

# Improving Galaxy provenance export using RO-Crate



Paul De Geest<sup>1,2</sup>, Frederik Coppens<sup>1,2</sup>, Stian Soiland-Reyes<sup>3,4</sup>, Ignacio Eguinoa<sup>1,2</sup>, Simone Leo<sup>5</sup> and the RO-Crate community<sup>6</sup>.

1. Department of Plant Biotechnology and Bioinformatics, Ghent University, 9052, Ghent, Belgium.
2. VIB Center for Plant Systems Biology, 9052, Ghent, Belgium.
3. Department of Computer Science, The University of Manchester, Manchester, United Kingdom
4. Informatics Institute, University of Amsterdam, Amsterdam, Netherlands
5. Center for Advanced Studies, Research, and Development in Sardinia (CRS4), Pula (CA), Italy
6. <https://www.researchobject.org/ro-crate/community.html>

## Galaxy

The **Galaxy** project provides a multi-user research environment for reproducible data analysis and workflow management by keeping track of all analytical steps and any related metadata as part of a history and allowing all of this to be exported and imported. However, these features are currently limited to the context of Galaxy instances. Therefore, the next natural step is to integrate the FDO concepts with Galaxy export capabilities and thereby increasing FAIRness and allowing for archiving, reproducing, publishing and sharing data analysis across platforms and repositories.

## RO-Crate

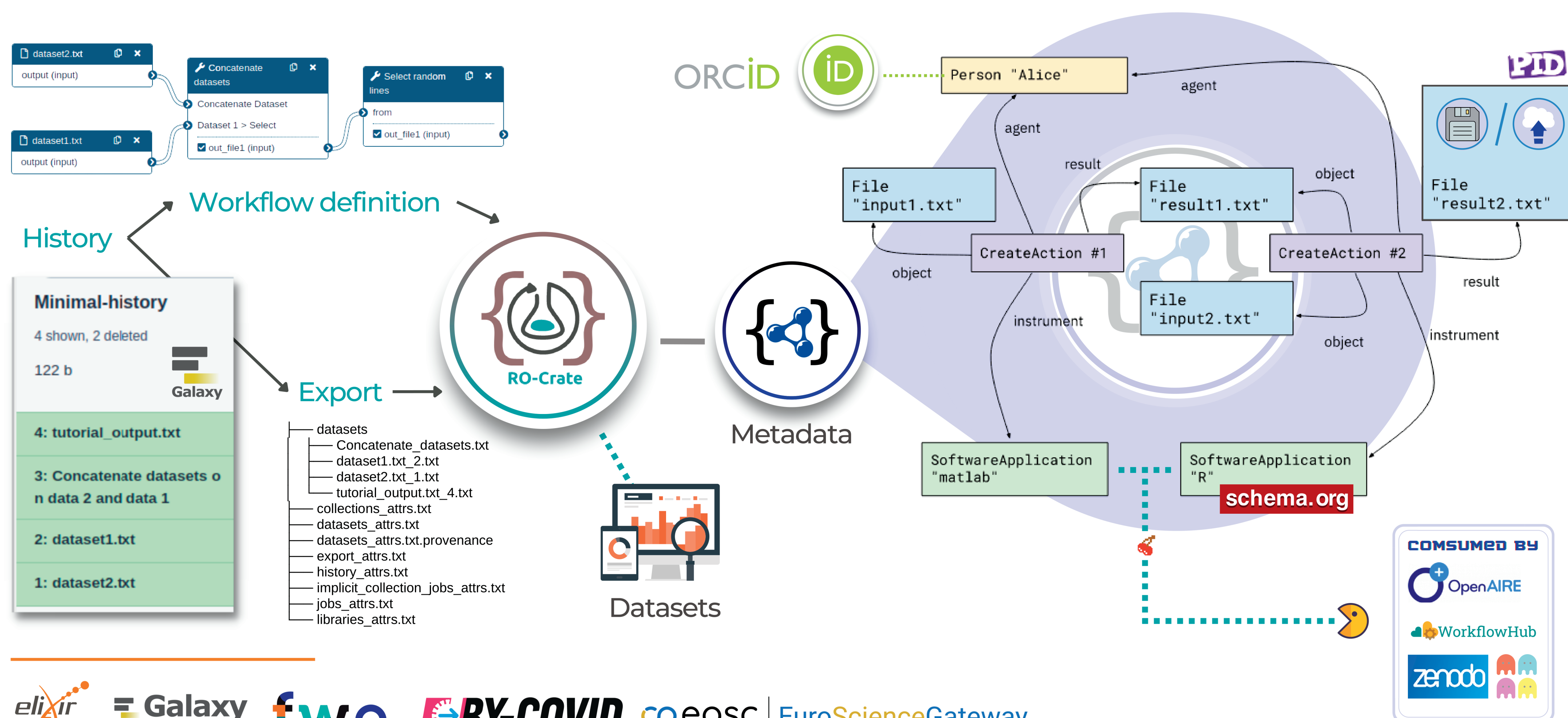
**RO-Crate** is a generic packaging format containing datasets and their description using standards for FAIR Linked Data. The format is based on schema.org annotations in JSON-LD, which allows for rich metadata representation representing (data-) artefacts as external references when necessary (FDO poster by Soiland-Reyes). The format has been implemented in wide range of services such as WorkflowHub and others, making it an interesting format to implement FDOs export features in Galaxy.

## Workflow Run RO-Crate

Towards this goal we are co-developing Workflow Run RO-Crate profile to store workflow execution data and metadata, including prospective and retrospective provenance, as well as associated tooling for its generation.

## Packaging a Galaxy history

We package a Galaxy history (set of executed analysis steps) as a Workflow Run RO-Crate by extracting the workflow steps definitions (prospective provenance) and the history contents (retrospective provenance) and combine these through tooling to fill out the Workflow Run RO-Crate profile and finally package these items with related (referenced) datasets in an RO-Crate.



## CONTACT

RO-Crate community:  
<http://researchobject.org/ro-crate>

ELIXIR Belgium:  
[info@elixirbelgium.org](mailto:info@elixirbelgium.org)

## FUNDING



This project has received funding from BY-COVID (EU Horizon Europe, grant agreement no. 101046203), FWO (grant agreement no. 1002819), EuroScienceGateway (EU Horizon Europe, grant agreement no. 101057388) and the SVDC project, funded by the Sardinian Regional Authority.