

LabID workflow integration

Documenting the computational provenance of “derived” data

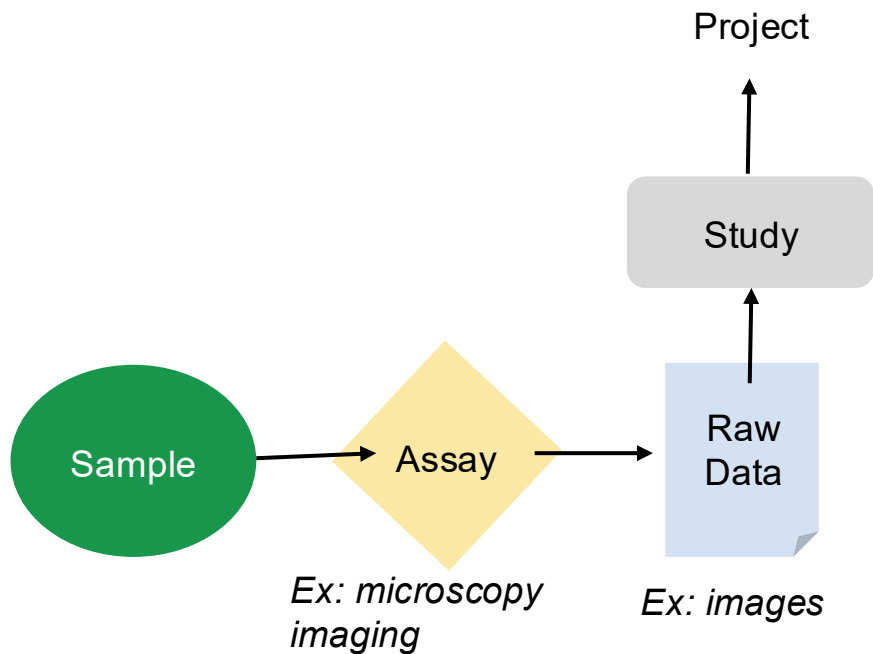
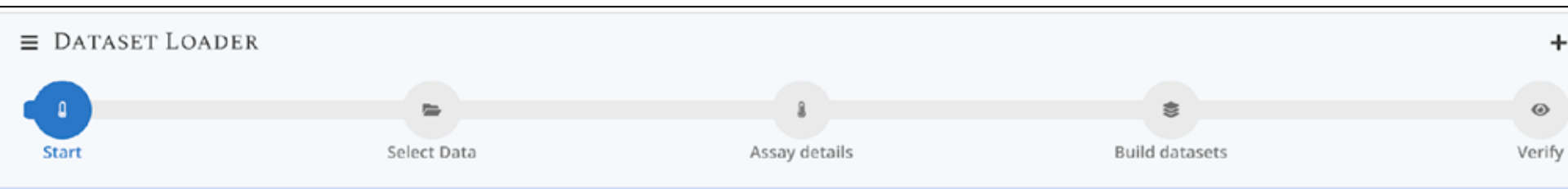
OSCARS Project :

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

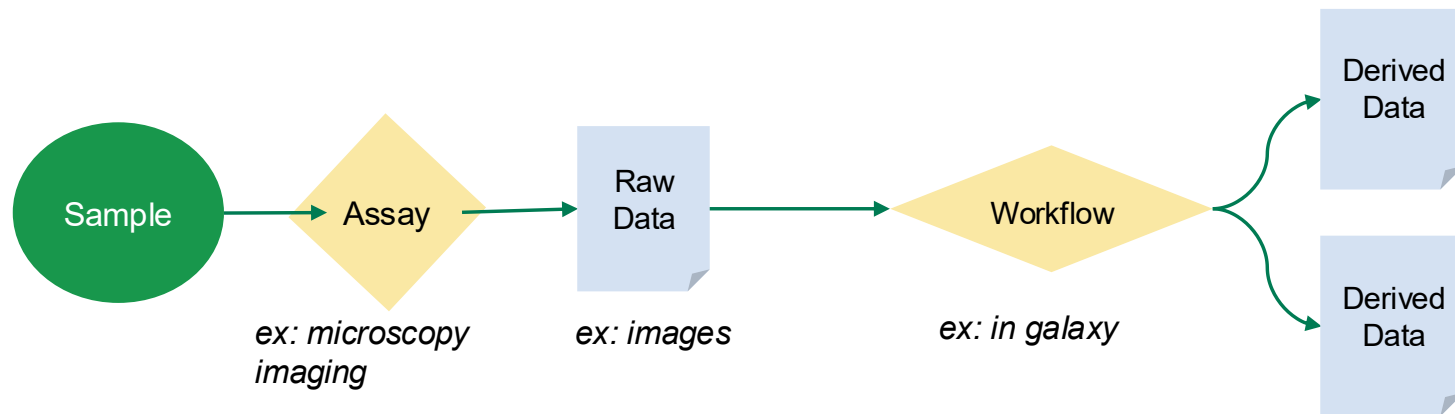


Laurent Thomas - 29/09/2026

Recording raw data provenance at data registration



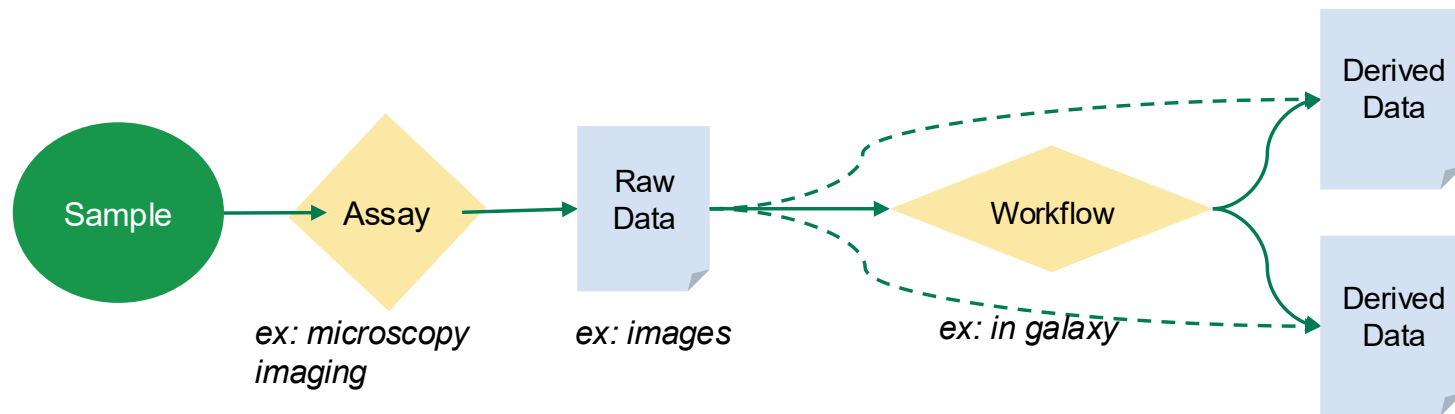
What about derived data ?



possible to link raw to derived data — — —

BUT missing all the information about the workflow execution / how the data was produced !

