



European Digital Twin Ocean



Earth Sciences  
Department



Barcelona  
Supercomputing  
Center  
Centro Nacional de Supercomputación



Funded by  
the European Union

# Workflow Provenance with RO-Crate in Autosubmit

Bruno De Paula Kinoshita, Daniel Beltrán Mora,  
Miguel Castrillo, **Manuel G. Marciani**, and Luiggi  
Tenorio Ku

[bruno.depaulakinoshita@bsc.es](mailto:bruno.depaulakinoshita@bsc.es), [manuel.gimenez@bsc.es](mailto:manuel.gimenez@bsc.es)

May 7, 2024

**Why?**

**How?**

**What?**

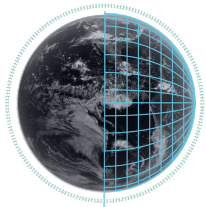


**Why** we implement provenance?

**How** did we implement provenance?

**What** did we achieve?

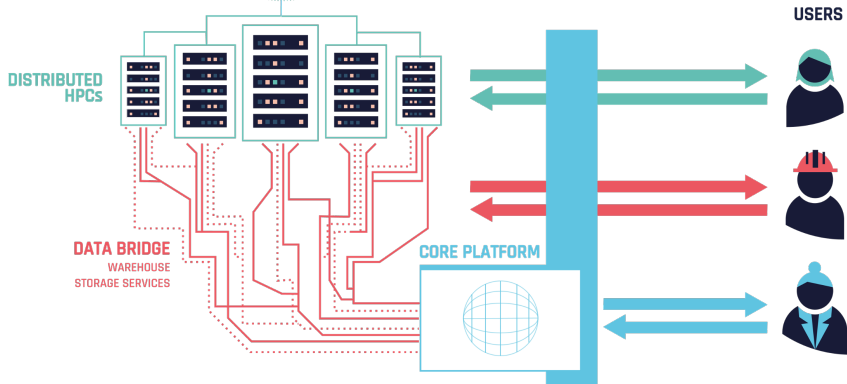




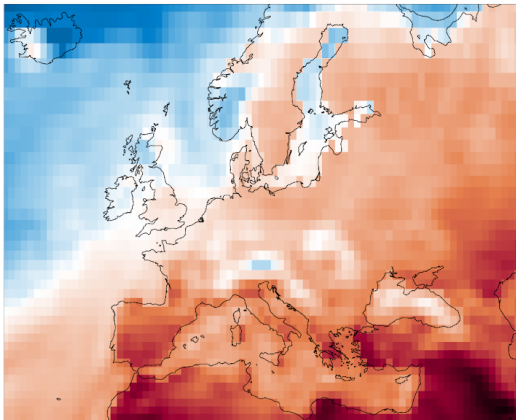
Funded by  
the European Union

# Destination Earth

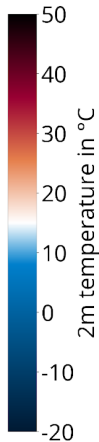
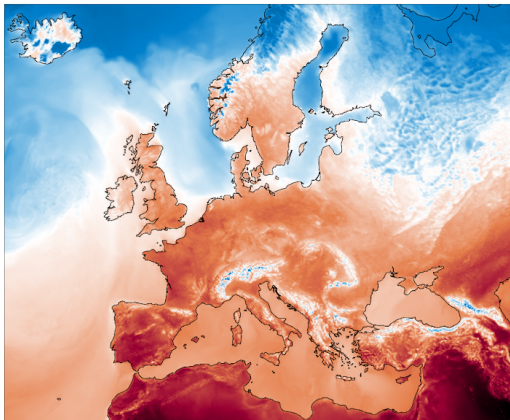
implemented by



