

27-29 May 2024 

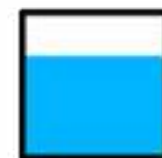


imdis

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



MARIS



National
Oceanography
Centre



The **GALAXY** platform as a
ground breaking **FAIR** tool
for **EARTH SYSTEM's**
analytical **WORKFLOWS**

Marie Jossé



DATA
TERRA

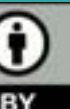
Jérôme Detoc
Erwan Bodéré



Ifremer



Funded by
the European Union



BY

OVERVIEW

01

Fair-Ease

02

The Earth System

03

Why Galaxy ?

04

Good practices

05

Our tools and workflows

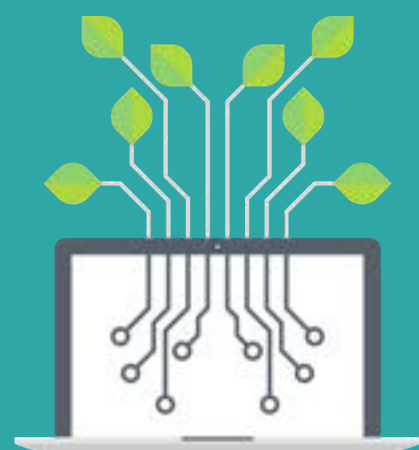
06

Try it !



FAIR-EASE

Building an interdomain digital architecture for distributed and integrated use of environmental data



FAIR-EASE Data Discovery
and Access
Interdisciplinary Service



FAIR-EASE Virtual
environments



THE EARTH SYSTEM

By Fair-Ease The Earth System is a complex and dynamic system that encompasses the interactions between the atmosphere, oceans, land, and biosphere.

Five pilots concerning applications for real-life science use-cases came with their specific needs in virtual research environments

Coastal Water Dynamics

Focuses on the coastal marine environment near river estuaries, where important processes take place.

Earth Critical Zone

Improve analytics model specific to the earth critical zone. Monitors land and soil degradation.

Volcano space observatory

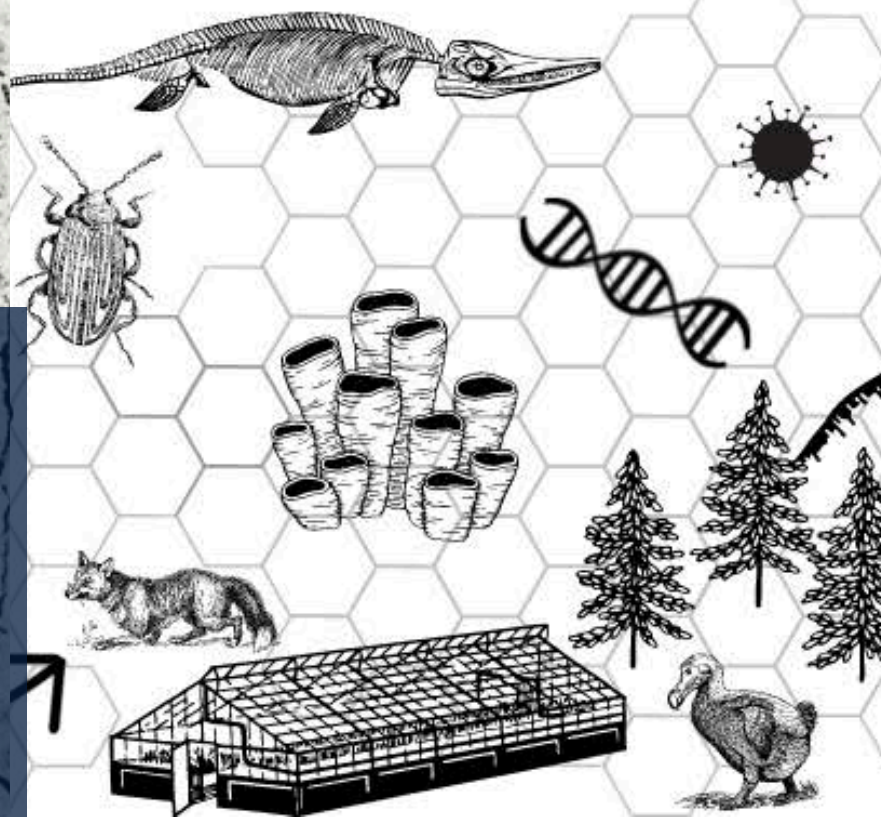
Monitors global volcanic activity, allowing the focus on any major volcanic eruption worldwide. Aggregating remote sensing observations from the solid earth and the atmosphere

Ocean Bio-Geochemical

Addresses fundamental scientific questions regarding the health of marine ecosystems (e.g. ocean acidification, ...) and needs for ocean resource management.

Marine Omics

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



WHY GALAXY

A community driven platform popular
in other domains

Reusing the existing : 2 subdomains of Galaxy
europe inspiring Fair-Ease, the climate and
ecology ones

FAIR by design


Opensource web platform for sharing and processing
research data

- Accessibility
 - Reproducibility
 - Transparency
 - Community work
-

Tools

Get Data
Send Data
Collection Operations
GENERAL TEXT TOOLS
Text Manipulation
Filter and Sort
Join, Subtract and Group
Convert Formats
GENERAL INTERACTIVE TOOLS
Interactive tools
EARTH AND ENVIRONMENTAL DYNAMICS
Water Coastal Dynamics
Earth Critical Zone
Volcano
ENVIRONMENTAL BIO-GEOCHEMICAL ASSETS
Bio-geochemical
BIODIVERSITY OBSERVATION
Marine Omics
WORKFLOWS
All workflows

Welcome to Galaxy for Earth System and environment



Galaxy for Earth System and environment was implemented within the project [Fair-Ease](#). It's a virtual platform to process, analyse and visualize Earth System, Environment and Biodiversity data. It is based on the [Galaxy framework](#), which guarantees simple access, easy extension, flexible adaption to personal and security needs, and sophisticated analyses independent of command-line knowledge.

