LabID-PROV: Tracking and Sharing Data Provenance with RO-Crate in Lab Integrated Data

OSCARS
Open Science Clusters' Action
for Research & Society

LS RI
Life Sciences

The LabID-PROV project aims to extend the Lab Integrated Data (LabID) data model to include accurate modelling of both workflow (WF) and WF runs, offering a unified application to manage derived data provenance independently from analysis procedure and platform, and providing a concrete solution to ensure the traceability of derived data.

Challenge

While primary research data is regularly shared in repositories, the sharing of final derived datasets remains inadequate, often due to the lack of comprehensive provenance metadata. Also, the diverse languages, tools, and computing environments used complicate the tracking of workflows and their metadata.

Solution

LabID-PROV will enhance the LabID platform by allowing accurate modelling of both WFs and WF runs. Leveraging several resources in the LS RI Science Cluster, it will streamline the import & export of datasets (and their metadata) described using Workflow Run RO-Crate profiles. The project will also implement use cases using both omics and imaging data.

Scientific Impact

LabID-PROV will bridge critical gaps in data traceability to facilitate FAIR sharing of derived data together with their provenance metadata, enhancing reproducibility and adoption of FAIR data management practices.

Partners

European Molecular Biology Laboratory - EMBL

https://www.oscars-project.eu/projects/labid-prov-tracking-and-sharing-data-provenance-ro-crate-lab-integrated-data