What about derived data ?

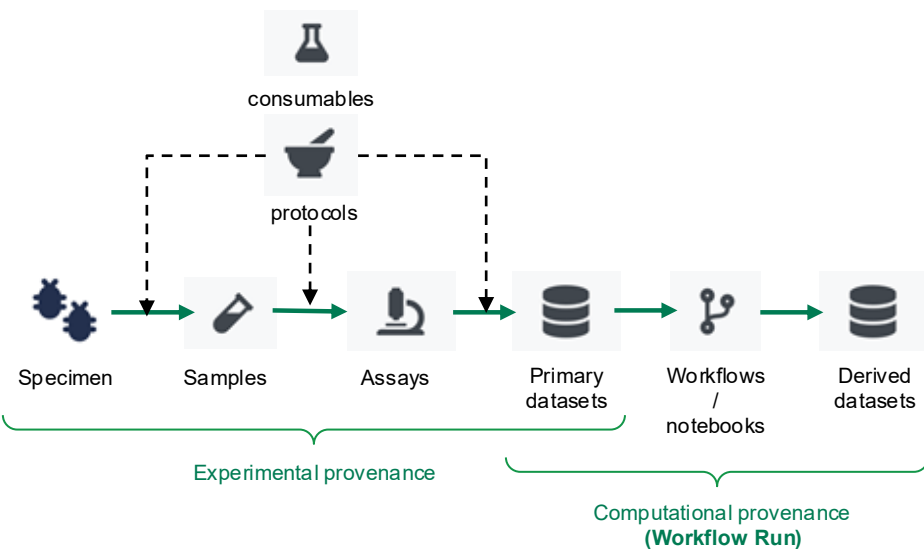


possible to link raw to derived data — — —

BUT missing all the information about the workflow execution / how the data was produced !

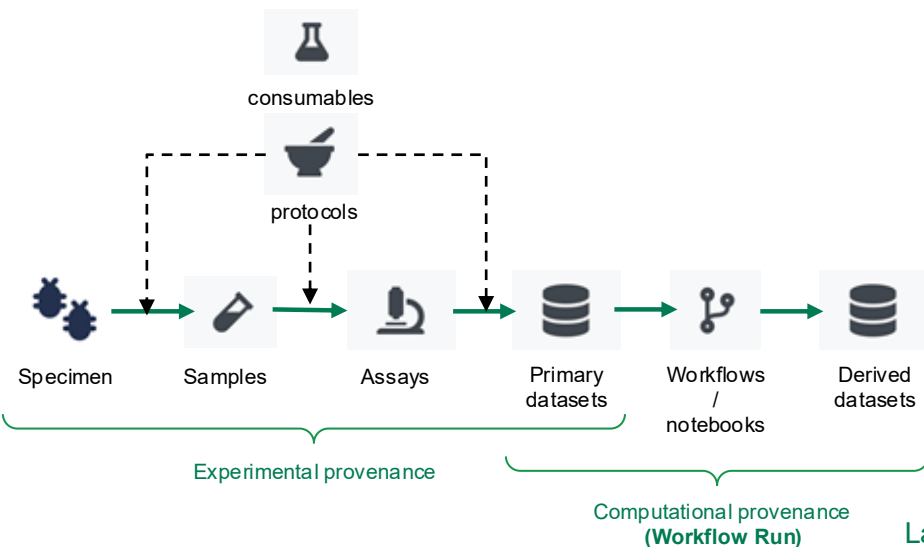
LabID workflow integration

A unified solution to document workflow execution and provenance of derived data



LabID workflow integration

A unified solution to document workflow execution and provenance of derived data



LabID-PROV project

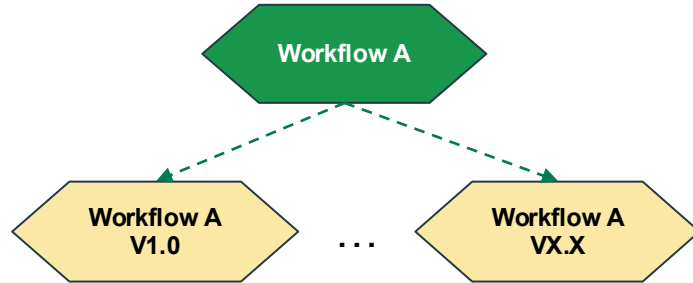
Extend LabID to :

- document workflow execution and provenance of derived data
- provide a **unified** and **interoperable** platform to **collect** and **share** the full data provenance
- leverage RO-Crate as a FAIR-compliant exchange format

Modelling workflow entities

Workflow

Workflow
Versions



Welcome, Laurent Sylvain Vincent
Thomas



v26.1.1 |

Import datasets ▾

- CONSUMABLES
- EQUIPMENT
- STORAGE EQUIPMENT
- SPECIMEN
- BIOMATERIALS
- PROTOCOLS
- LAB NOTEBOOK
- ASSAYS

- WORKFLOWS
 - Workflow
 - Workflow Version
 - Workflow Run

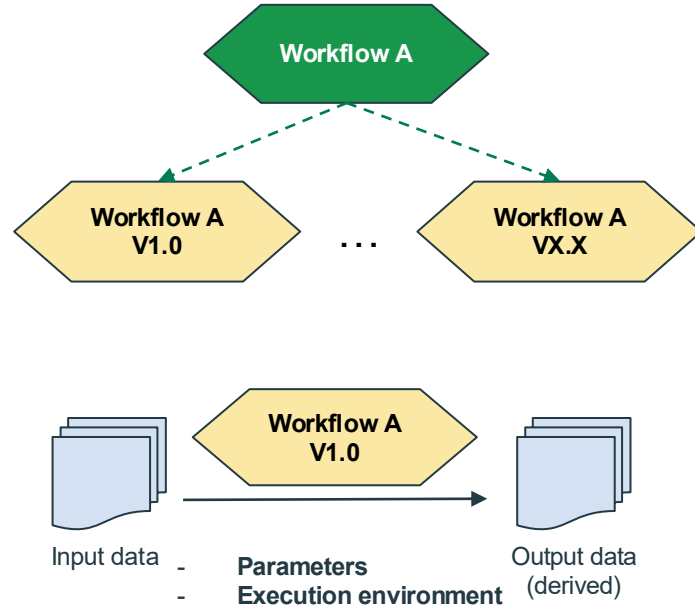
DATASET MANAGEMENT

Modelling workflow entities

Workflow

Workflow
Versions

Workflow
Run



Welcome, Laurent Sylvain Vincent
Thomas



v26.1.1 |

Import datasets ▾

- CONSUMABLES
- EQUIPMENT
- STORAGE EQUIPMENT
- SPECIMEN
- BIOMATERIALS
- PROTOCOLS
- LAB NOTEBOOK
- ASSAYS

- WORKFLOWS
 - Workflow
 - Workflow Version
 - Workflow Run

DATASET MANAGEMENT

Import, creation and export of workflows versions in LabID



Galaxy API / BioBlend



Import, creation and export of workflows versions in LabID



Import Workflow from usegalaxy.eu

Repository platform
usegalaxy.eu

Import from usegalaxy.eu URL Search https://usegalaxy.eu

Please provide a valid usegalaxy.eu workflow URL.

