Earth and Environmental Dynamics

Biodiversity Observation

History

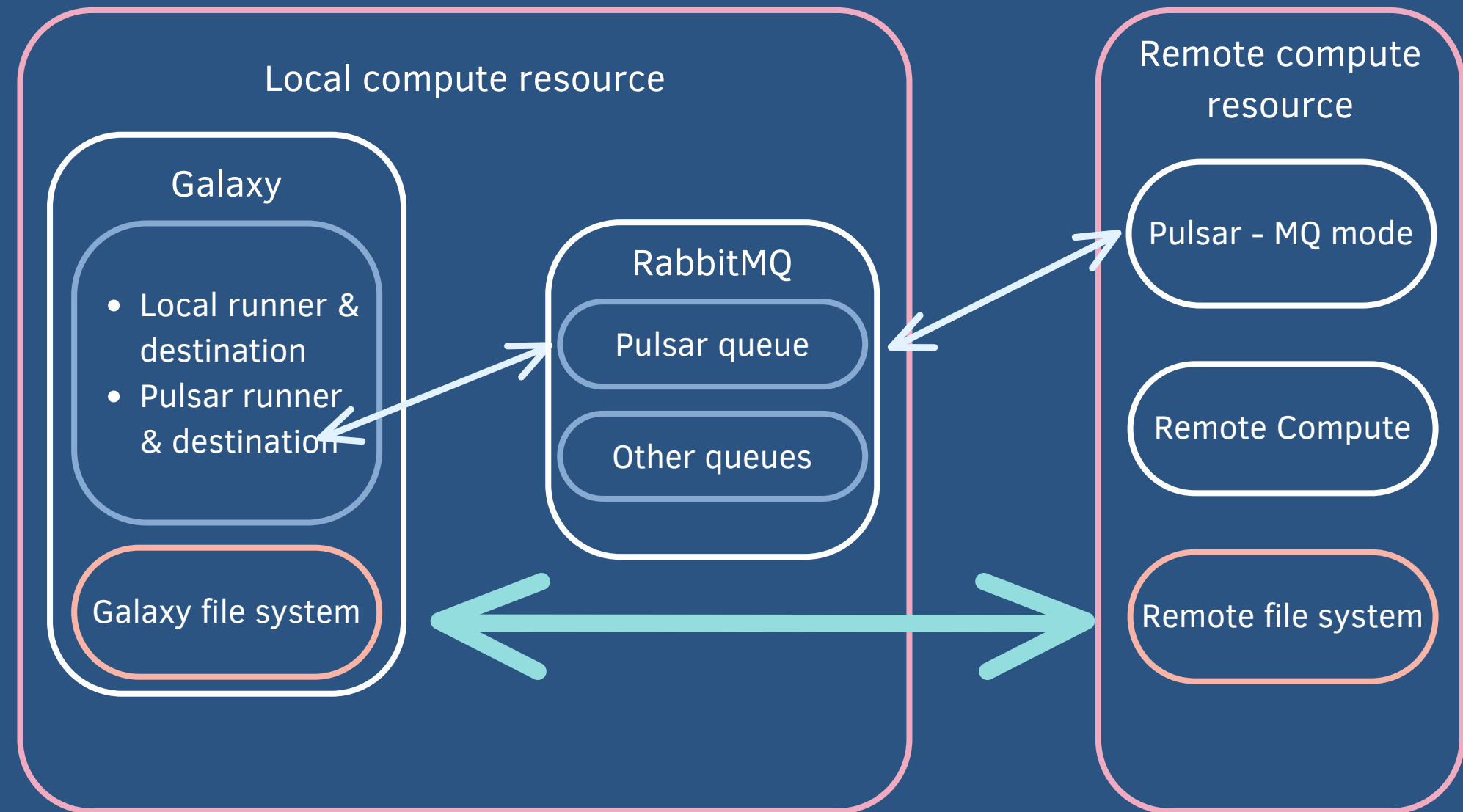