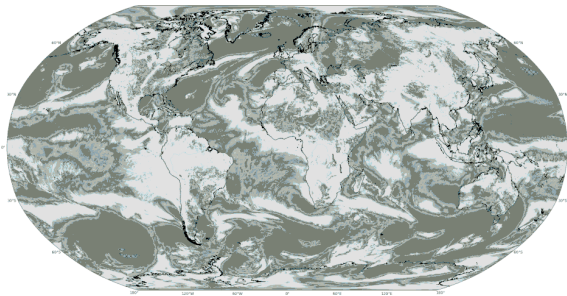
IPCC AR6 (2021), 100km

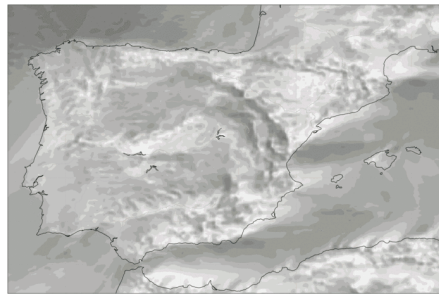
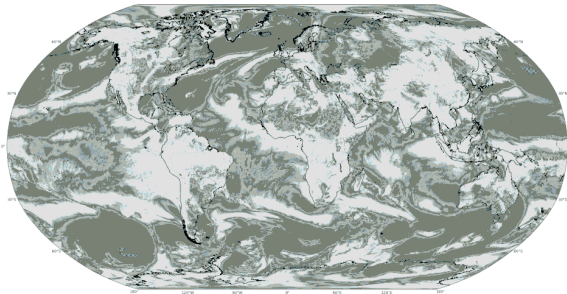


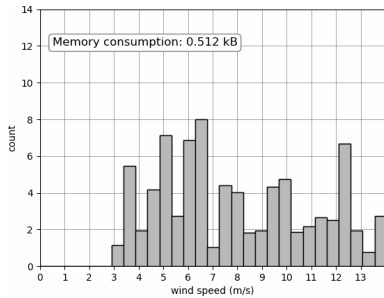
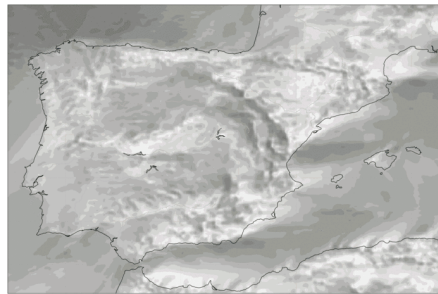
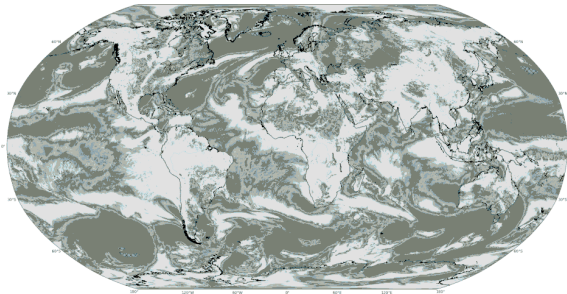
Climate DT, 5km



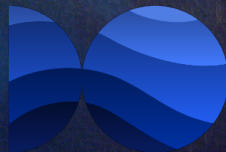












# EDITO

European Digital  
Twin Ocean

Image Courtesy of NASA GHRC

**AUTOSUBMIT** 



Used to **create** workflows for **both projects**



Used to **create** workflows for **both projects**

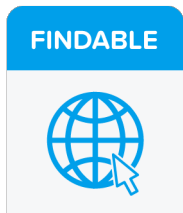
**Orchestrates** novel workflows on **flagship HPC platforms**



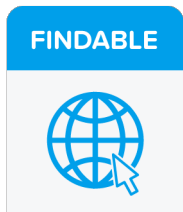


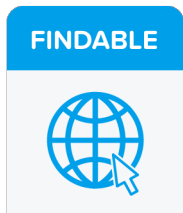
Image courtesy of AXVECO B.V.