Content

History

Ocean's variables
#training
206 MB 21

- 35: Interactive JupyterTool and notebook on data 32
- 34: Executed JupyterTool Notebook
- 33: JupyterTool output collection
a list with 0 datasets
- 32: CMEMS OPeNDAP TEST.ipynb
- 31: CMEMS MOTU Test.ipynb
- 30: CMEMS FTP Test.ipynb
- 29: Executed Holoviz
- 28: Holoviz notebook output collection
a list with datasets
- 14: openEO_2022-05-26Z
- 13: openEO_2022-05-18Z
- 12: openEO_2022-05-15Z



GALAXY FEATURES



FAIR

Galaxy

Tool name, id,
description, ...



Open source
platform for
everyone



One tool for
multiple kind of
datasets



To be chained in
different
workflows



Best practices



Atomisation

Divided in elementary
bricks

Generalisation

Standardised tools

Rigorous

Peer review

Standardisation

AN XML FILE FOR EACH TOOL

- Each script (python, R, java ...) is linked to a “wrapper”, the xml file, where some necessary info must be written.

```
<tool id="obis_data" name="OBIS occurrences" version="@VERSION@" profile="20.01">
  <description>retrieve data</description>
  <macros>
    <import>macro_obis.xml</import>
  </macros>
  <expand macro="topic"/>
  <expand macro="requirements">
    <requirement type="package" version="2.11.3">r-robis</requirement>
  </expand>
  <required_files>
    <include type="literal" path="robis.r"/>
  </required_files>
```

```
<command detect_errors="exit_code"><![CDATA[
Rscript
  '$__tool_directory__/_robis.r'
  '$species'
  '$taxon'
  '$lat_min'
  '$lat_max'
  '$long_min'
  '$long_max'
  '$output'
]]>
```

```
<inputs>
  <param name="species" type="text" format="character" label="Scientific name of the species"
    <validator type="regex">^[A-Za-z ]*$</validator>
  </param>
  <param name="taxon" type="text" label="Taxon ID"
    <validator type="regex">^[0-9]*$</validator>
  </param>
  <param name="lat_min" type="float" min="-90" max="90" value="0" label="Input latitude min (+north/-south)"
  <param name="lat_max" type="float" min="-90" max="90" value="0" label="Input latitude max (+north/-south)"
  <param name="long_min" type="float" min="-90" max="90" value="0" label="Input longitude min (+east/-west)"
  <param name="long_max" type="float" min="-90" max="90" value="0" label="Input longitude max (+east/-west)"
</inputs>
```

```
<outputs>
  <data name="output" format="tabular" from_work_dir="output.tab" label="Species occurrence data"
</outputs>
```

```
<tests>
  <test expect_num_outputs="1">
    <param name="species" value="Scomber scombrus"/>
    <output name="output">
      <assert_contents>
        <has_text text="Scombridae"/>
        <has_n_columns n="163"/>
      </assert_contents>
    </output>
  </test>
  <test expect_num_outputs="1">
    <param name="lat_min" value="6"/>
```

OBIS occurrences retrieve data (Galaxy Version 6.9.2)

Tool Parameters

Scientific name of the species - optional

Genus species format, eg : Scomber scombrus

Taxon ID - optional

Input latitude min (+north/-south): - optional

Input latitude max (+north/-south): - optional

Input longitude min (+east/-west): - optional

Input longitude max (+east/-west): - optional

Additional Options

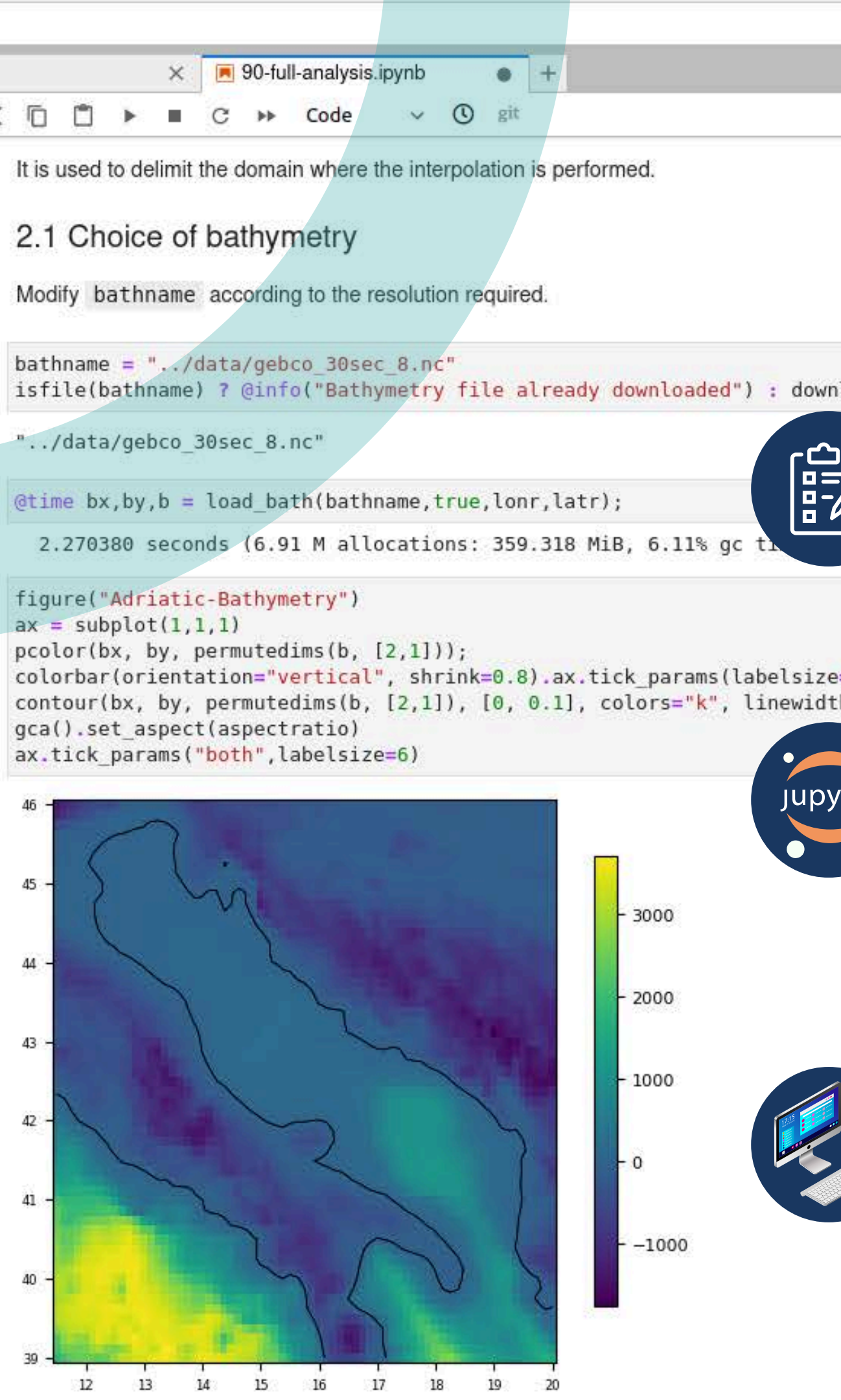
Email notification

No

Send an email notification when the job completes.

