# OSCARS
## Open Science Clusters' Action for Research & Society

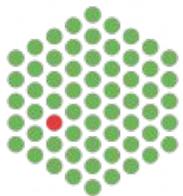# Funded Project

---

## LabID-PROV: Tracking and sharing data provenance
## with RO-Crate in Lab Integrated Data
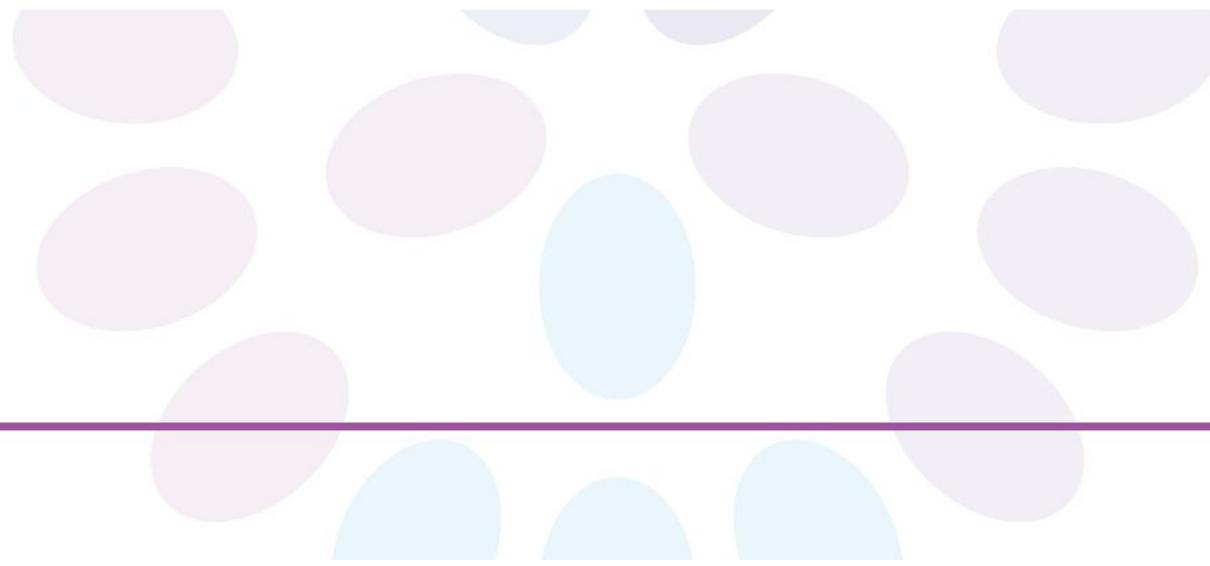
Lauriane THOMAS, EMBL Heidelberg, ORCID: 0000-0001-7661-1449

Implemented by

EMBL
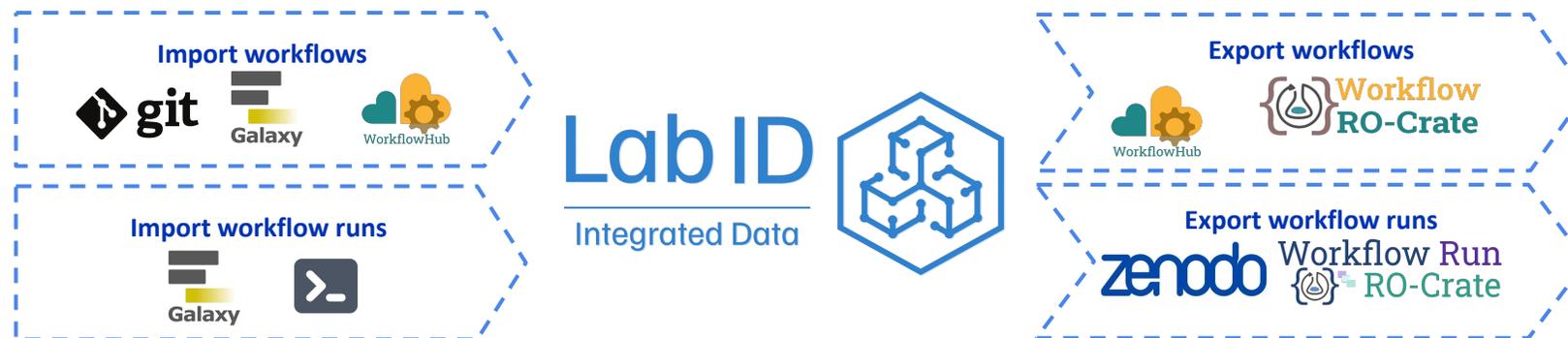
Workflows commonly used to process research data…

**BUT** sharing data resulting from workflow execution is not yet common practice

and when shared often **lacks comprehensive metadata (research context)**

> **How to simplify tracking and sharing of derived data with full data provenance?**

1) Extend the LabID data management platform to document **workflows**, **workflow runs** with **derived data**