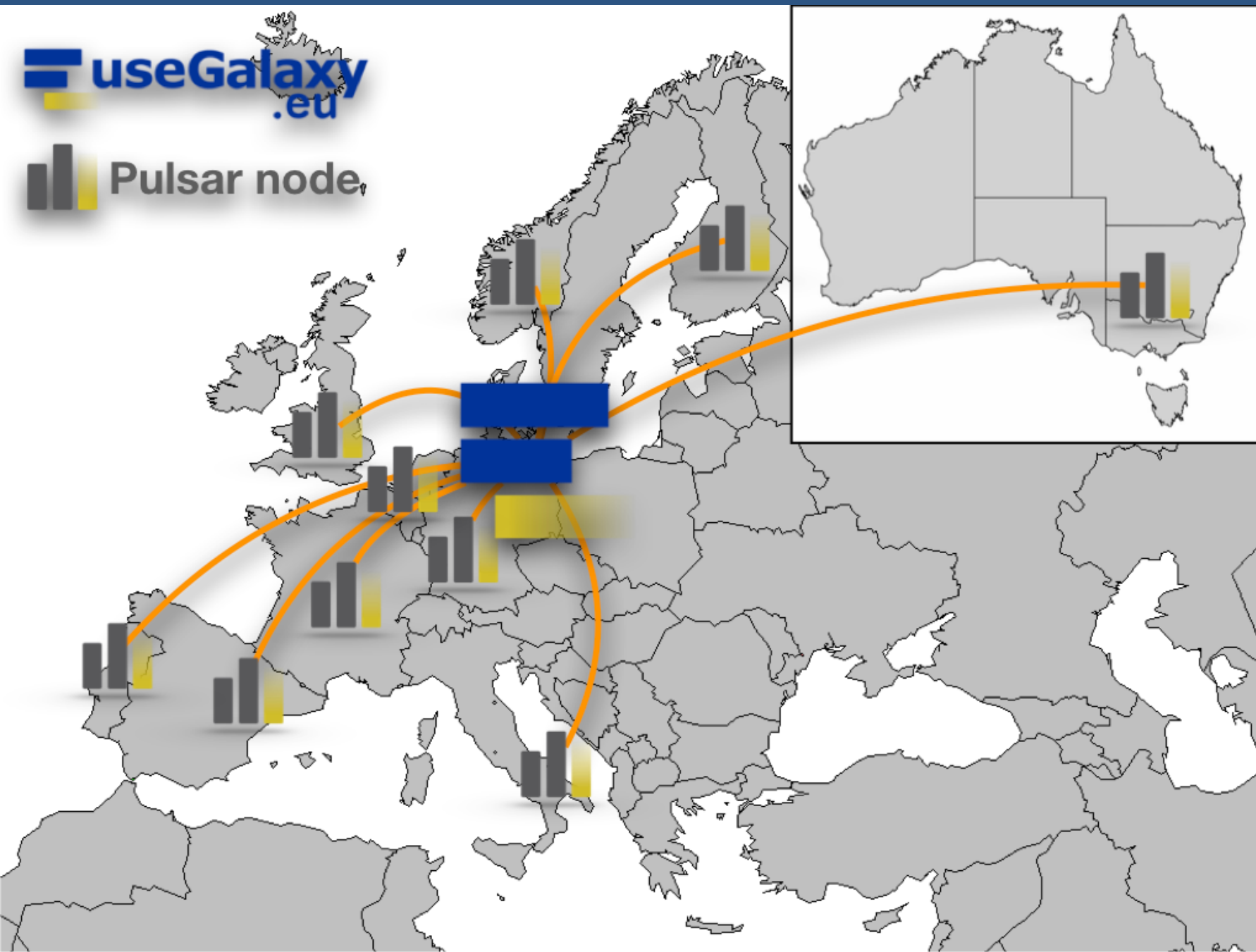
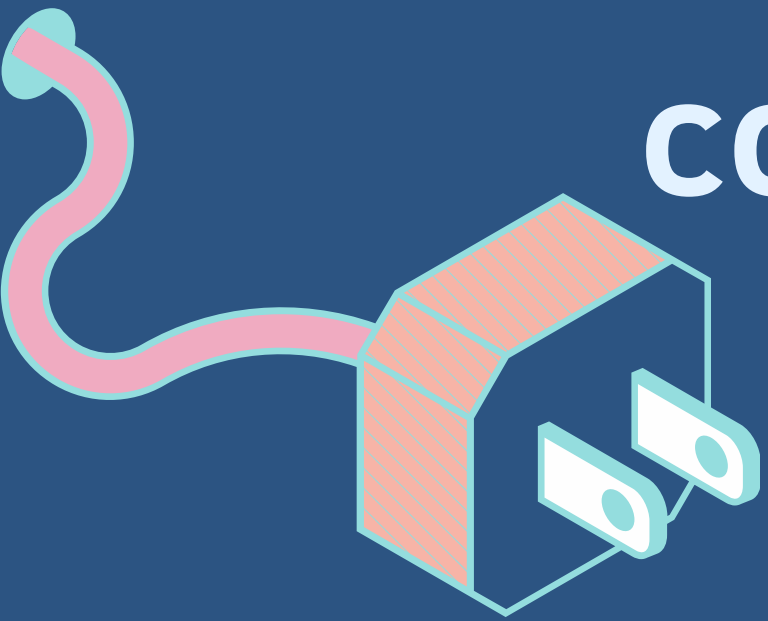
Rechercher des données