**FINDABLE**



**ACCESSIBLE**



**INTER-  
OPERABLE**



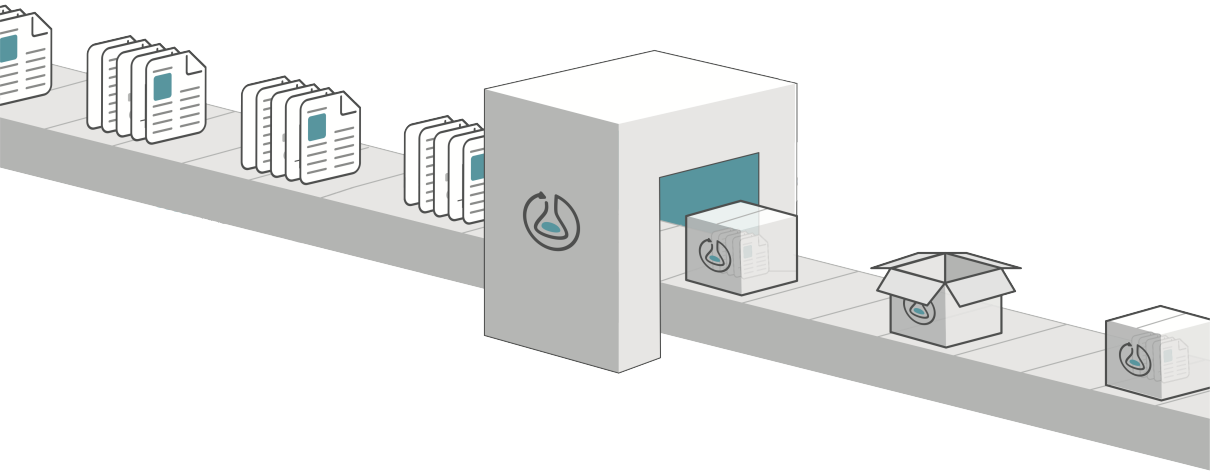
**REUSABLE**

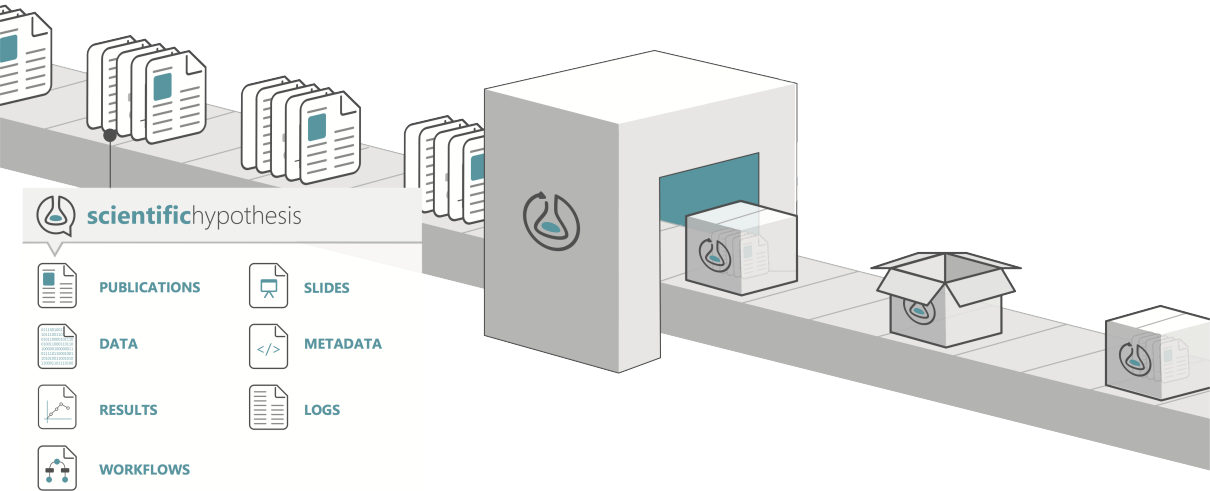


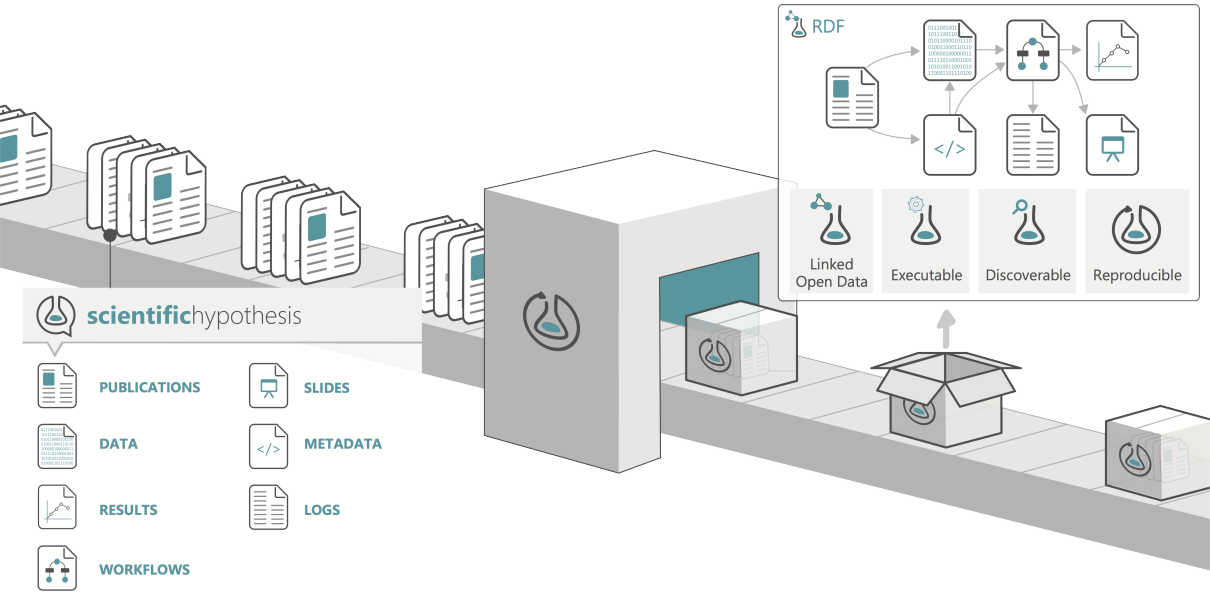
**REPRO-  
DUCIBLE**











 **scientific hypothesis**

-  **PUBLICATIONS**
-  **SLIDES**
-  **DATA**
-  **METADATA**
-  **RESULTS**
-  **LOGS**
-  **WORKFLOWS**

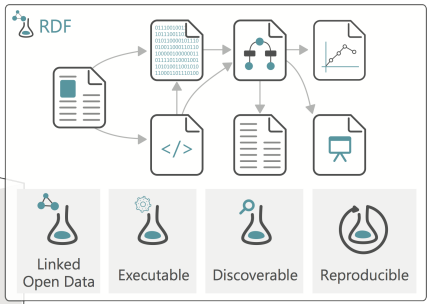
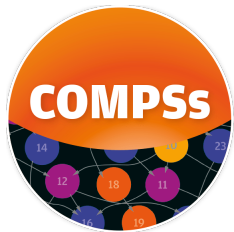
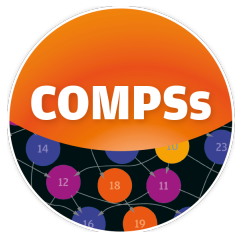
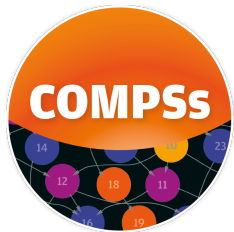


Image adapted from Research Object









COMMON  
WORKFLOW  
LANGUAGE



COMMON  
WORKFLOW  
LANGUAGE

zenodo



COMMON  