Run Tool

GALAXY & TOOLS



Batch tools

Inputs parameters into a formular, execute the tool and directly retrieve the wanted outputs. **NO need for any informatic skills**



Jupyterlab tools

Interactive tools that can be launched with specific environment and set of notebooks (DIVAnd, Holoviz, Copernicus Data Space Ecosystem, Pangeo, ...)

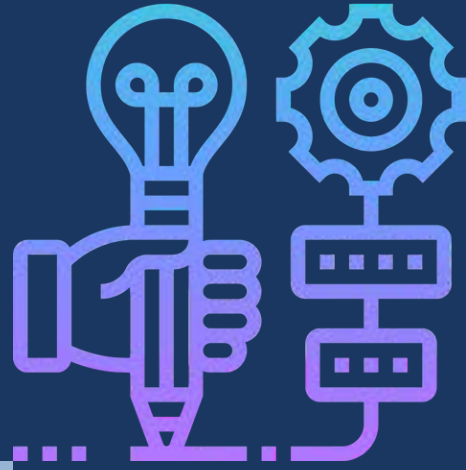


Desktop tools

Intercative tools that can be launched through Galaxy and allow you to have desktop applications launched in a browser window (QGIS, ODV, ...)

Atomisation

CREATE WORKFLOWS FROM ELEMENTARY BRICKS



Workflow Preview ▶ Run

1: Eutrophication_Med_profiles_2022_unrestricted_SNAPSHOT_2023-10-24T16-39-44.zip
output (input)

2: ODV
ODV collection in a zip folder.
 outputs_netcdf (input)
 outputs_all (input)

3: Interactive DIVAnd Notebooks
Include data into the environment
 output_netcdf (input)
 output_all (input)
 jupyter_notebook (ipynb)

4: ODV
Netcdf or tabular text file. For text file, odv format is recommended.
 outputs_netcdf (input)
 outputs_all (input)

About This Workflow

Ocean's variables - Version 2

Author
marie.josse

All published Workflows by marie.josse

Creators
Marie Jossé

Description
Subset data on the Mediterranean see and extract and visualise the Phosphate variable

Tags
earth-system Ocean

Allows better understanding of an analysis

Smaller scripts easier to read

Allows researchers to combine tools

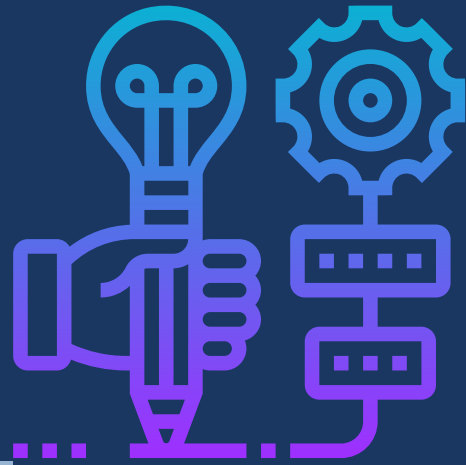
Can combine tools to have a full operational workflow

Allows to create different kind of analysis

Can use the tool with different ones and have multiple workflows

Atomisation

CREATE WORKFLOWS FROM ELEMENTARY BRICKS



Compute and analyze biodiversity metrics with PAMPA toolsuite

Name
Compute and analyze biodiversity metric

Version
1: Oct 13th 2023, 40 steps

Annotation
Compute and analyze biodiversity metrics with PAMPA toolsuite
These notes will be visible when this workflow is viewed.

License
Specify a license for this workflow.

Creator
Add a new creator - either a person or an organization.

Tags
ecology
Add Tags

Apply tags to make it easy to search for and find items with the same tag.

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

The 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

Galaxy Training!
Climate Learning Pathways Help Settings Search Tutorials

Ocean's variables study

Author(s) Marie Josse

Introduction

Managing ODV Galaxy Interactive tool

Ocean Data View

DIVAnd : Data-Interpolating Variational Analysis in n dimensions

Conclusion

Extra Information

Frequently Asked Questions

References

Feedback

Citing this Tutorial

There your data should be opening an you can now visualise them!

[Link to here](#) | [FAQs](#) | [Gitter Chat](#) | [Help Forum](#)

Galaxy Training!
Contributors Learning Pathways Help Extras Search Tutorials

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

See Open Data stewardship

Question

1. What are the longitude and latitude of the red dot?

Solution

GALAXY JOURNEY STEPS



Create tools

As explained you can easily create tools (interactive or not) with an xml file



Set up a workflow

- Can be shared, published, runned, ...
- Workflows with human in the loop : jupyterlab, desktop applications, ...
- Fully automatic workflows : click button tools without any knowledge in programmation



Write a tutorial

- Easily wrote in markdown
- Lot of guides on how to write the perfect tutorial
- Can even be extracted from the workflow.

eosc

FAIR-EASE

OUR TOOLS AND WORKFLOWS



OCEAN VARIABLES STUDY :

Ocean Data View & DIVAnd

Workflow Preview

```
graph LR; A["1: Eutrophication_Med_profiles_2022_unrestricted_SNAPSHOT_2023-10-24T16-39-44.zip"] --> B["2: ODV  
ODV collection in a zip folder.  
- outputs_netcdf (input)  
- outputs_all (input)"]; B --> C["3: Interactive DIVAnd Notebooks  
Include data into the environment  
- output_netcdf (input)  
- output_all (input)  
- jupyter_notebook (ipynb)"]; C --> D["4: ODV  
Netcdf or tabular text file.  
For text file, odv format is recommended.  
- outputs_netcdf (input)  
- outputs_all (input)"]; style A fill:#2c5e8c,color:#fff; style B fill:#2c5e8c,color:#fff; style C fill:#2c5e8c,color:#fff; style D fill:#2c5e8c,color:#fff;
```

This aggregated dataset contains all unrestricted EMODnet Chemistry data on eutrophication and acidity and covers the Mediterranean Sea.

