Visualizing ClimateDT workflow provenance with RO-Crate and METACLIP

MARTÍ ORTÍS SUNYER, COMPUTATIONAL EARTH SCIENCES, BARCELONA SUPERCOMPUTING CENTER
marti.ortis@bsc.es

Workflow context and scientific detail, explored together through the Provenance Viewer.

1 / WHY WE NEED PROVENANCE

Keep the execution context

Modern climate workflows are increasingly complex, involving multiple models, processing steps, dependencies and computing environments. Capturing provenance helps explain how a particular workflow ran and supports traceability and reproducibility. However, provenance metadata can be difficult for scientists and other users to inspect and interpret in its raw form.

2 / CAPTURE THE WORKFLOW

Autosubmit + RO-Crate

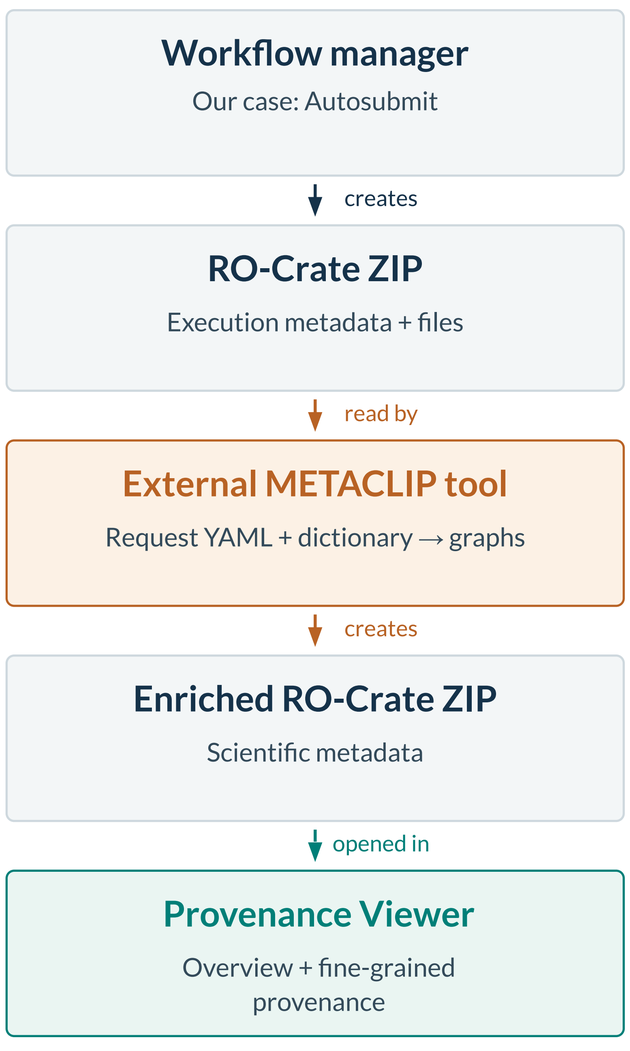
Autosubmit manages ClimateDT jobs and dependencies. When provenance generation is enabled, it can automatically create an RO-Crate ZIP: a package of related resources and structured metadata. The Workflow Run Crate profile provides conventions for describing a workflow execution. The `ro-crate-metadata.json` file describes files, software, people and execution actions, and records the relationships between them. This approach supports the FAIR principles by making metadata easier to exchange across tools and documenting the context needed for reuse and reproducibility.

3 / ADD SCIENTIFIC DETAIL

GSV + OPA through METACLIP

Within the ClimateDT workflow, GSV (Generic State Vector) requests specify which ClimateDT data to retrieve, including the model, variables, dates, region and grid. OPA (One-Pass Algorithms) requests specify how to process those data, including statistical operations, their parameters and frequencies. Our external tool maps these request configurations into knowledge graphs using METACLIP, a climate-domain ontology. This connects workflow-level provenance with detailed descriptions of data selection and processing, which users can explore in the Provenance Viewer.