WORKFLOW  
LANGUAGE



zenodo



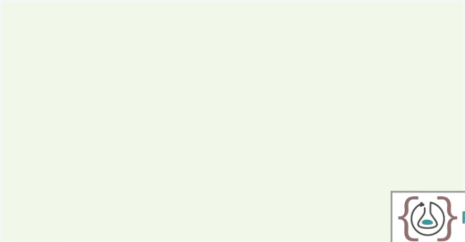
COMMON  
WORKFLOW  
LANGUAGE

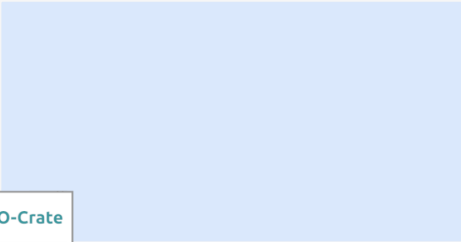
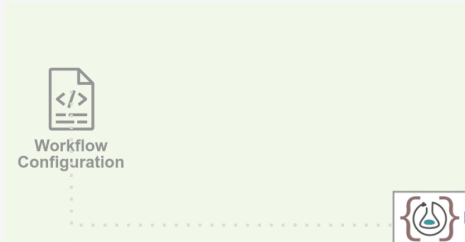


**EUROPEAN OPEN  
SCIENCE CLOUD**

**zenodo**

```
{
    "@id" : "file:///conf/experiment.yml",
    "@type" : "File",
    "contentSize" : 996,
    "dateModified" : "2023-11-03T22:41:31",
    "name" : "experiment.yml",
    "sdDatePublished" : "2023-11-03T22:43:56+00:00" }
```



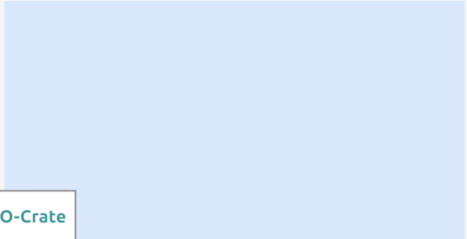




Two document icons with code symbols (</>) are shown. The first is labeled "Workflow Configuration" and the second is labeled "Run Files".