Import

Import, creation and export of workflows versions in LabID



Galaxy API / BioBlend



Import, creation and export of workflows versions in LabID



Galaxy API / BioBlend



Workflow RO Crate

Lab ID

Integrated Data



Import, creation and export of workflows versions in LabID



Import Workflow from RO-Crate

Repository platform

- ☒ workflowhub.eu
- ☐ dev.workflowhub.eu
- ☐ usegalaxy.eu
- ☐ galaxy.embl.de

Search <https://workflowhub.eu>

The search targets the [name](#) of each workflow. You can refine your search by using metadata modifiers, such as [description:workflow](#) or [author:johs.doe](#).

segmentation

Nucleoli segmentation using CellProfiler (EOSC-Life D6) ID: 41

This workflow has been created as part of Demonstrator 6 of the project EOSC-Life (within WP3) and is focused on reusing publicly available RNAi screens to gain insights into the nucleus...

Delineating Regions-of-interest for Mass Spectrometry Imaging by Multimodally Corroborated Spatial Segmentation ID: 437

ROIforMSI
Source codes for manuscript "Delineating Regions-of-interest for Mass Spectrometry Imaging by Multimodally Corroborated Spatial Segmentation"
https://doi.org/10.1101/111111

fluorescence-nuclei-segmentation-and-counting/main ID: 771

This workflow performs segmentation and counting of cell nuclei using fluorescence microscopy images. The segmentation step is performed using Otsu thresholding (Otsu, 1979). The workflow is based on...

Import, creation and export of workflows versions in LabID



Galaxy API / BioBlend



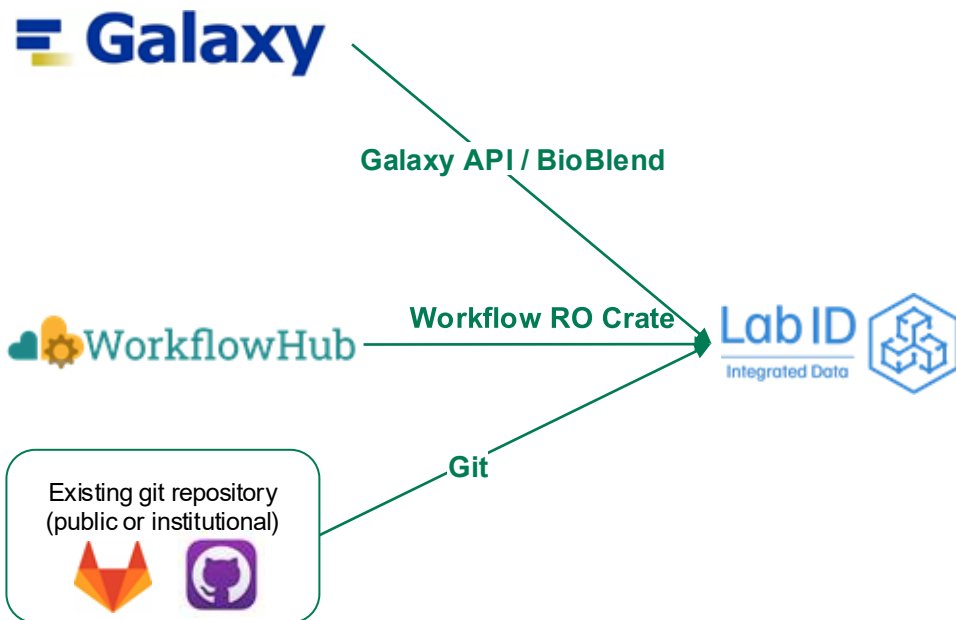
Workflow RO Crate

Lab ID

Integrated Data



Import, creation and export of workflows versions in LabID



Import, creation and export of workflows versions in LabID



Existing git repository
(public or institutional)



Import a workflow from a Git repository

Provide a valid Git repository URL, and select a version commit

Repository URL

Workflow version (commit)