About This Workflow

Ocean's variables - Version 2

Author
marie.josse

All published Workflows by marie.josse

Creators
Marie Jossé

Description
Subset data on the Mediterranean sea and extract and visualise the Phosphate variable

Tags
earth-system Ocean

License
Creative Commons Attribution 4.0 International

Last Updated
Monday Mar 25th 9:30:52 2024 GMT+1

This aggregated dataset contains all unrestricted EMODnet Chemistry data on eutrophication and acidity and covers the Mediterranean Sea.

OCEAN VARIABLES STUDY : Ocean Data View & DIVAnd

Workflow Preview

```
graph LR; A["1: Eutrophication_Med_profiles_2022_unrestricted_SNAPSHOT_2023-10-24T16-39-44.zip"] --> B["2: ODV  
ODV collection in a zip folder.  
- [ ] outputs_netcdf (input)  
- [ ] outputs_all (input)"]; B --> C["3: Interactive DIVAnd Notebooks  
Include data into the environment  
- [ ] output_netcdf (input)  
- [ ] output_all (input)  
- [ ] jupyter_notebook (ipynb)"]; C --> D["4: ODV  
Netcdf or tabular text file.  
For text file, odv format is recommended.  
- [ ] outputs_netcdf (input)  
- [ ] outputs_all (input)"]; style A fill:#2c5e8c,color:#fff; style B fill:#2c5e8c,color:#fff; style C fill:#2c5e8c,color:#fff; style D fill:#2c5e8c,color:#fff;
```

1. First, use of Ocean Data View (ODV) to subset the data

2. Second, Use of DIVAnd in a notebook and its proper jupyterlab to create a climatology

3. Third, Visualise the phosphate variable with ODV

About This Workflow

Ocean's variables - Version 2

Author: marie.josse

Creators: Marie Jossé

Description: Subset data on the Mediterranean sea and extract and visualise the Phosphate variable

Tags: earth-system, Ocean

License: Creative Commons Attribution 4.0 International

Last Updated: Monday Mar 25th 9:30:52 2024 GMT+1



OCEAN VARIABLES STUDY : Ocean Data View & DIVAnd

Workflow Preview

```
graph LR; A["1: Eutrophication_Med_profiles_2022_unrestricted_SNAPSHOT_2023-10-24T16-39-44.zip"] --> B["2: ODV  
ODV collection in a zip folder.  
- outputs_netcdf (input)  
- outputs_all (input)"]; B --> C["3: Interactive DIVAnd Notebooks  
Include data into the environment  
- output_netcdf (input)  
- output_all (input)  
- jupyter_notebook (ipy nb)"]; C --> D["4: ODV  
Netcdf or tabular text file.  
For text file, odv format is recommended.  
- outputs_netcdf (input)  
- outputs_all (input)"]; style A fill:#2c5e8c,color:#fff; style B fill:#2c5e8c,color:#fff; style C fill:#2c5e8c,color:#fff; style D fill:#2c5e8c,color:#fff;
```

About This Workflow
Ocean's variables - Version 2

Author
marie.josse

All published Workflows by marie.josse

Creators
Marie Jossé

Description
Subset data on the Mediterranean sea and extract and visualise the Phosphate variable

Tags
earth-system Ocean

License
Creative Commons Attribution 4.0 International

Last Updated
Monday Mar 25th 9:30:52 2024 GMT+1

Phosphate @ depth [meters]=first

Map showing Phosphate concentration (0 to 0.8) across the Mediterranean Sea region (32°N to 38°N, 20°E to 30°E).

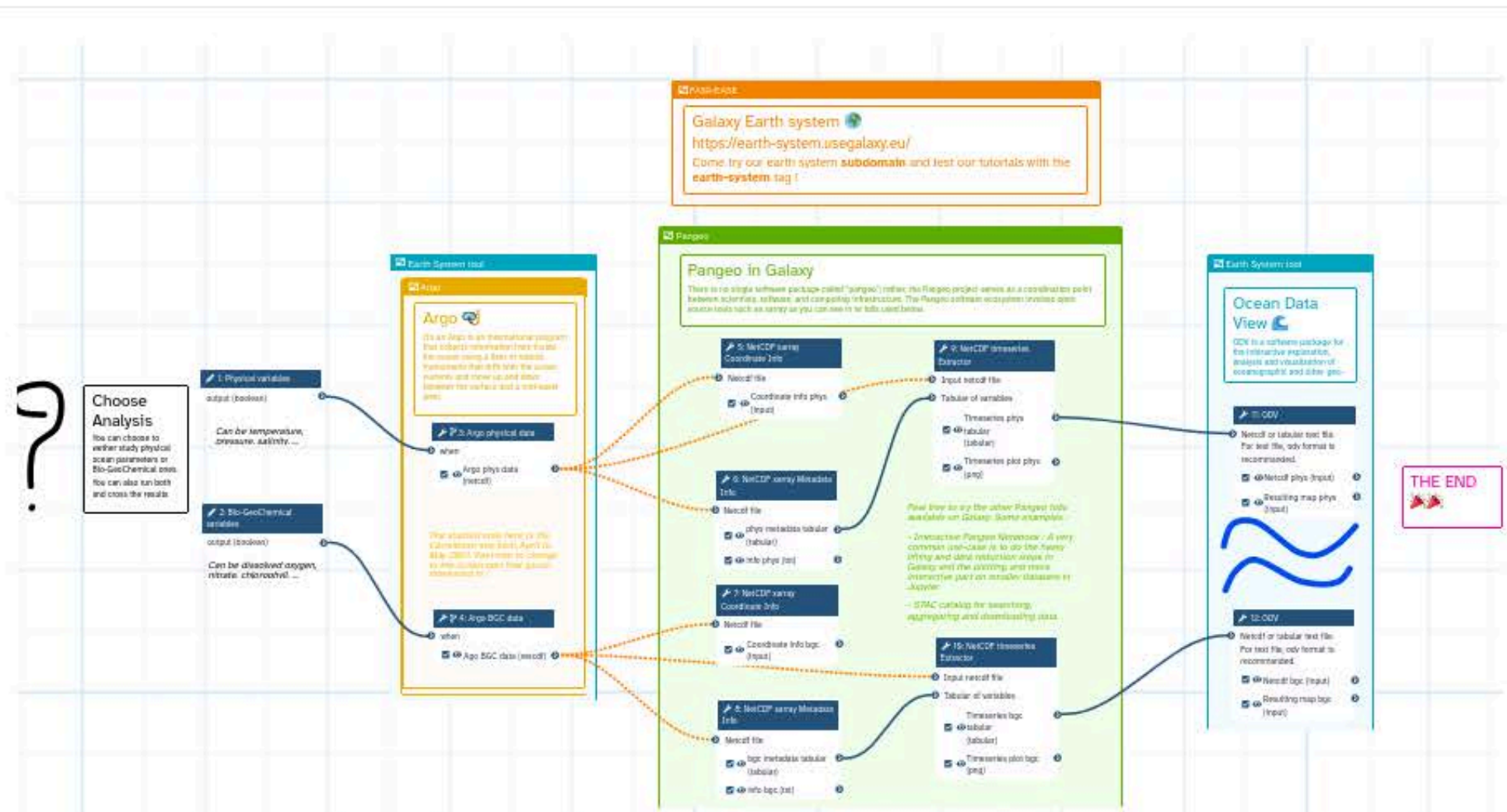
History panel:

- 50: Water_body_Phosphate_Mediterranean_ViewOnP3xW0_1
- 40: Water_body_Phosphate_Mediterranean_ViewOnSurfaceW0_1
- 45: ODV outputs
- 47: Water_body_Phosphate_Mediterranean

ANALYSE ARGO DATA WITH PANGEO TOOLS & ODV

Workflow Preview

▶ Run



About This Workflow

Full Analyse Argo data - Version 14

Author

marie.josse



All published Workflows by marie.josse

Creators

Marie Jossé

Description

Process argo data with the Pangeo Ecosystem and visualise them with Ocean Data View (ODV)

Tags

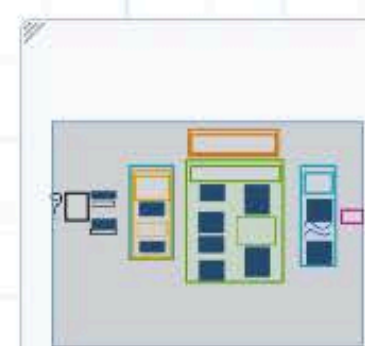
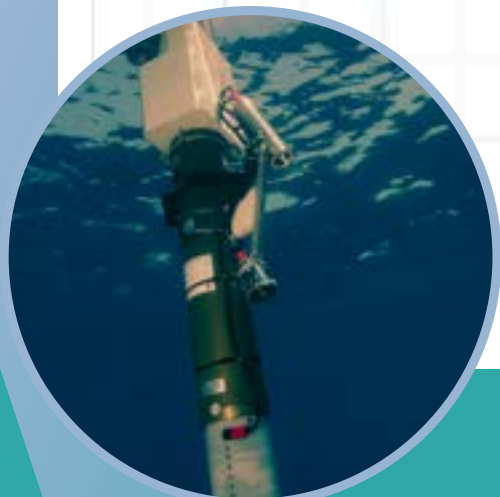
Ocean Pangeo earth-system