Unnamed history

This history is empty. You can load your own data or Charger des données depuis une source externe.

Galaxy EUROPE

Galaxy deals with remote compute resources and remote files system



↔ AMQP messages

↔ File transport (curl)

How to deploy Galaxy

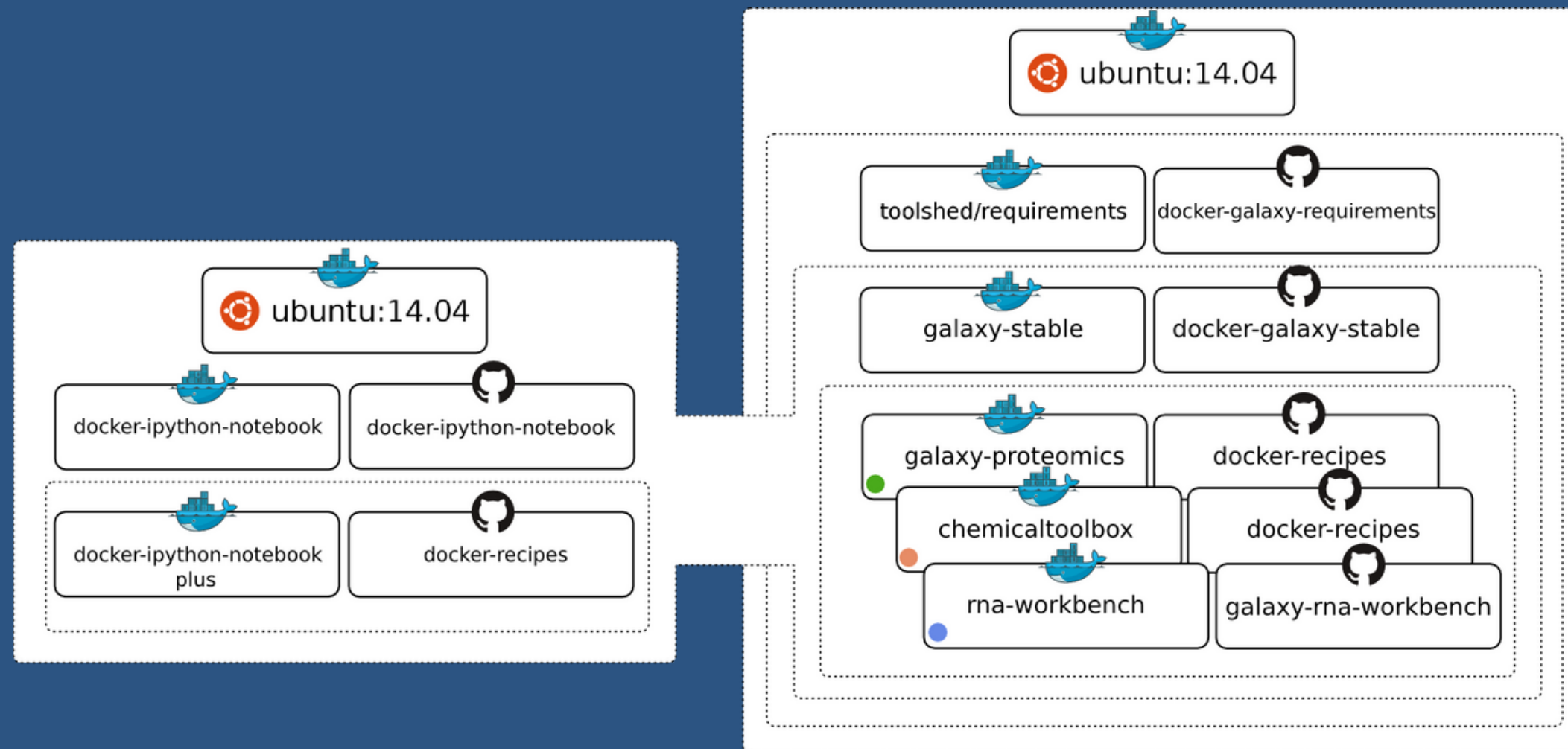
Ansible components to automate almost every aspect of Galaxy installation and maintenance.