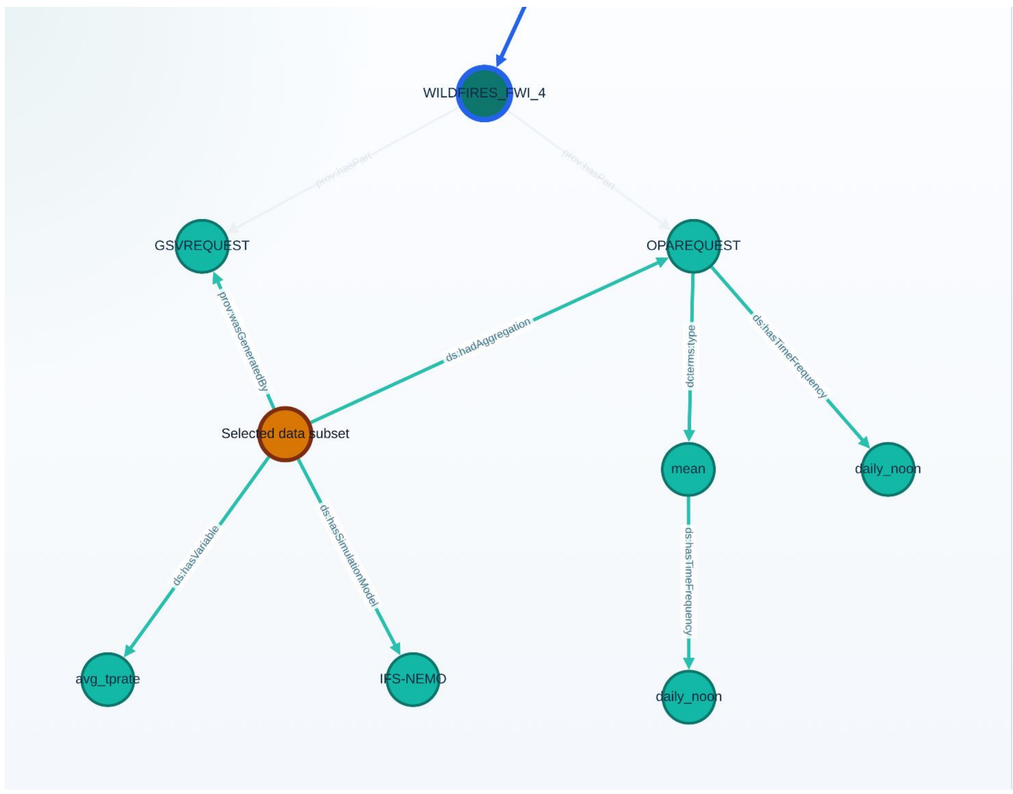
CURRENT METADATA ENRICHMENT PROCESS AND VISUALIZATION



4 / EXPLORE THE METADATA

From execution overview to scientific detail

The Provenance Viewer brings together RO-Crate metadata and METACLIP request graphs. Users can search for a variable and inspect the model, data selection and statistical processing specified in each request. Here, the graph shows the relationships within the WILDFIRES_FWI_4 request for the time-mean total precipitation rate (avg_tprate), while the side panel presents its configuration.



5 / FOLLOW ONE REQUEST

WILDFIRES_FWI_4 requests a noon-to-noon daily mean of avg_tprate from IFS-NEMO

The WILDFIRES_FWI_4 application requests a noon-to-noon daily mean of the time-mean total precipitation rate (avg_tprate) from IFS-NEMO. The viewer lets us inspect the data selection and processing settings, including the grid, interpolation, and statistical and output frequencies, without having to locate and read the underlying metadata files manually.

Which data?

- **Time-mean total precipitation rate**

Which processing?

- **Mean**

Which selection?

- **1 Jan 1991 · 00:00–23:00, hourly · Surface**

Which statistical frequency?

- **Daily, noon to noon**

Which grid and interpolation?

- **Grid: 0.1 / 0.1 · Nearest neighbour**

Which output frequency?

- **Daily, noon to noon**

6 / NEXT STEPS

Generate fine-grained provenance within the workflow

Generate and attach METACLIP request graphs during workflow execution.

External enrichment → **Workflow integration**

Explore it through existing tools

Bring provenance exploration into the Autosubmit GUI.

Standalone viewer → **Exploration in existing tools**

Explore multiple experiments together

Visualize provenance from multiple experiments and search across them.

Single experiment → **Multiple-experiment exploration**

Earth Sciences Department
Barcelona Supercomputing Center
Centro Nacional de Supercomputación
Funded by the European Union
Destination Earth
implemented by ECMWF esa EUMETSAT