

# Workflow Provenance with RO-Crate in Autosubmit

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

June 12, 2024

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

**Why?**

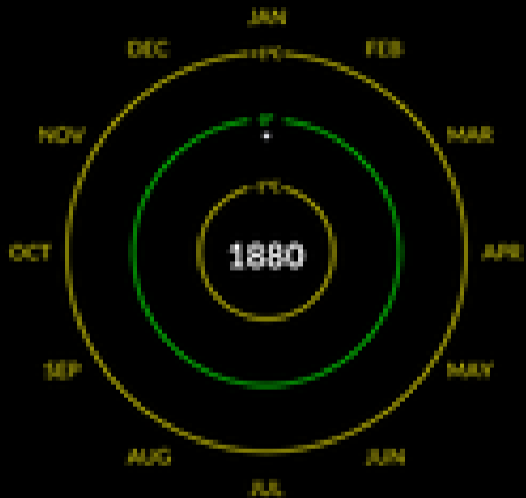
**How?**

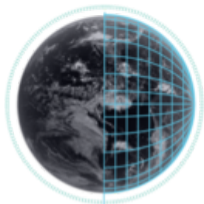
**What?**

**Why** we implement provenance?

**How** did we implement provenance?

**What** did we achieve?

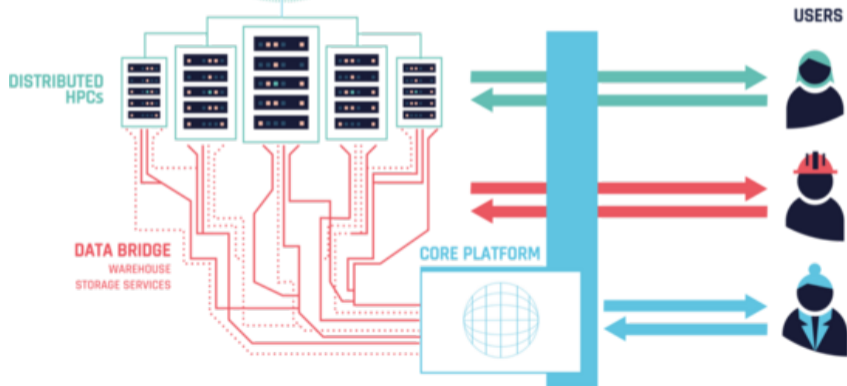




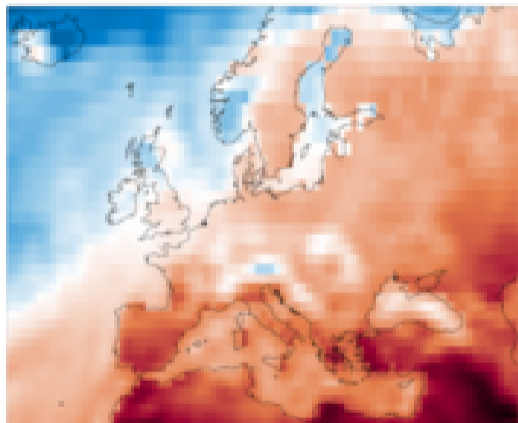
Funded by  
the European Union

**Destination Earth**

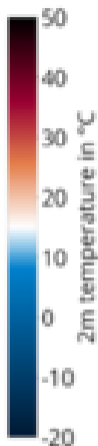
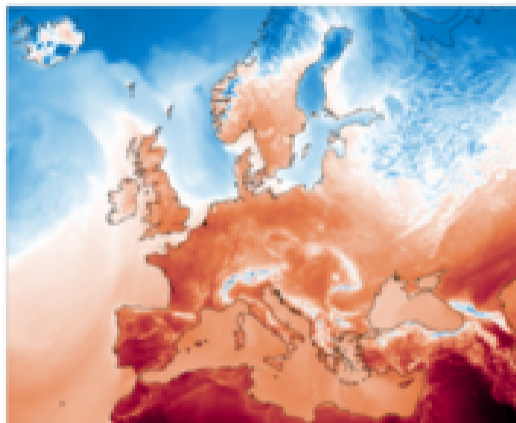
implemented by