Ansible is an advanced configuration management system.

These **playbooks** are used to maintain Galaxy main, cloud and **Docker** images, virtual machines, ...



ANSIBLE



High quality **Docker** containers for stable Galaxy environments.

Releases corresponding to each new version of Galaxy.

Many flavors available.

Galaxy - Earth System as a new service of D4Science




Blue-Cloud2026
A federated European FAIR and Open Research Ecosystem for oceans, seas, coastal and inland waters




[Home](#) [Explore](#) [Catalogue](#) [Sign In](#)

D4Science Labs a series of free-to-use applications to generate new knowledge from data comprising support for tabular data validation, data enrichment, and efficient analytical tools.




AnalyticsLab

[Access this VRE](#) [Info](#)



CollabResearch

[Access this VRE](#) [Info](#)



RStudioLab

[Request Access](#) [Info](#)




Galaxy EUROPE

EarthSystemGalaxy


[Access this VRE](#) [Info](#)

Marine Labs a series of free-to-use applications to generate new knowledge from biotic and abiotic marine datasets comprising support for data validation, data enrichment, and efficient analytical tools.



AlienAndInvasiveS...

[Access this VRE](#) [Info](#)




BiodiversityLab

[Request Access](#) [Info](#)



EnvironmentalData...

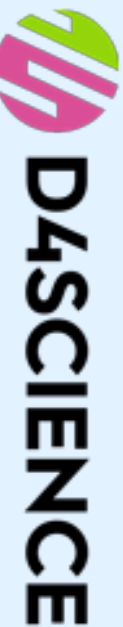
[Access this VRE](#) [Info](#)