Bump version: 2.4.0 → 2.5.0

friendsofstrandseq/mosaiccatcher-pipeline SNAKEMAKE 5.25

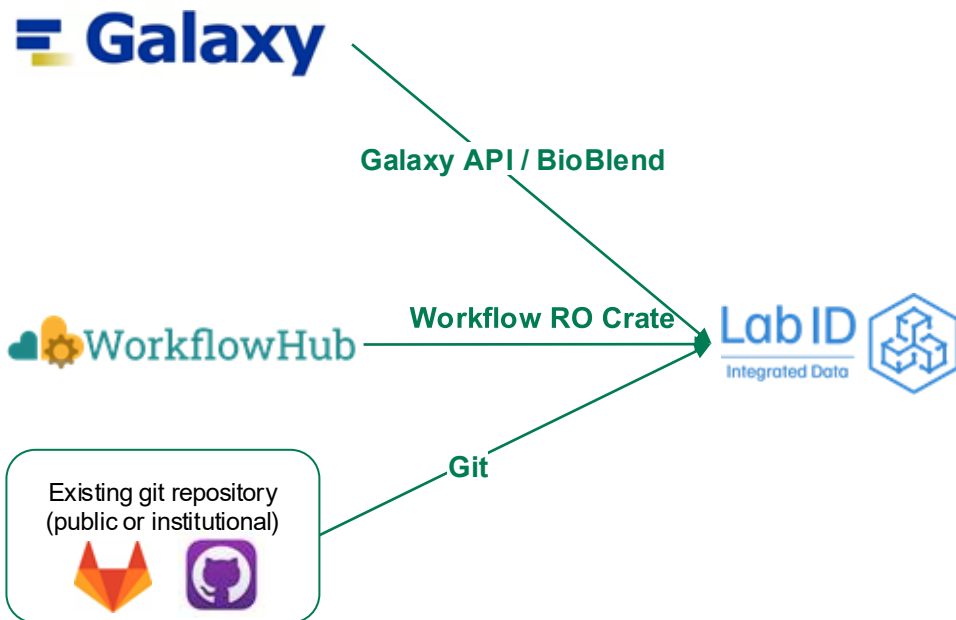
Workflow Name

Workflow Files Information

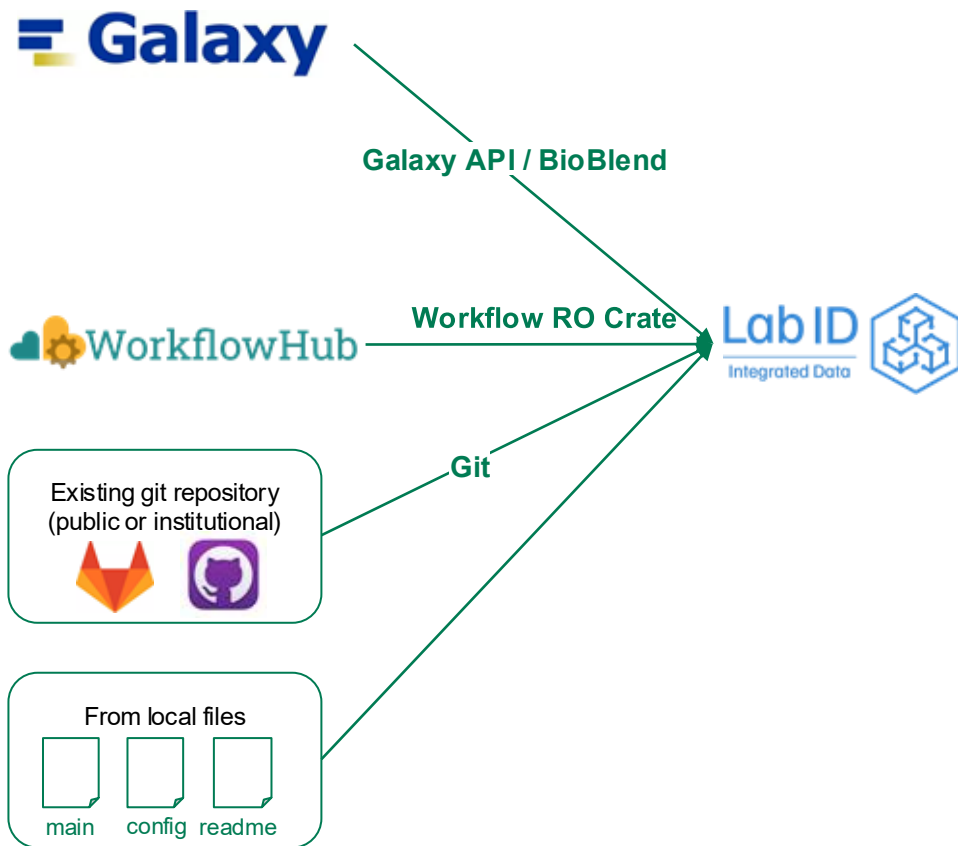
Add and type workflow files for this workflow version. Workflow files are the essential set of files that define the workflow, describe it or are necessary for its execution.

Files	File Types
✓ README.md	README
✓ LICENSE	License
✓ afac/update_sm_tag_pipeline/snakefile	Main
✓ afac/update_sm_tag_pipeline/config.yaml	Config
✓ config/config.yaml	Config
✓ dev/snakemake-v9-features/config.yaml	Config
✓ dev/snakemake-v9-features/profile/config.yaml	Config

Import, creation and export of workflows versions in LabID



Import, creation and export of workflows versions in LabID



Import, creation and export of workflows versions in LabID

The screenshot displays the LabID interface for managing workflow versions. The top bar shows the version identifier "Version 2026-03-19T12:31:04.794Z" and buttons for "Save & exit" and "Cancel".

The main content area is divided into two panels. The left panel, titled "Item Details", contains a form for creating or editing a workflow version. The form includes fields for ID, NAME, DESCRIPTION, WORKFLOW, COMPUTER LANGUAGE, LICENSE, PUBLISHED AS, WORKFLOW NAME, COMMENT NAME, COMMENT DATE, and IS LATEST. The right panel contains sections for "Annotations" and "Attachments".

The "Item Details" form fields are as follows:

- ID: `5586363-646f-4c81-857e-9a11525c20af`
- NAME: Version 2026-03-19T12:31:04.794Z
- DESCRIPTION:
- WORKFLOW: my test wf
- COMPUTER LANGUAGE: KNIME (KNIME workflow language)
- LICENSE: MIT (MIT License)
- PUBLISHED AS:
- WORKFLOW NAME:
- COMMENT NAME:
- COMMENT DATE:
- IS LATEST: ☒ Yes ☐ No

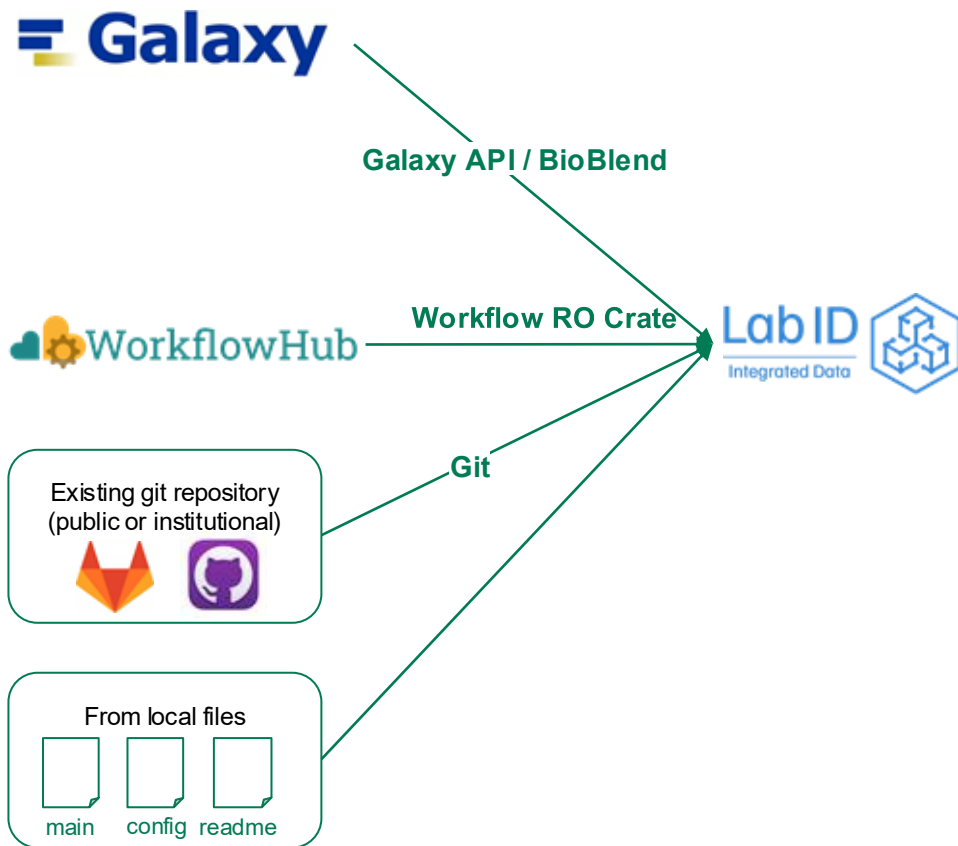
The "Annotations" section has a dropdown menu to "Pick an annotation type to add". The "Attachments" section has a "Drag & Drop or Browse files (max 500MB)" area.