License

Creative Commons Attribution 4.0 International

Last Updated

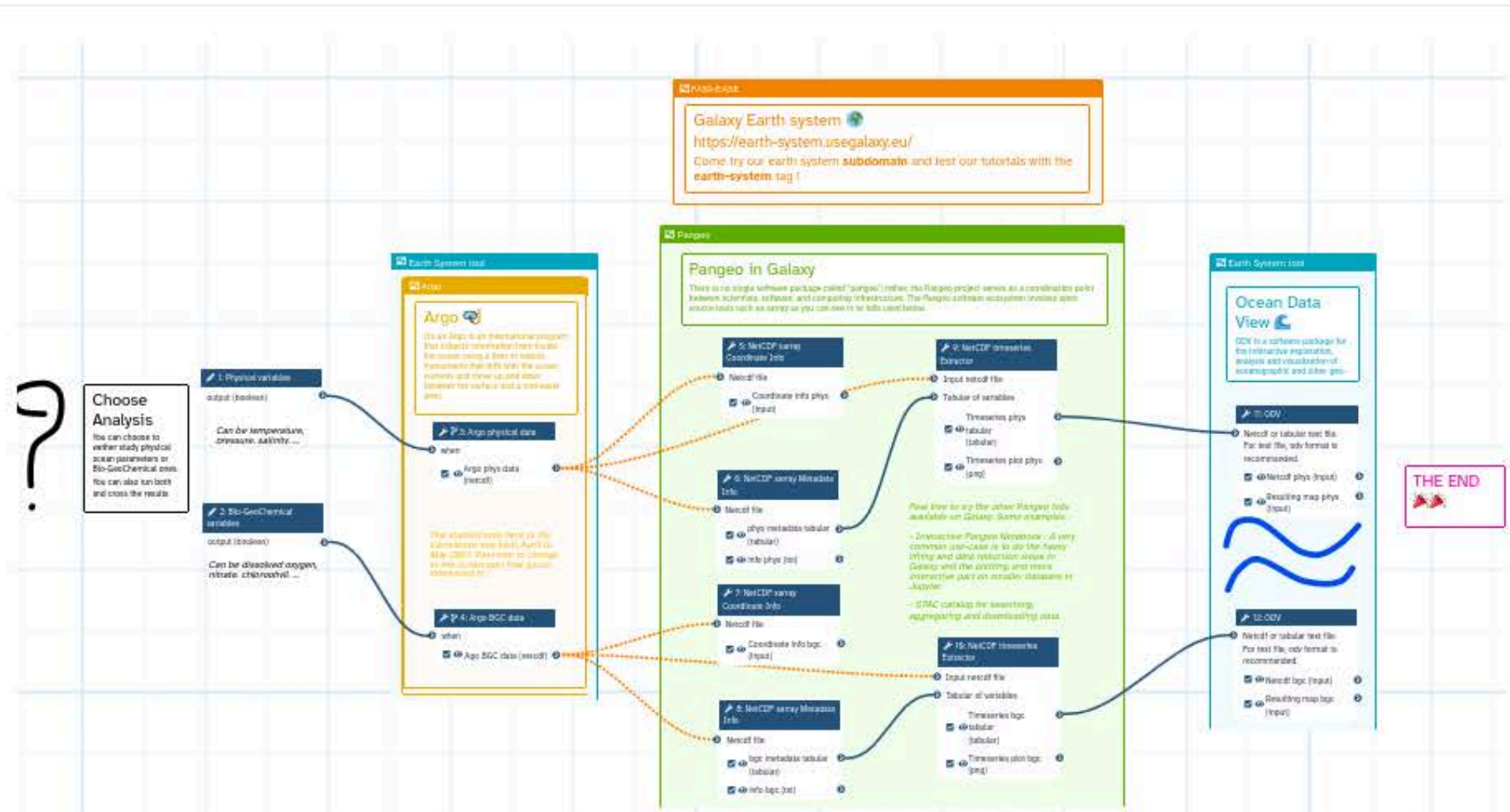
Monday Apr 8th 11:52:07 2024 GMT+2



ANALYSE ARGO DATA WITH PANGEO TOOLS & ODV

Workflow Preview

▶ Run



About This Workflow

Full Analyse Argo data - Version 14

Author

marie.josse



All published Workflows by marie.josse

Creators

Marie Jossé

Description

Process argo data with the Pangeo Ecosystem and visualise them with Ocean Data View (ODV)

Tags

Ocean Pangeo earth-system

License

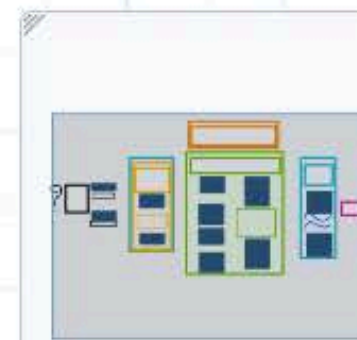
Creative Commons Attribution 4.0 International

Last Updated

Monday Apr 8th 11:52:07 2024 GMT+2



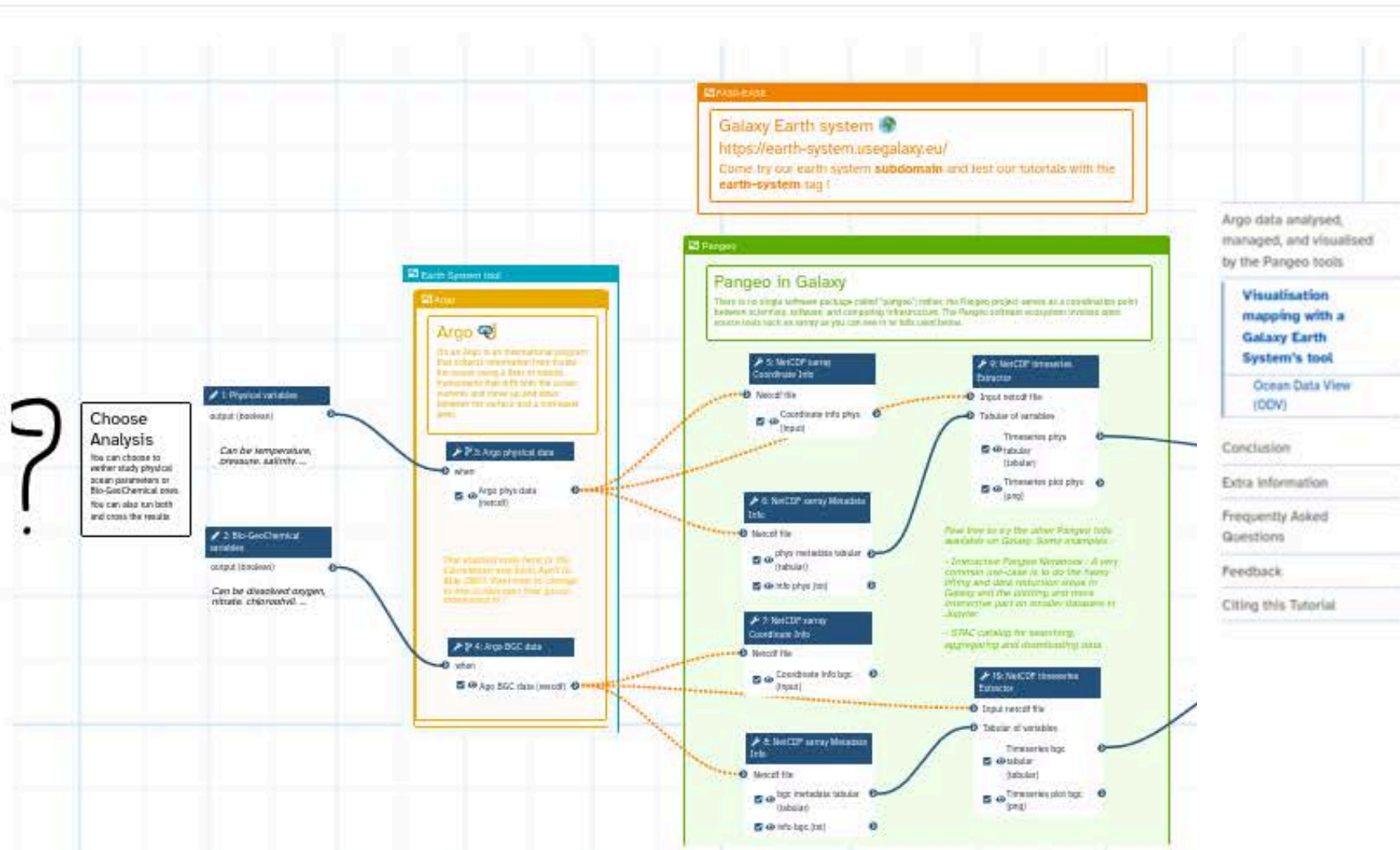
1. First, choose if you want biogeochemical or physical Argo data and retrieve them
2. - Second, use of pangeo tools to clean the data, extract the metadata and subset the data
3. Third, Visualise the Argo data with ODV