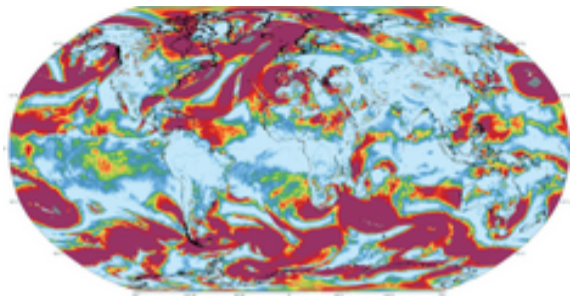


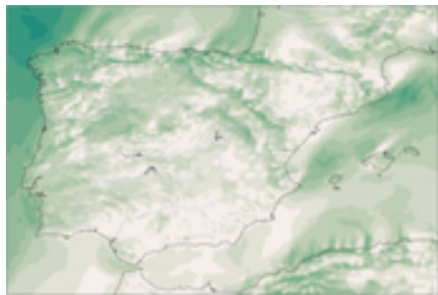
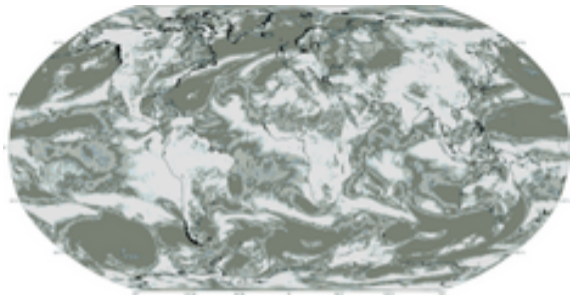
IPCC AR6 (2021), 100km



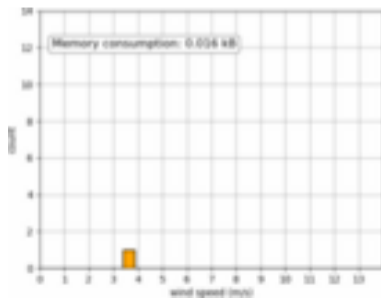
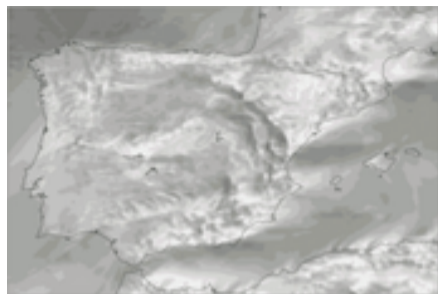
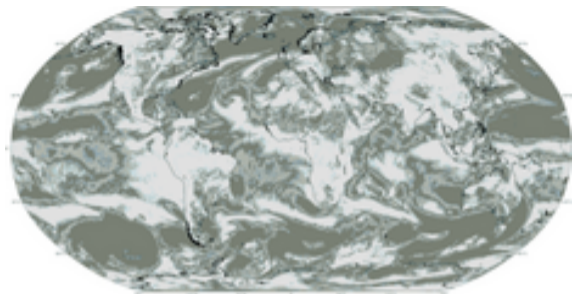
Climate DT, 5km

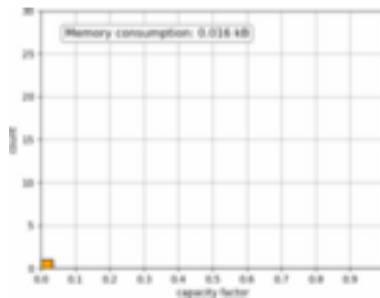
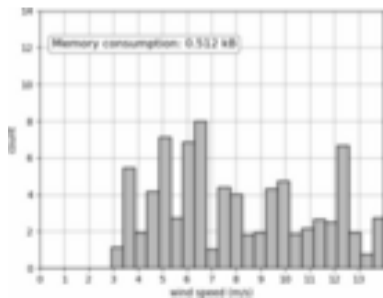
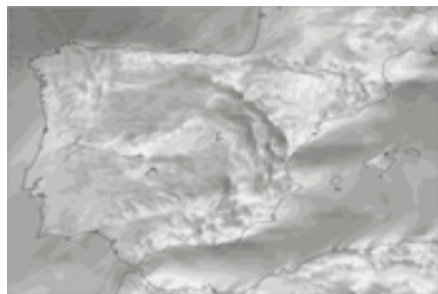
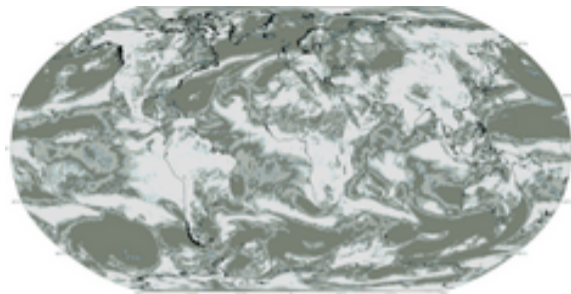