2) Leverage the **Workflow profiles** of **Research Object Crate (RO-Crate)**
   - to automate ingestion of workflows and workflow runs
   - to share workflows and workflow runs with full provenance (WorkflowHub, Zenodo, ROHub)

3) Foster adoption by documenting **concrete use-cases from different disciplines** (genomics and imaging)



**Import workflows**
git  Galaxy  WorkflowHub

**Import workflow runs**
Galaxy

**Lab ID**
Integrated Data

**Export workflows**
WorkflowHub  Workflow RO-Crate

**Export workflow runs**
zenodo  Workflow Run RO-Crate

Status

| | |
|---|---|
| Public release of LabID-PROV supporting the management of Workflows and Workflow Runs | � are filled purple bars |
| Integration with workflow ecosystem (import/export from Git, Galaxy, WorkflowHub) | |
| Automated ingestion of workflow runs for concrete use-cases<br>- nf-core run as ro-crate<br>- image-conversion to OME Zarr | |
| Documentation and training material illustrating the LabID-PROV features | |
| Publication of example Workflow Run Crates including complete data provenance | |
| Training workshop for end-users | |

Public release of LabID-PROV supporting the management of Workflows and Workflow Runs

Integration with workflow ecosystem (import/export from Git, Galaxy, WorkflowHub)

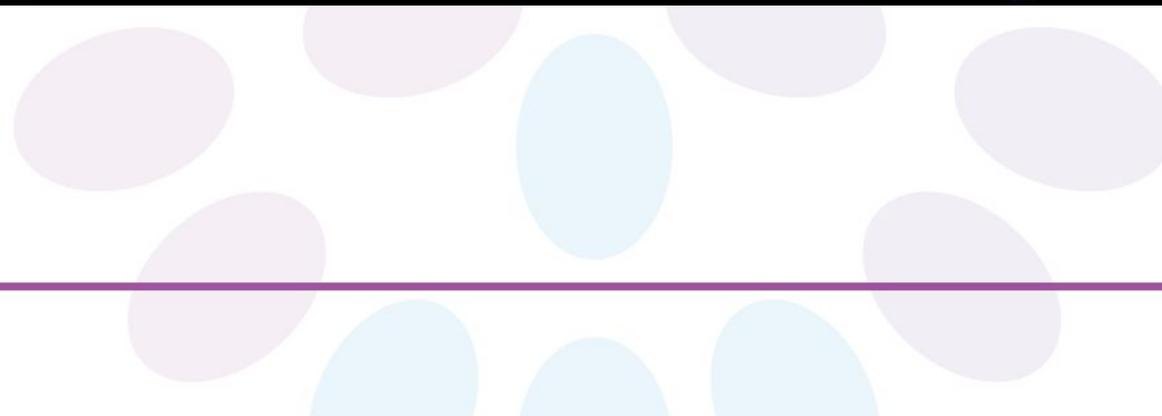Automated ingestion of workflow runs for concrete use-cases
- nf-core run as ro-crate
- image-conversion to OME Zarr

Documentation and training material illustrating the LabID-PROV features

Publication of example Workflow Run Crates including complete data provenance

Training workshop for end-users

## Software sustainability

LabID: established data-management platform across EMBL sites (5 countries)

Developed and maintained by **dedicated team** of the EMBL Data Science Center (MODIS)

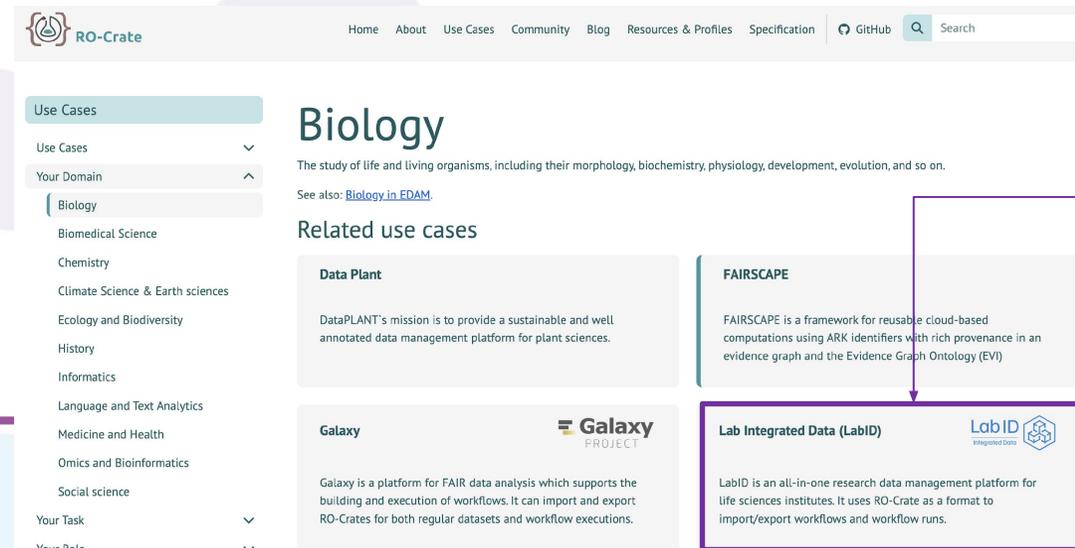Support service commercially available from **LabRise**

Discussion for follow-up collaborations and installations at other institutes

## Dissemination

Engagement and contributions to various communities (RO-Crate, Galaxy, deNBI)

Whitepaper describing features and use-cases

Referencing of software and training material in research repositories

RDMkit    elixir TeSS    RO-Crate



RO-Crate — Biology

The study of life and living organisms, including their morphology, biochemistry, physiology, development, evolution, and so on.