The "Ownership and Lifecycle" and "Sharing" sections are collapsed. The "Add files" section is expanded, showing a "Drag & Drop your workflow files or Browse local files" area. Below this, a list of files is shown, including "test_rocrate.py" added by "lathomas" 6 days ago.

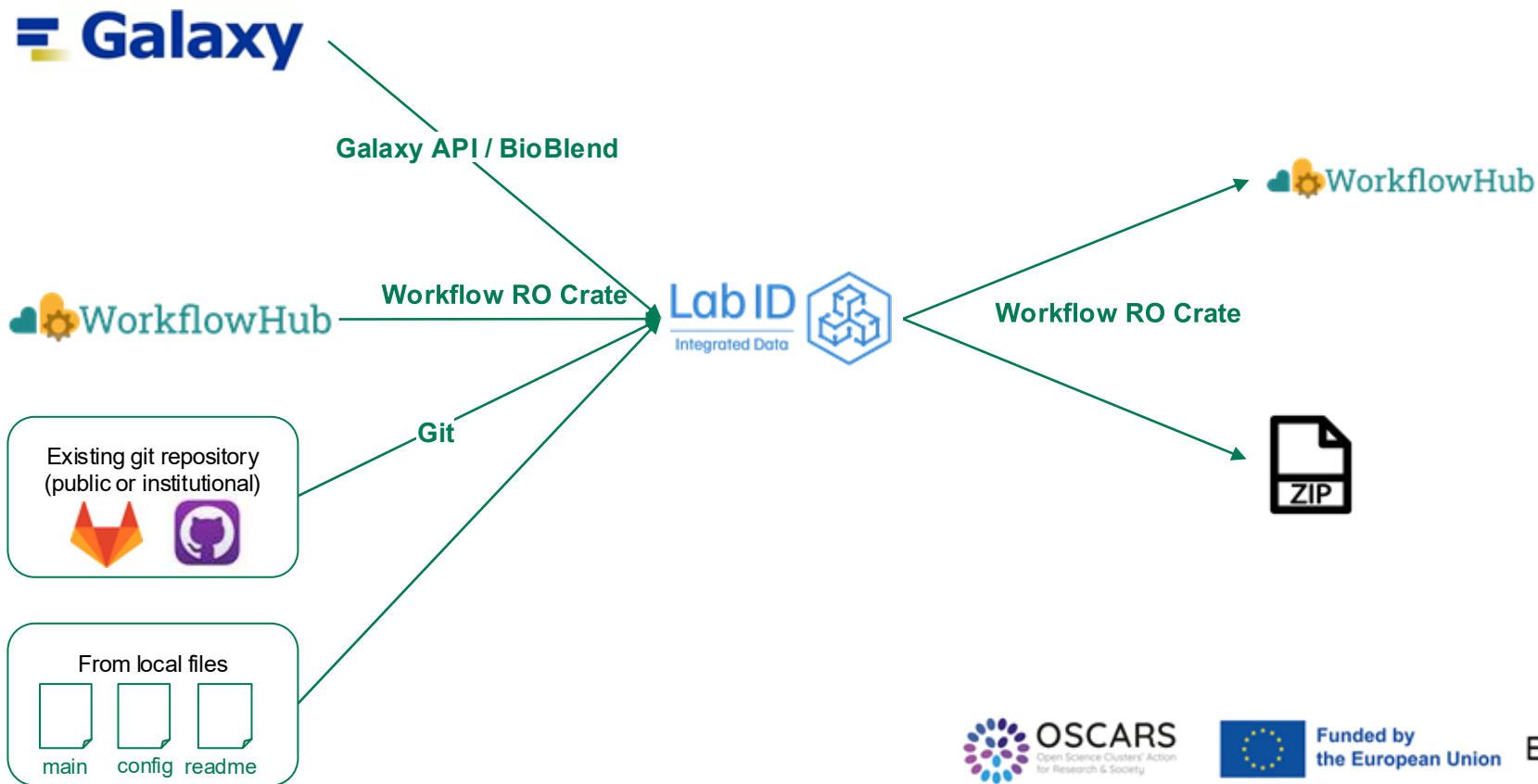
On the left side, there are icons for "Existing (public)" and "From". Below these, there are three icons labeled "main", "config", and "readme".

On the right side, a dropdown menu is open, showing options: "Other", "Config", "DAG PNG", "Input", "License", "Log", "Main", and "Select workflow file type". The "Other" option is selected.