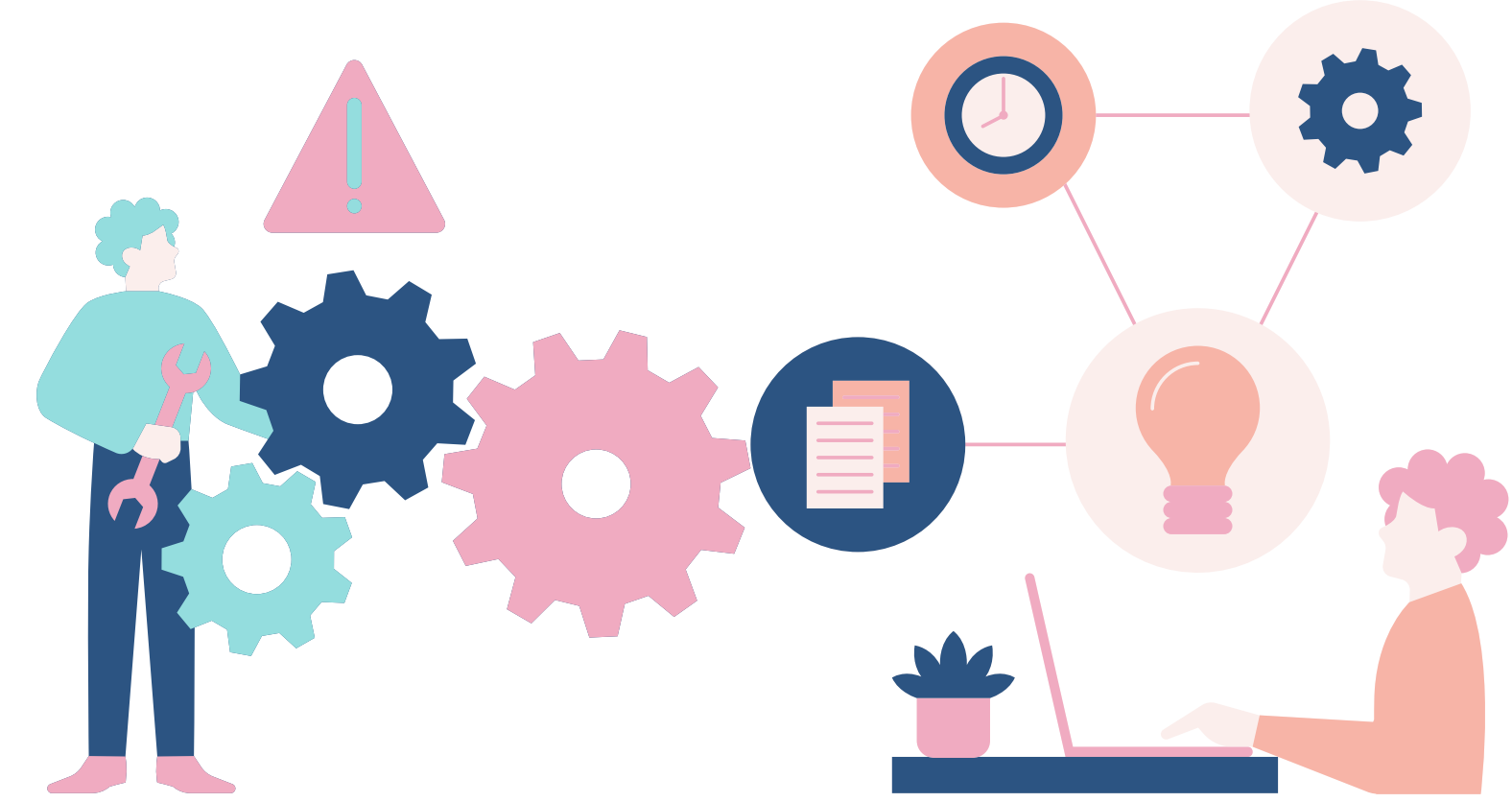