Workflow Configuration

Run Files

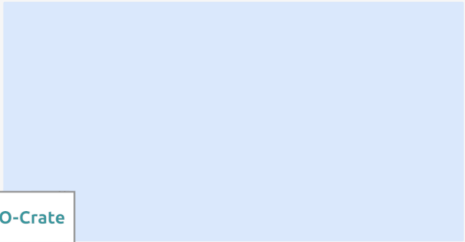


Three document icons, each containing a code symbol (</>), representing different types of files. Below each icon is a label: "Workflow Configuration", "Run Files", and "Model Config".

Workflow Configuration

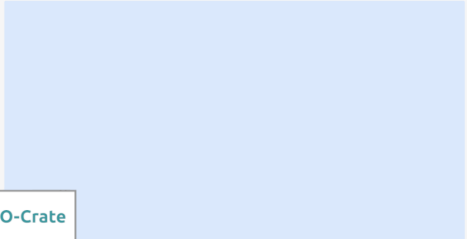
Run Files

Model Config





A light green rectangular box containing four document icons, each with a code symbol (&lt;/&gt;). Below each icon is a label: "Workflow Configuration", "Run Files", "Model Config", and "Authorship". Dotted lines connect each icon to a central point above the RO-Crate logo.





A light green rectangular box containing four document icons, each with a code symbol (&lt;/&gt;). Below each icon is a label: "Workflow Configuration", "Run Files", "Model Config", and "Authorship". Dotted lines connect each icon to a central point below the "Authorship" label.

A light blue rectangular box containing one document icon with a code symbol (&lt;/&gt;). Below the icon is the label "Logs". A dotted line connects the icon to the "RO-Crate" logo below.





Workflow Configuration

Run Files

Model Config

Authorship

Logs

Dependencies






Workflow Configuration

Run Files

Model Config

Authorship




Dotted lines connect these four categories to the RO-Crate logo below.

Logs

Dependencies

Resource Usage



Dotted lines connect these three categories to the RO-Crate logo below.





Workflow Configuration

Run Files

Model Config

Authorship

RO-Crate

This block contains four document icons with code symbols (</>). Below each icon is a label: "Workflow Configuration", "Run Files", "Model Config", and "Authorship". A dashed line connects these four icons to the "RO-Crate" logo, which is a stylized blue and red icon with the text "RO-Crate" next to it.

Logs

Dependencies

Resource Usage

Model Outputs

This block contains four document icons with code symbols (</>). Below each icon is a label: "Logs", "Dependencies", "Resource Usage", and "Model Outputs". A dashed line connects these four icons to the "RO-Crate" logo.





Workflow Configuration

Run Files

Model Config

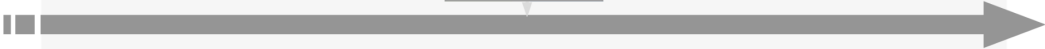
Authorship

Logs

Dependencies

Resource Usage

Model Outputs



Atmo Model

Ocean Model

Data Quality app

Stats app

Data Assim

Custom Apps



```
autosubmit archive --rocrate $expid
```

```
autosubmit archive --rocrate $expid
```

```
autosubmit archive --rocrate $expid
```





# Autosubmit mHM test domains

Version 1

[View on GitHub](#)[Download RO Crate](#)[Overview](#)[Files](#)[Related items](#)

Total size: 1.74 MB

- ▲ conf
  - ▲ metadata
    - experiment\_data.yml Main Workflow
    - minimal.yml
  - ▶ pkl
  - ▶ plot
  - ▶ proj
  - ▶ status
  - ▶ tmp
  - ro-crate-metadata.json
  - ro-crate-preview.html





## References

BEDIA, Joaquín, et al. The METACLIP semantic provenance framework for climate products.



LEO, Simone, et al. Recording provenance of workflow runs with RO-Crate.



MANUBENS-GIL, Domingo, et al. Seamless management of ensemble climate prediction experiments on HPC platforms.



SOILAND-REYES, Stian, et al. Packaging research artefacts with RO-Crate.





European Digital Twin Ocean



*This project has received funding from the European High-Performance Computing Joint Undertaking (JU) under grant agreement N° 955558.*

*This project has received funding from the European Union's Horizon Europe research and innovation programme under the grant N° 101093293.*

Earth Sciences  
Department



Barcelona  
Supercomputing  
Center  
Centro Nacional de Supercomputación



Funded by  
the European Union

# Questions?

Bruno De Paula Kinoshita, Daniel Beltrán  
Mora, Miguel Castrillo, **Manuel G.  
Marciani**, and Luigi Tenorio Ku

[bruno.depaulakinoshita@bsc.es](mailto:bruno.depaulakinoshita@bsc.es), [manuel.gimenez@bsc.es](mailto:manuel.gimenez@bsc.es)