

Bioinformatics Workflows: Representation and Query



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

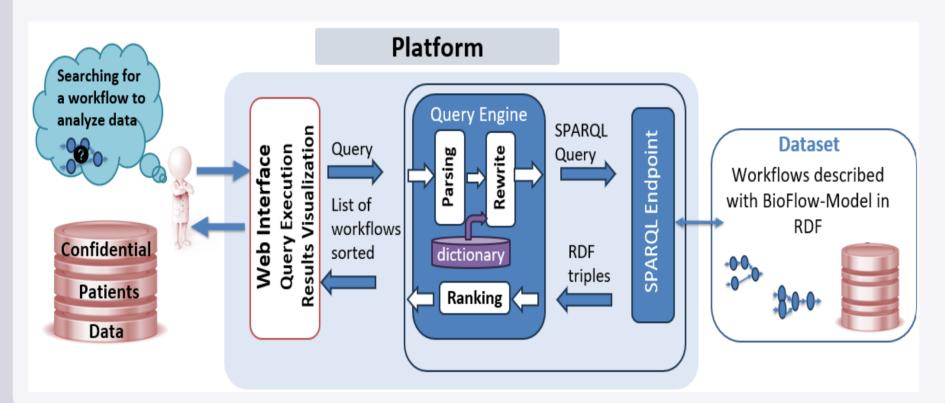
Laboratoire d'InfoRmatique en Image et Systèmes d'information LIRIS UMR 5205 CNRS / INSA de Lyon / Université Claude Bernard Lyon 1 / Université Lumière Lyon 2 / Eccole Centrale de Lyon

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 by enabling the retrieval of the workflows.

Scenario

From the user's query to the retrieval of relevant workflows.



Problematic

- How to sufficiently describe scientific workflows for reuse and sharing?
- How to build queries to retrieve the workflows?

State of the art

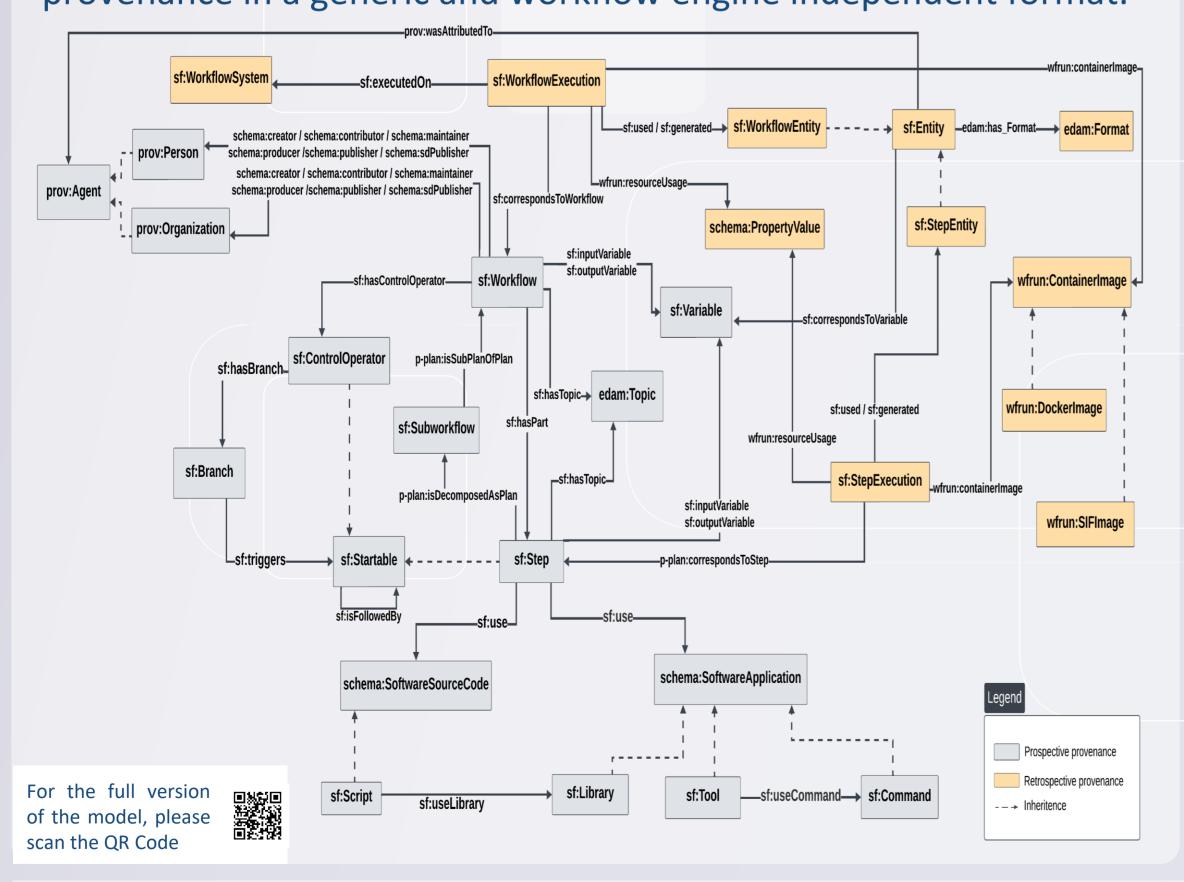
Overview of the features of the BioFlow-Model, compared to the CWL and P-Plan models, and WRROC profiles of RO-Crate.

Category	Subcategory	CWL	P-Plan	WRROC	BioFlow-Model
Scientific context	Workflow design	~	~	~	~
	Entity annotations	~	~	-	~
	Workflow execution ann.	-	~	-	~
Data	Data identification	~	~	-	~
	File characteristics	~	-	~	~
	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	-	-	-	~
	Workflow engine	-	-	~	~
	Human agent	-	~	~	~
Control flow		-	-	-	~

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 queries with BioFlow-Model

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

FIND wf

FROM Workflow wf, Step step1, Step step2

WHERE step1 IN wf

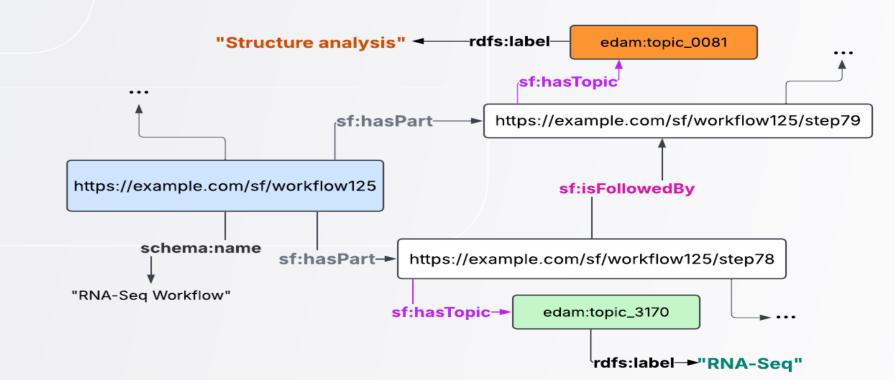
AND step2 IN wf

AND step1.topic XLIKE "RNA-Seq"

AND step2.topic XLIKE "Structure analysis"

AND step1 ISJUSTBEFORE step2;

Graph of the workflow matching to the query:



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