Import, creation and export of workflows versions in LabID



Import, creation and export of workflows versions in LabID

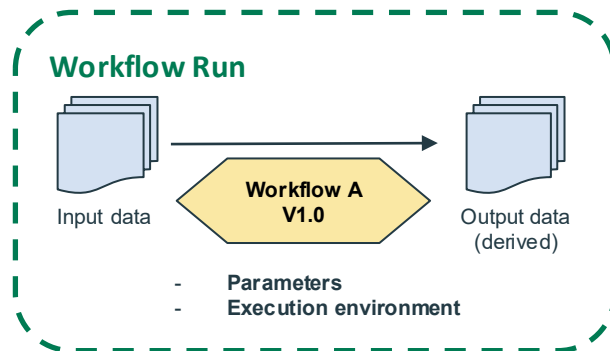


Importing/creating workflow runs in LabID

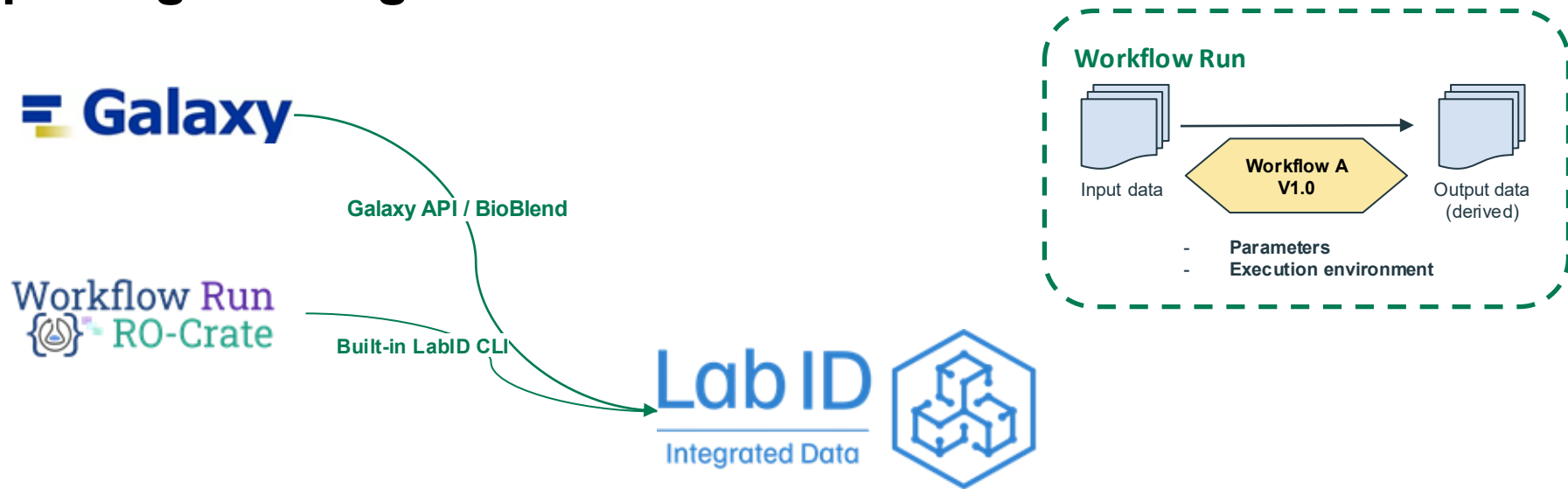


Galaxy API / BioBlend

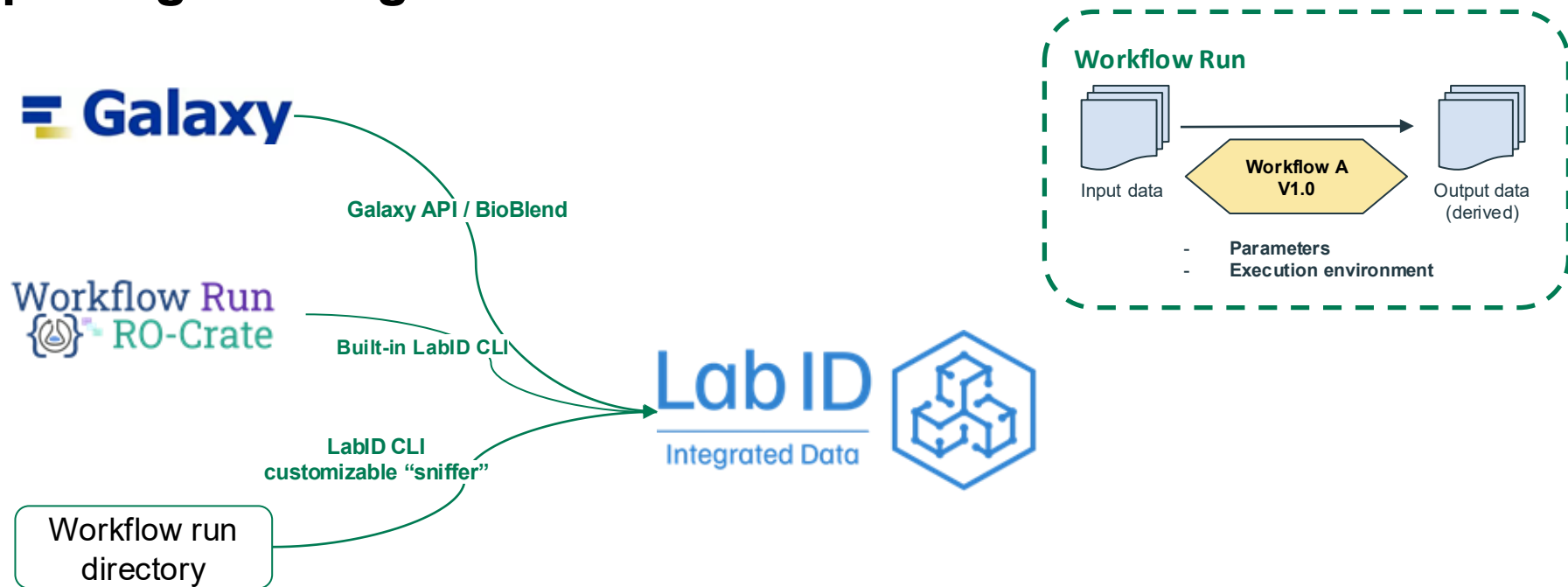
Lab ID
Integrated Data



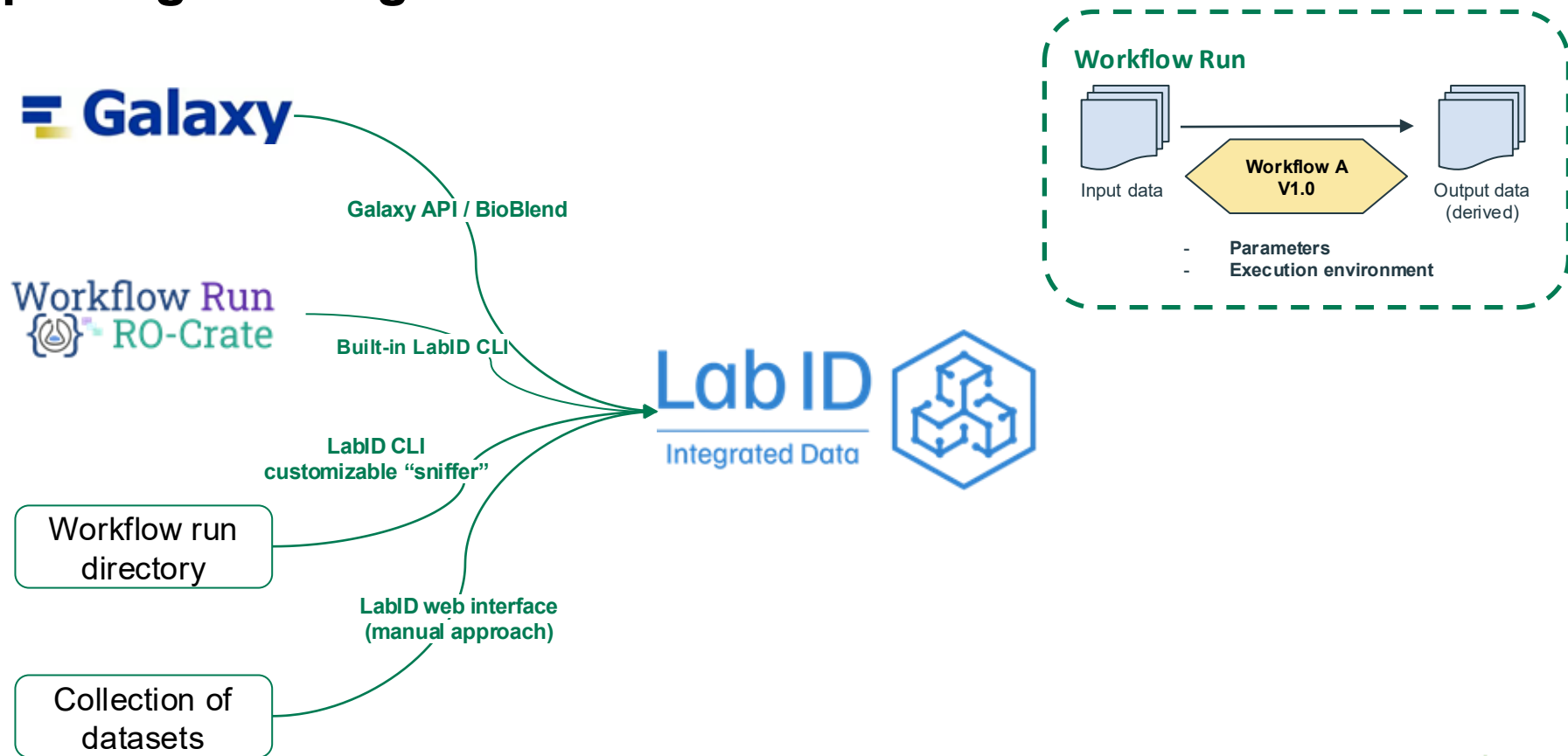
Importing/creating workflow runs in LabID



Importing/creating workflow runs in LabID



Importing/creating workflow runs in LabID



Example : image-conversion as Workflow runs

WORKFLOW RUN

conversion_czi_to_zarr_20260121_162346_20260121_162405

Details Command Line Config Inputs Outputs Reports Logs

Success

ID 856883fb-b506-45ea-a3bc-29cb33a3893d

Name conversion_czi_to_zarr_20260121_162346_20260121_162405

Description —

Workflow Euro-Biolmaging/EuBI-Bridge

Workflow Version Version jdk-v11 (4b2d1f7)

Workflow Manager OTHER

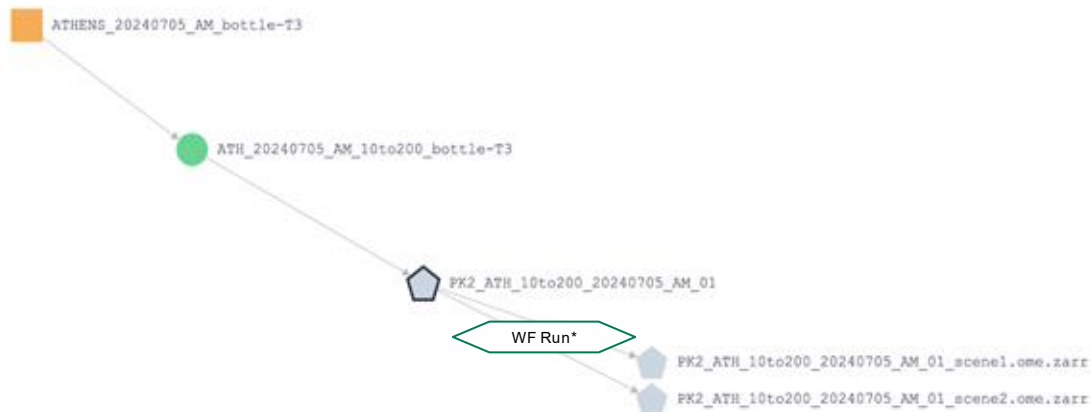
Status SUCCESS

Started at —

Completed at —

Is Imported ✗ no

Is Imported ROCrate ✗ no



(* not yet depicted in the current version of LabID)

- ✓ Capture OME-Zarr conversion metadata e.g. tool version, parameters and execution environment
- ✓ Complete provenance information, up to original biological material

Perspective: **Automate** image conversion to OME-Zarr **upon registration of raw data** (czi image)

Example workflow run (Nextflow)

Workflow version executed

Workflow run details for `scnaseq-rocrate_20250818_172055`.

Workflow: `nf-core/scnaseq`
Workflow Version: `Version 4.0.0 (af0dddb8)`
Workflow Manager: NEXTFLOW
Status: SUCCESS
Started at: 2025-03-18 11:45:05
Completed at: —
Is Imported: ☒ no
Is Imported ROcRate: ☒ no

Annotations: 0
