HMC Conference
> Metadata in Action, 28.–30.04.2026, Heidelberg

<HMC> HELMHOLTZ
Metadata Collaboration

Beitrag ID: **17**   Typ: **Talk**

LabID-Prov : Unifying experimental and computational data provenance with LabID

LabID is an open-source platform designed to streamline research data management for scientists, research groups, and core facilities of life-science institutes. By integrating sample and dataset management, inventory tracking, and an electronic lab notebook, LabID enables users to organize, annotate, and share experimental data in compliance with FAIR principles. At its core, LabID uses a relational database to document entities— such as instruments, samples, and resulting datasets—and their relationships, constructing a comprehensive knowledge graph that traces data provenance from biological specimens to raw outputs.

We introduce **LabID-Prov**, an extension that enhances this knowledge graph by capturing post-acquisition data processing steps—typically executed via computational workflows. While a variety of platforms and software exist to process raw data with workflows and scripts, there is no central solution to document these processes while preserving data provenance.

With LabID-Prov, the knowledge graph of data provenance has been expanded to incorporate **derived datasets produced by computational workflows**. These datasets are directly linked to their source raw data, while the software, parameters, and processing methods applied are systematically documented. This extension is supported by newly implemented data models for workflow versions and runs, ensuring standardized and comprehensive metadata capture.

To facilitate integration with existing platforms, LabID-Prov supports importing from Git repositories, Galaxy instances, and WorkflowHub, while leveraging **Workflow and Workflow Run RO-Crate** specifications to simplify sharing and deposition in scientific repositories (WorkflowHub...). Interoperability is further strengthened through an API, a command-line utility, and a Python library, enabling automation and customization. For example, users can import a workflow from Git, enrich its metadata (e.g. license, authors) via LabID's interface, and export it to WorkflowHub, all within a unified ecosystem.

By unifying experimental and computational provenance, LabID-Prov guarantees that derived datasets retain the same level of contextual information as raw data, fostering reproducible computational research.

ONLY WORKSHOPS - Proposed interaction format

Alternative Track

1. From Harmonisation to Action(ability)

ONLY WORKSHOPS - Tentative audience

ONLY WORKSHOPS - Maximum number of participants

ONLY WORKSHOPS - Special technical requirements

Autoren: SCHOLTALBERS, Jelle (EMBL, Heidelberg / LabRise Consulting); MONFORT, Matthias (EMBL, Heidelberg); REZA, Nayeem (EMBL, Heidelberg); THOMAS, Laurent (EMBL Heidelberg); Dr. GIRARDOT, Charles (EMBL, Heidelberg)

Vortragende(r): THOMAS, Laurent (EMBL Heidelberg)

Sitzung Einordnung: TALK SESSION

Track Klassifizierung: HMC Conference 2026 Track Topics: 2. Software Interoperability for (Meta)data Acquisition