See also: Biology in EDAM.

### Related use cases

**Data Plant**

DataPLANT's mission is to provide a sustainable and well annotated data management platform for plant sciences.

**FAIRSCAPE**

FAIRSCAPE is a framework for reusable cloud-based computations using ARK identifiers with rich provenance in an evidence graph and the Evidence Graph Ontology (EVI)

**Galaxy**

Galaxy is a platform for FAIR data analysis which supports the building and execution of workflows. It can import and export RO-Crates for both regular datasets and workflow executions.

**Lab Integrated Data (LabID)**

LabID is an all-in-one research data management platform for life sciences institutes. It uses RO-Crate as a format to import/export workflows and workflow runs.

## Software adoption

LabID: established data-management platform across EMBL sites (5 countries)

Discussion for follow-up collaborations and installations at other institutes

Interactive documentation / demo server with concrete use-cases (ex: imaging, sequencing…)

Thematic training workshop(s)

## Software sustainability

Developed and maintained by dedicated team of the EMBL Data Science Center

Support service commercially available from LabRise

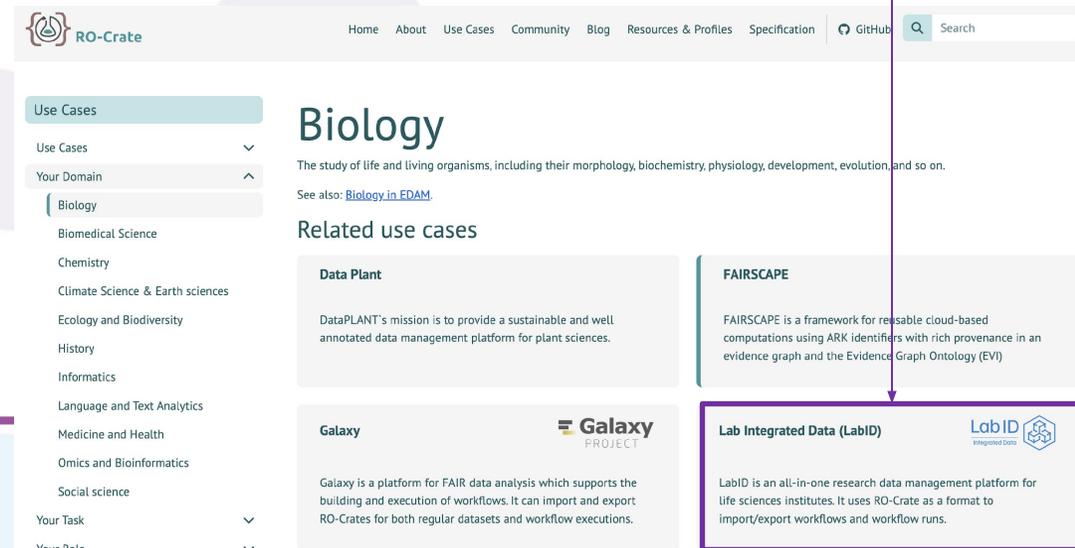Open-source software (MIT licensed, on gitlab.com)

Adoption of *FAIR for Research Software* (FAIR4RS) ex: source code archived on Zenodo

## Dissemination

Engagement and contributions to various communities (RO-Crate, Galaxy, deNBI)

Poster and talks at conference (Galaxy summit, HMC Heidelberg)

Whitepaper describing features and use-cases

Referencing of software and training material in research repositories

RDMkit  elixir TeSS  RO-Crate

**EMBL**

Prof. Eileen Furlong
Head of the Genome Biology dept and Senior Scientist at EMBL

## Multimodal Open Data Integration Support (MODIS) team

**Charles Girardot**
Head of Multimodal Open Data Integration Support
ORCID: 0000-0003-4301-3920

**Laurent Thomas**
Postdoctoral Fellow
ORCID: 0000-0001-7686-3249

**Matthias Monfort**
Web Developer / Workflow Management

**Md Nayeem Reza**
Senior Software Engineer
ORCID: 0000-0003-2068-5812

**Marco Wetter**
Full Stack Developer

**lab∧rise** CONSULTING

**Jelle Scholtalbers** · 1er
Freelancer / Full stack developer
France · **Coordonnées**
https://labrise-consulting.com
ORCID : 0000-0002-6090-2482