# EDITO

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



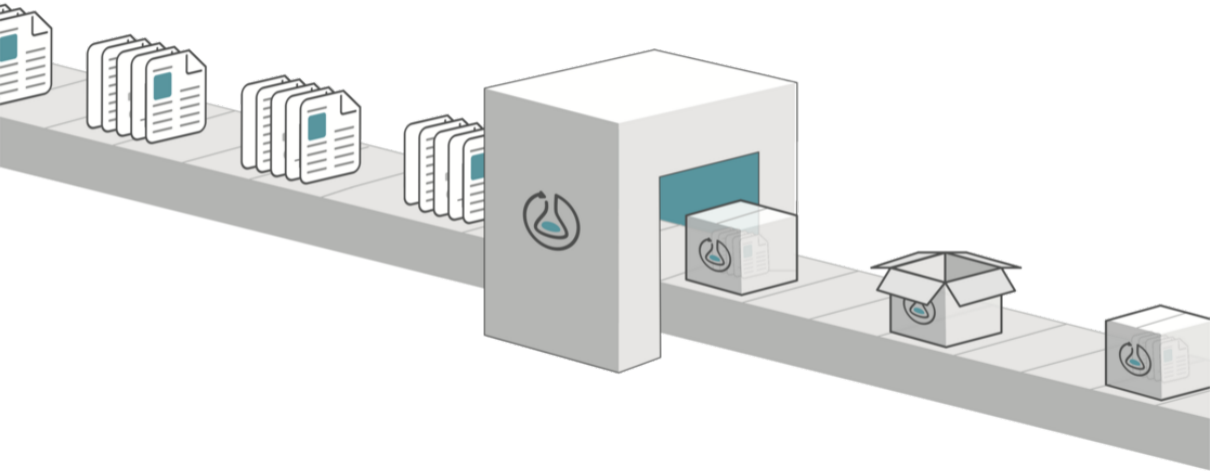
REUSABLE





Image adapted from AXVECO B.V.