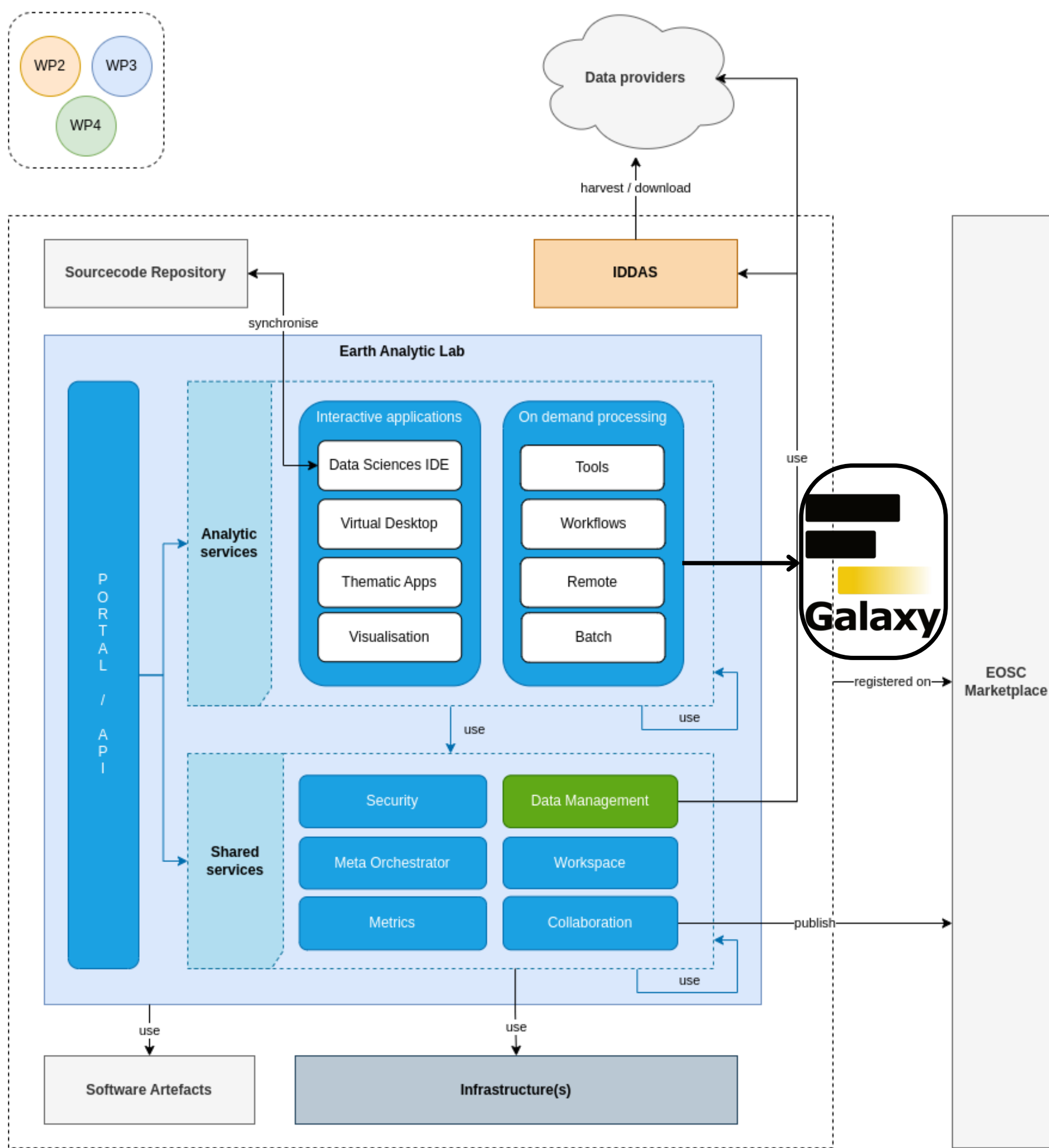