ANALYSE ARGO DATA WITH PANGEO TOOLS & ODV

Workflow Preview

▶ Run



About This Workflow

Full Analyse Argo data - Version 14

Author

marie.josse

Hands-on: Save your visualisation map

1. Click right on the map select **Save Plot As...**
2. In the pop-up screen go to the folder **ODV, galaxy, outputs**.
3. In **File name** rename your view (for example **subset_temp_argo_data**)
4. In **Files of type** select **PNG (*.png *.PNG)** and **Save** then **OK** and **OK**. Once you're finished with ODV :
5. On th top left click on **File** select **Exit**
6. If you want to save the other window also click on **Yes**. Here we don't need it so click **No**.

[Link to here](#) | [FAQs](#) | [Gitter Chat](#) | [Help Forum](#)

Argo data analysed, managed, and visualised by the Pangeo tools

Visualization mapping with a Galaxy Earth System's tool
Ocean Data View (ODV)

Conclusion

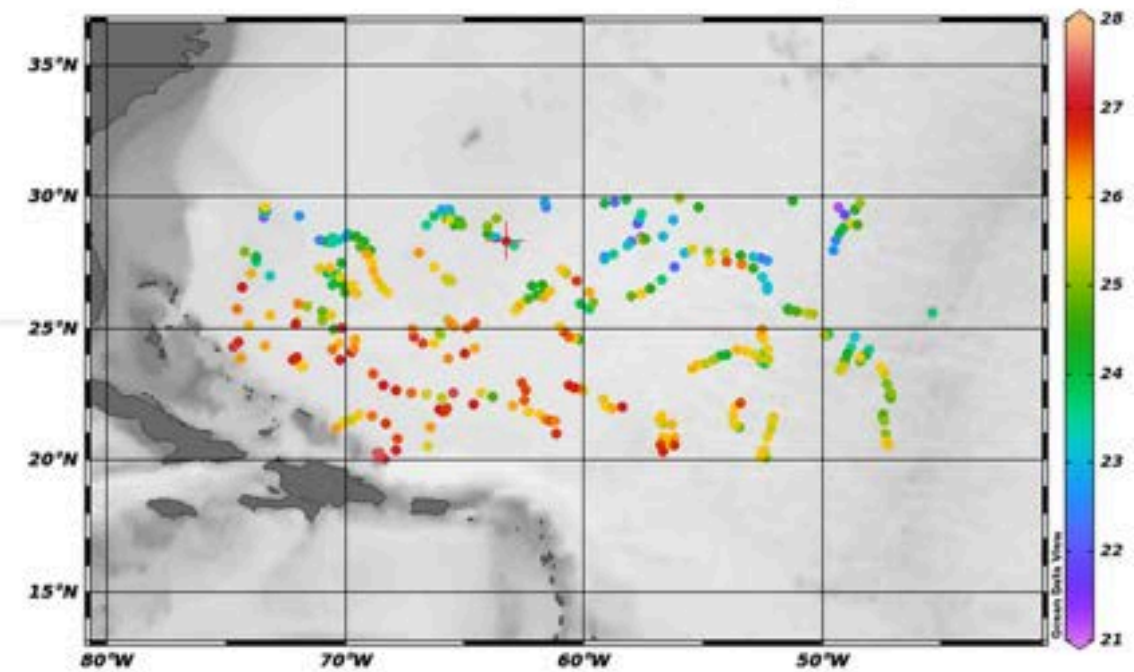