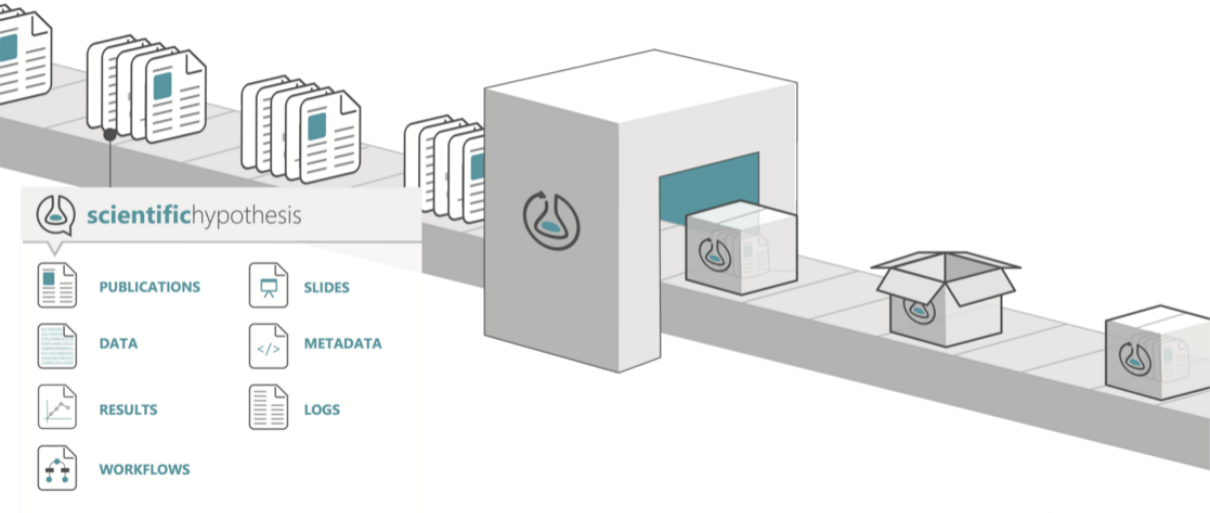


Image adapted from Research Object



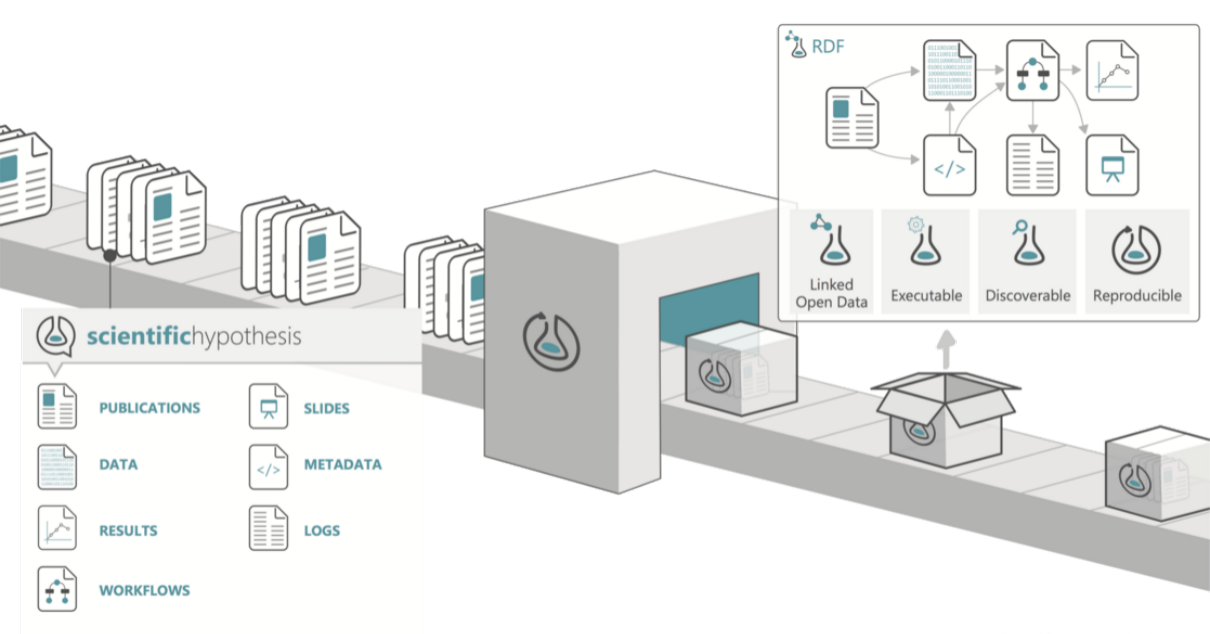


Image adapted from Research Object







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" }
```





**AUTO**SUBMIT



Workflow  
Configuration



RO-Crate

ZIP

100

**AUTO**SUBMIT



Workflow  
Configuration



Run Files



RO-Crate

ZIP

100

**AUTO**SUBMIT



Workflow  
Configuration



Run Files



Model Config



RO-Crate

ZIP

**AUTO**SUBMIT



Workflow  
Configuration



Run Files



Model Config



Authorship



RO-Crate

ZIP

**AUTO**SUBMIT



Workflow  
Configuration



Run Files



Model Config



Authorship



Logs



RO-Crate

ZIP

100

**AUTO**SUBMIT



Workflow  
Configuration



Run Files



Model Config



Authorship



Logs



Dependencies



RO-Crate

ZIP

100

**AUTO**SUBMIT



Workflow  
Configuration



Run Files



Model Config



Authorship



Logs



Dependencies



Resource Usage



RO-Crate

ZIP

100



**AUTOSUBMIT**



Workflow Configuration



Run Files



Model Config



Authorship



Logs



Dependencies



Resource Usage



Model Outputs



RO-Crate

ZIP

RO-C



Workflow Configuration



Run Files



Model Config



Authorship



Logs



Dependencies



Resource Usage



Model Outputs



RO-Crate

ZIP



Atmo Model



Ocean Model



Data Quality app



Stats app



Data Assim



Custom Apps



META CLIP

```
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



# AUTOSUBMIT Autosubmit mHM test domains

Version 1

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

Preview

✕

Path: ro-crate-metadata.json

Size: 40.3 KB

[Download](#)[Raw](#)

Total size: 1.74 MB

- conf
  - metadata
    - experimenter
    - minimal.yml
  - pkl
  - plot
  - proj
  - status
  - tmp
  - ro-crate-metadata.json
  - ro-crate-preview.html

```
{
  "@context": "https://w3id.org/ro/crate/1.1/context",
  "@graph": [
    {
      "@id": "ro-crate-metadata.json",
      "@type": "CreativeWork",
      "about": {
        "@id": "./"
      },
      "author": [
        {
          "@id": "https://orcid.org/0000-0001-8250-4074"
        }
      ],
      "conformsTo": [
        {
          "@id": "https://w3id.org/ro/crate/1.1"
        },
        {
          "@id": "https://w3id.org/workflowhub/workflow-ro-crate/1.0"
        }
      ]
    }
  ]
}
```





## 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.



*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.*

# 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)