Notes: 2

Notes:

- The run was executed with the `nf-prov` plugin to generate a ROcRate (specified with a custom config file).
- command to execute the workflow `"nextflow run nf-core/scnaseq -profile test,docker -c custom.config --outdir / + 4.0.0"`

Data used as input, config and generated by the run (output, logs...)

Outputs

Dataset	Type	Owner	Workflowrun	Owned by Group	Public
<code>Sample_X_1_fastq.html</code>	Output	@lathomas	<code>scnaseq-rocrate_20250818_172055</code>	Computational Support Genome Biology	✓ yes
<code>Sample_X_2_fastq.html</code>	Output	@lathomas	<code>scnaseq-rocrate_20250818_172055</code>	Computational Support Genome Biology	✓ yes
<code>Sample_Y_raw_matrix.seurat.rds</code>	Output	@lathomas	<code>scnaseq-rocrate_20250818_172055</code>	Computational Support Genome Biology	✓ yes
<code>Sample_Y_3_fastq.zip</code>	Output	@lathomas	<code>scnaseq-rocrate_20250818_172055</code>	Computational Support Genome Biology	✓ yes
<code>versions.yml</code>	Output	@lathomas	<code>scnaseq-rocrate_20250818_172055</code>	Computational Support Genome Biology	✓ yes
<code>Sample_X_S0_out.tab</code>	Output	@lathomas	<code>scnaseq-rocrate_20250818_172055</code>	Computational Support Genome Biology	✓ yes
<code>Sample_Y_filtered_matrix.seurat.rds</code>	Output	@lathomas	<code>scnaseq-rocrate_20250818_172055</code>	Computational Support Genome Biology	✓ yes
<code>versions.yml</code>	Output	@lathomas	<code>scnaseq-rocrate_20250818_172055</code>	Computational Support Genome Biology	✓ yes
<code>Sample_Y_2_fastq.html</code>	Output	@lathomas	<code>scnaseq-rocrate_20250818_172055</code>	Computational Support Genome Biology	✓ yes

How to share/export workflow runs including all the provenance ?