RPrototypingLab

[Request Access](#) [Info](#)

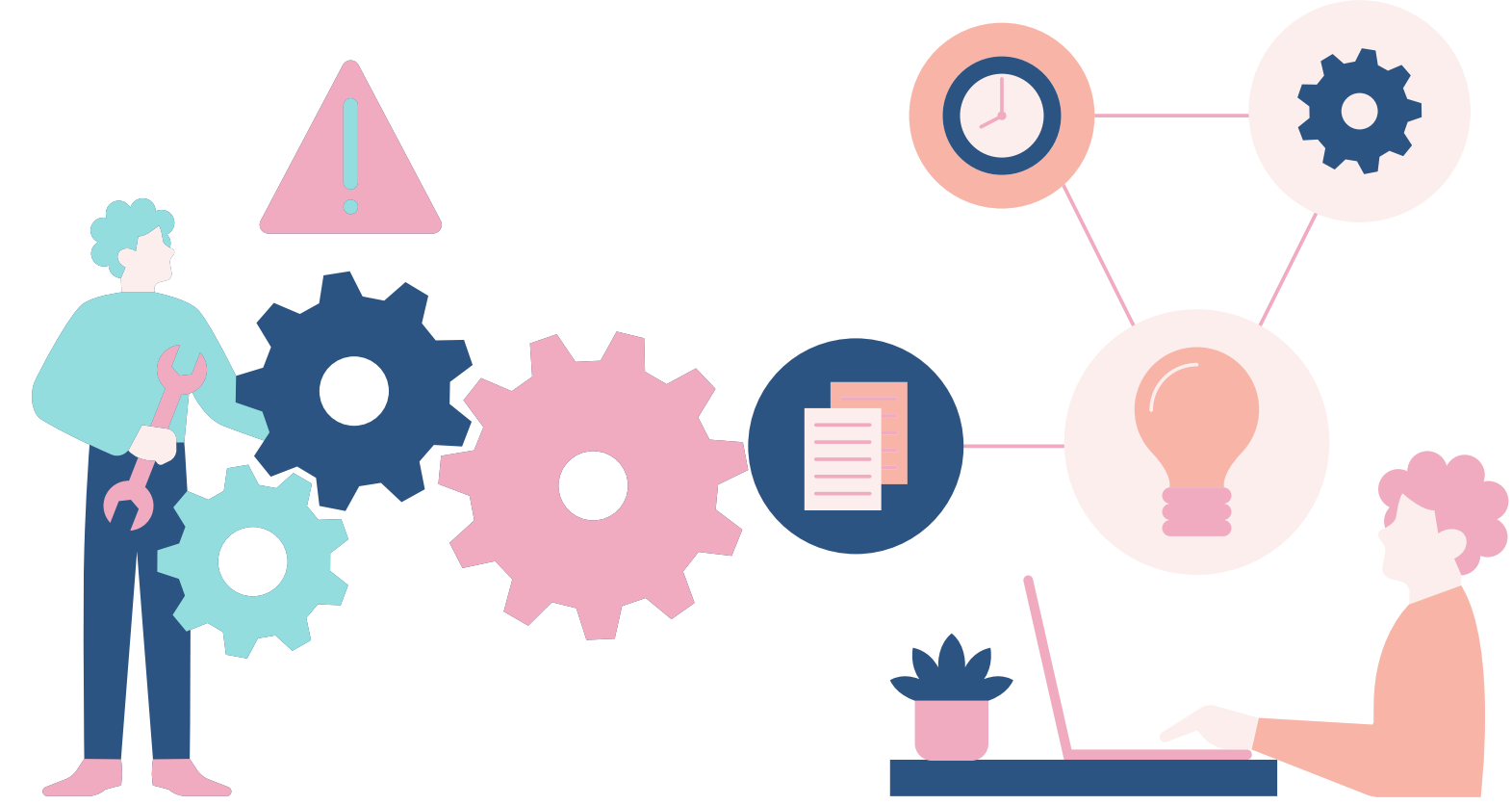
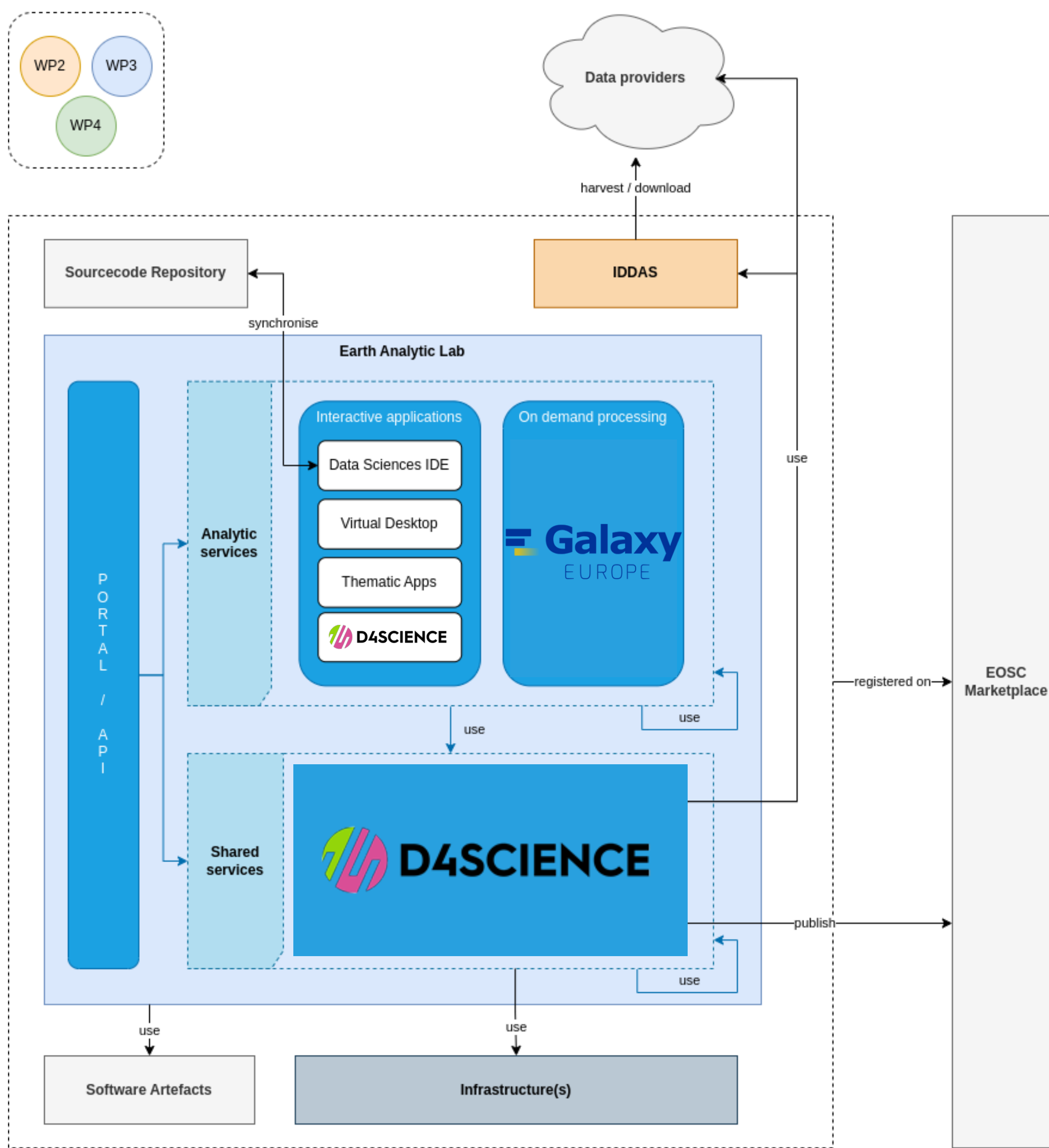
[Cookie Settings](#)





Propose an EAL as a D4Science virtual lab

- An easy way to visualise, analyse and process environmental and biodiversity data on-demand
- Improve data access both in terms of data harmonisation and in terms of technical efficiency of data access.
- Galaxy a main component of the EAL

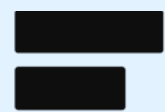


Propose an EAL as a D4Science virtual lab

Do you have any questions?

Send it to us!

marie.josse@ifremer.fr



Some use full links:



Galaxy <https://usegalaxy.eu/>



<https://training.galaxyproject.org/>