Extra information

Frequently Asked Questions

Feedback

Citing this Tutorial

TEMP @ N_POINTS=first

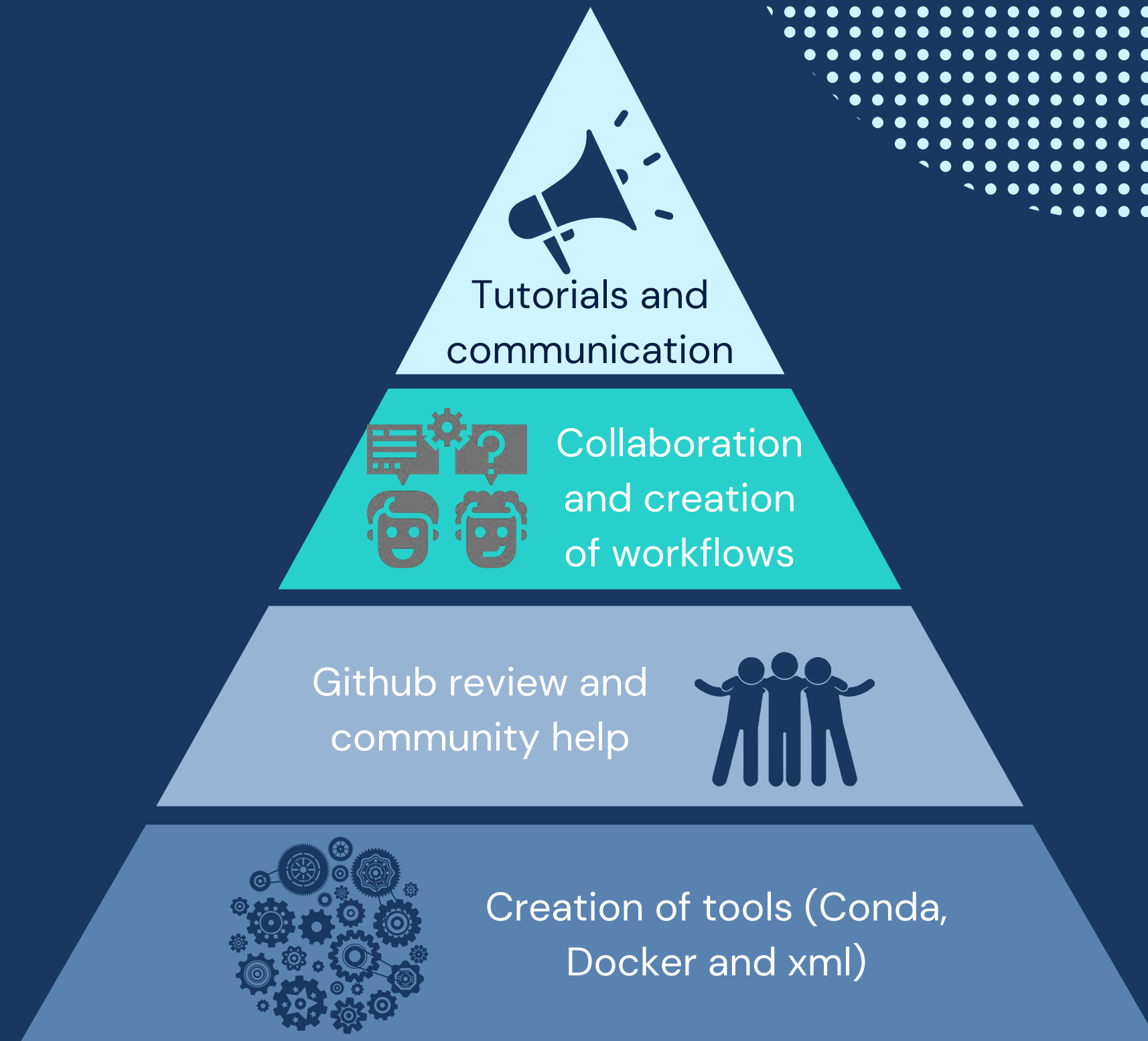


What you need to keep in mind !

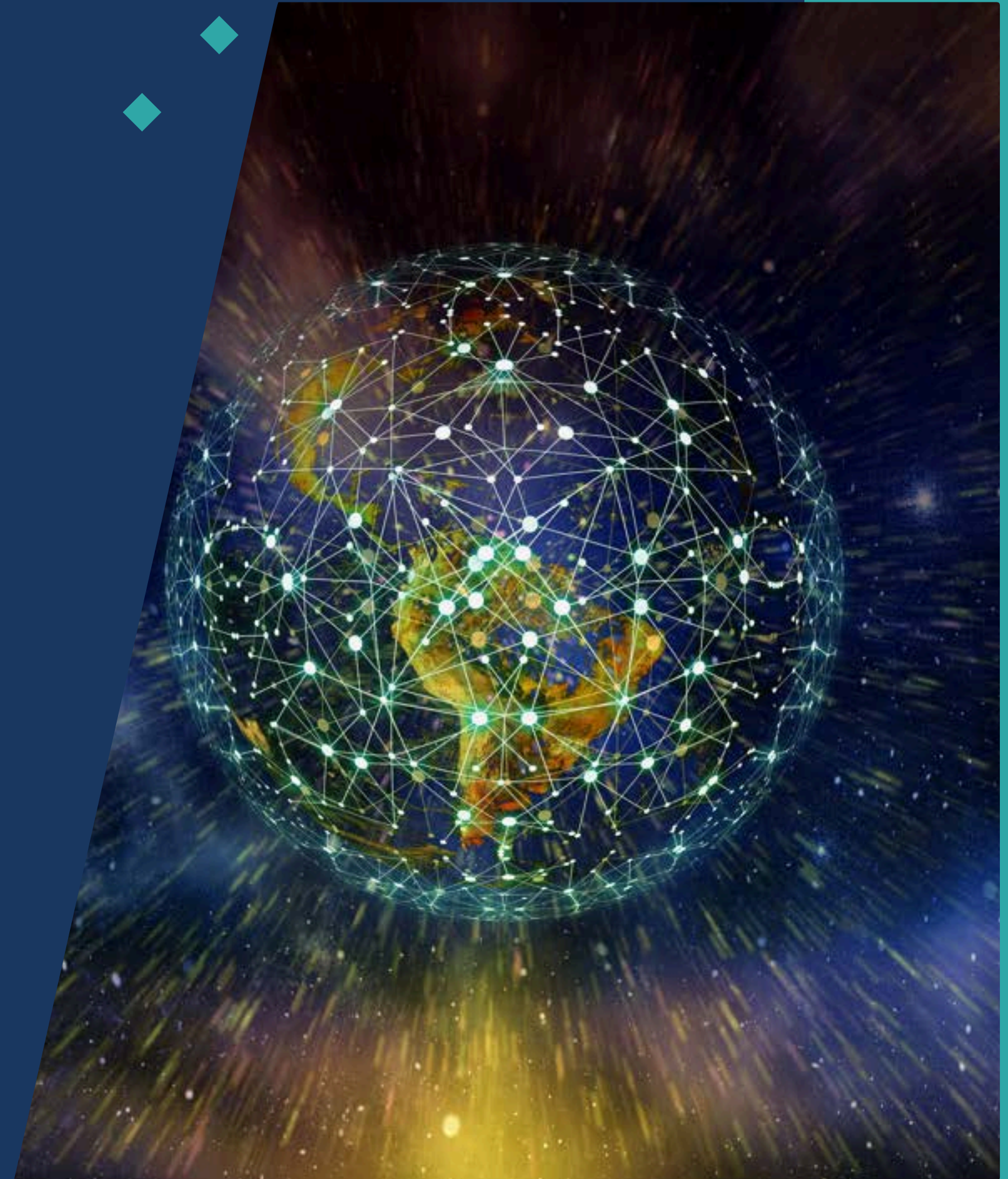
Galaxy is an **open-source** platform for **FAIR** data analysis that enables users to:

- 01** • Use **tools** from various domains (that can be plugged into workflows) through its graphical web interface.
- 02** • Run code in **interactive environments** (RStudio, Jupyter...) along with other tools or **workflows**.
- 03** • Manage data by **sharing and publishing** results, workflows, and visualizations.
- 04** • 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**.



THANK YOU



27-29 May 2024 

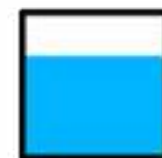


imdis

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



MARIS



National
Oceanography
Centre



eosc
Blue-Cloud2026