

# **Bioinformatics Workflows: Representation and Query**



Mouna El Garb, Emmanuel Coquery, Fabien Duchateau, Nicolas Lumineau

Laboratoire d'InfoRmatique en Image et Systèmes d'information LIRIS UMR 5205 CNRS / INSA de Lyon / Université Claude Bernard Lyon 1 / Université Lumière Lyon 2 / Ecole Centrale de Lyon

#### Context

**ShareFAIR** project aims to develop a biomedical data analysis platform for the sharing and reuse of scientific workflows, with a query engine designed to address the findable aspect of FAIRness.

#### **Scenario**

From the user's query to relevant workflows.



### State of the art

Overview of the features of the BioFlow-Model, compared to the CWL and P-Plan models, and WRROC profiles of RO-Crate.

Category	Subcategory	CWL	P-Plan	WRROC	<b>BioFlow-Model</b>
Scientific context	Workflow design	~	~	~	~
	Entity annotations	~	~	~	~
	Workflow execution ann.	-	~	~	~
Data	Data identification	~	~	~	~
	File characteristics	~	2	~	~
	Data access	~	-	~	~
	Parameter mapping	~	~	~	~
Software	Software identification	~	-	~	~
	Software documentation	~	-	~	~
	Software access	~	-	~	~
Workflow	Workflow software	~	-	~	~
	Workflow parameters	~	~	~	~
	Workflow requirements	~	-	~	-
Computational	Software environment	-	-	-	-
environment	Hardware environment	-	-	-	-
	Container image	~	-	~	~
Execution	Execution timestamps	-	~	~	~
detail	Consumed resources	-	1.52	~	~
	Workflow engine	12	-	~	~
	Human agent	-	~	~	~
Control flow	1775) -	-	_		~

## **Problematic**

- How to sufficiently **describe** scientific workflows for reuse and sharing ?
- How to write and build queries to retrieve relevant workflows?

**Legend:** ✓ fully represented, ~ partially represented, - not covered

# **BioFlow-Model**

A structured model to describe workflows using metadata and provenance in a generic and workflow engine independent format.



## **Example of query with BioFlow-Model**

Query: Find the workflows having an RNA-Seq step followed by a **Sequence analysis** step.

• Graph of the workflow matching to the query:



**Perspectives Experimental validation of BioFlow-Model** 

- Rewrite the queries to extract workflows that closely match the user's query
- Rank the workflows returned by query execution based on their relevance