Tabular / MAGE-TAB / XML (ENA-EGA)

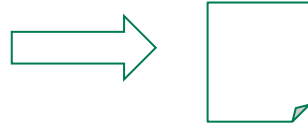
- simple but limited support for derived data
- no widely adopted standard (mainly suitable for genomics)



- **Graph** of data-provenance
(biological and computational)
- arbitrary metadata



How to share/export workflow runs including all the provenance ?



Tabular / MAGE-TAB / XML (ENA-EGA)

- simple but limited support for derived data
- no widely adopted standard (mainly suitable for genomics)



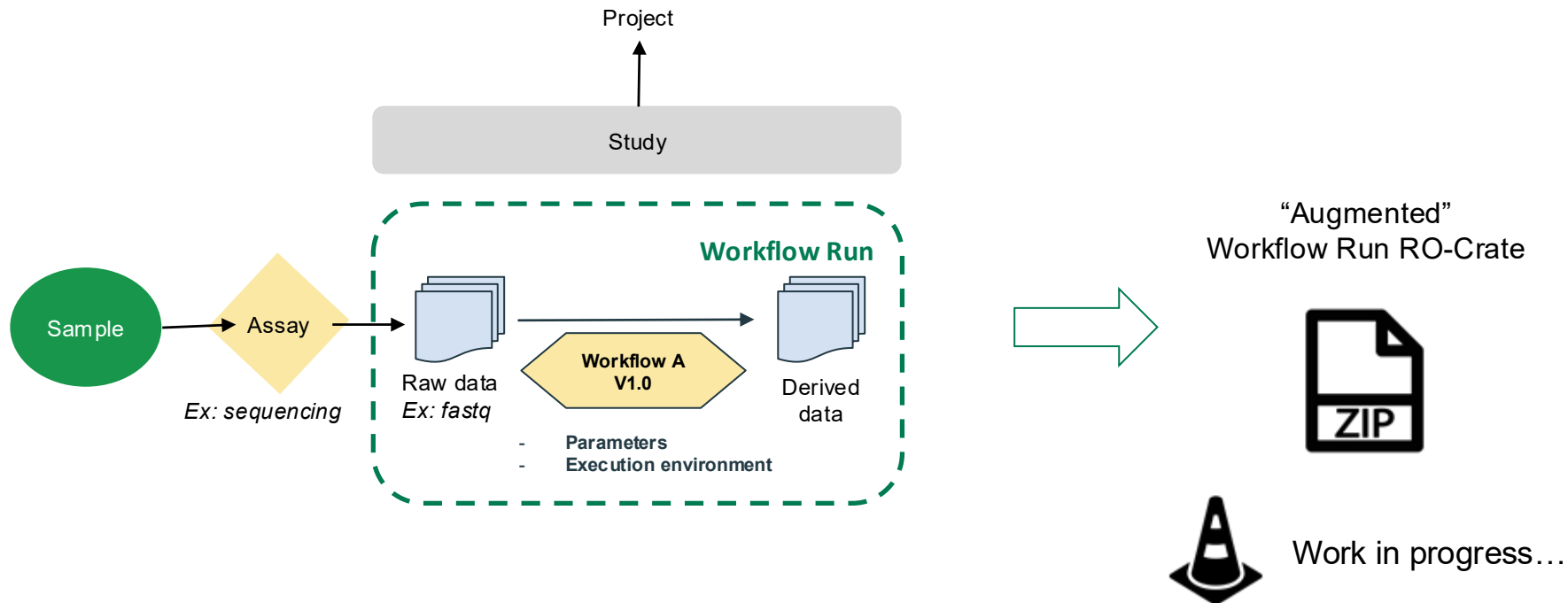
- **Graph** of data-provenance
(biological and computational)

- arbitrary metadata

- JSON describing a **graph** of **data** and **context entities**
- **standardised** with ontologies / controlled vocabularies
- **interoperable** / **machine-actionable** (indexing, search)
- **extensible**
ex : ISA-RO-Crate specs to document biological provenance (assay, sample prep.)
- already **adopted by other platforms and communities** (WorkflowHub, Nextflow, Galaxy, ARC...)

Exporting workflow runs from LabID

- LabID generates a custom **Workflow Run RO Crate** (unless imported as RO-Crate)
- Including **computational AND biological data-provenance** (e.g sample LabID identifiers of exported entities)
- Can be published to e.g Zenodo, RO-Hub...



1-click export of Workflow Run RO-Crate

- agnostic of workflow engine

Example Run

Edit

RO-Crate

Delete

Details

Command line

Configs

Inputs

Outputs

Logs & Reports

Success

ID

81e9a278-9ce1-4dbe-b591-8e72e6f8a6f4

Name

Example Run

Description

—

Workflow

Assembly-Hifi-Trio-phasing-VGP5/main

Workflow Version

Assembly-Hifi-Trio-phasing-VGP5/main: 0.9.2

Workflow Manager

GALAXY

Status

SUCCESS

Started at

2025-05-22 14:19:04

Completed at

—

Is Imported

✗ no

Annotations 0

Pick an annotation type to add

Notes 0

Write a note

Add

Outlook - Workflow scheduling

Use LabID to trigger workflow runs and automatically register them

1. Select workflow and input data
2. Configure parameters



Conceptually simple but technically challenging....

Outlook - Workflow scheduling

Use LabID to trigger workflow runs and automatically register them

1. Select workflow and input data
2. Configure parameters



3. Send for execution



Conceptually simple but technically challenging....

Outlook - Workflow scheduling

Use LabID to trigger workflow runs and automatically register them

1. Select workflow and input data
2. Configure parameters



3. Send for execution



4. Register data and toggle
run status (success/failed)

Conceptually simple but technically challenging....

LabID workflow integration - summary

A unified solution to document workflow execution and provenance of derived data

- Document **computational provenance** for derived data, with same standard as raw data
- **Unified platform** to document workflow and script execution, regardless of the workflow/programming language
- End-to-end provenance (specimen to derived data)
- Encouraging FAIR practices around workflows (versioning, sharing...)
- Interoperable (workflow repositories and engines)



Lab ID
Integrated Data

