

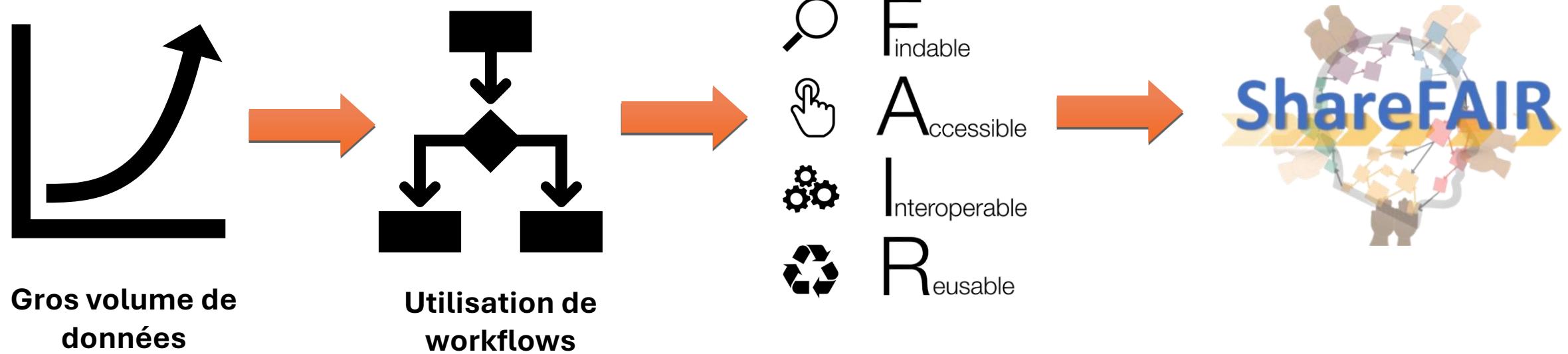


Conception d'une interface web de visualisation de workflows et de leurs métadonnées

Salma EL AOUDATI

Dr. Emmanuel COQUERY – Dr. Fabien DUCHATEAU

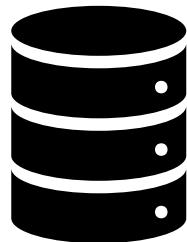
Contexte scientifique



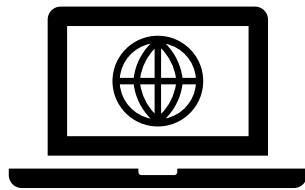
Sarah Cohen-B., and al. Scientific workflows for computational reproducibility in the life sciences : Status, challenges and opportunities. Future Generation Computer Systems, 75 :284–298, 2017.

Casper Visser, and al. Ten quick tips for building fair workflows. PLOS Computational Biology, 19 :e1011369, 09 2023

Objectifs du stage



Modélisation des workflows en base de données



Développement d'une interface web de visualisation des workflows

Systèmes de workflows

Il existe différents systèmes de gestions de workflow :

- **Snakemake**
- **Nextflow**
- Constelab
- SAGA
- YAWL
- ...

Etat de l'art sur les représentations des workflows

Il existe différents standards de représentation des workflows :

- **CWL**
- **WDL**
- **EDAM**



Besoin d'un métamodèle



Ontologie CWL. Common Workflow Language (CWL) Workflow Description, v1.1 (commonwl.org)

Ontologie WDL. https://docs.openwdl.org/en/latest/WDL/Base_structure/

Ontologie EDAM. EDAM: Ontology of bioscientific data analysis and data management (edamontology.org)

Conception du métamodèle

	CWL	EDAM	Nouveau
# classes	2	1	11
# prédictats	1	1	45
# mapping de classes et prédictats	10	12	34

61 concepts

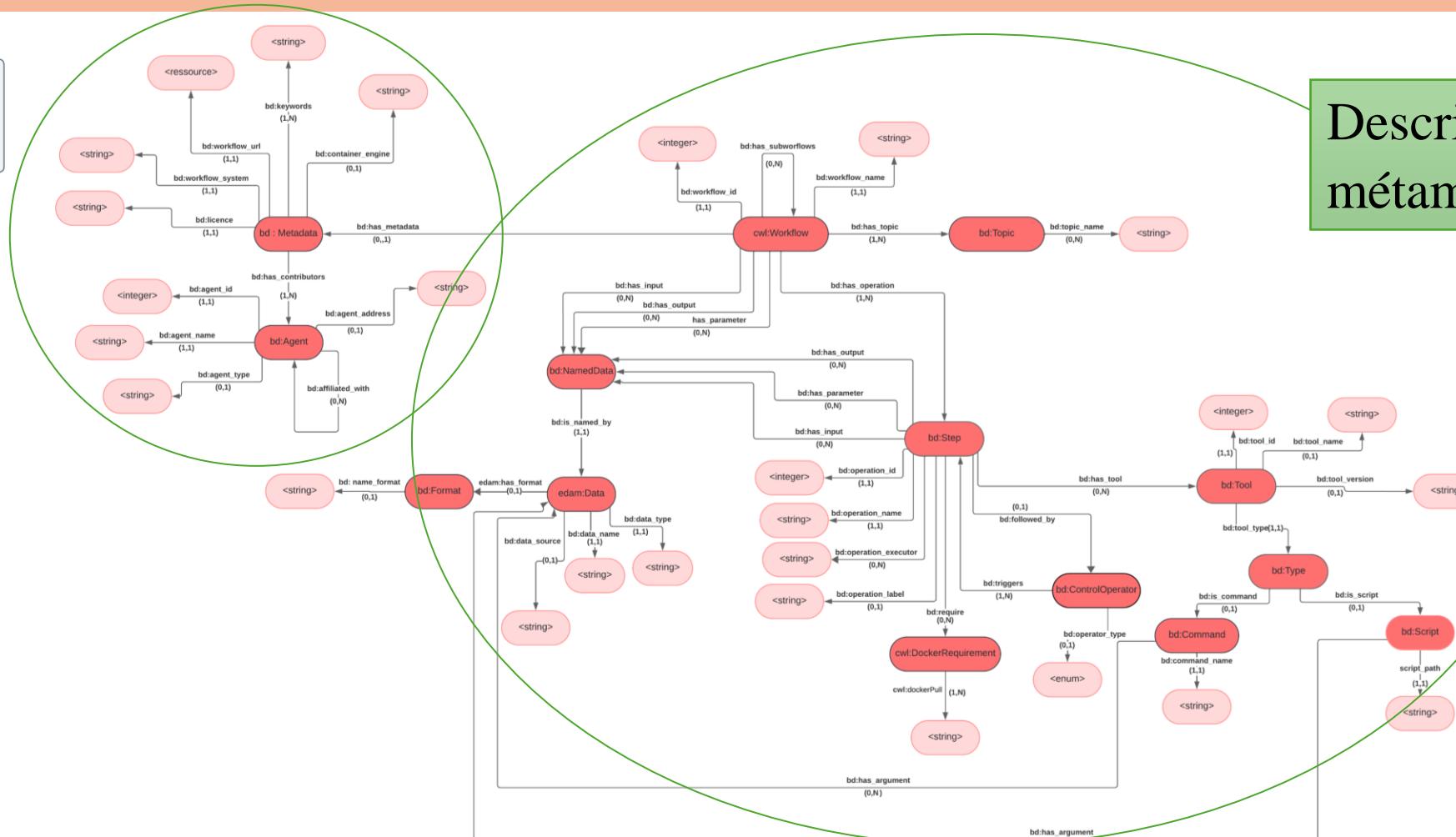
Ontologie CWL. Common Workflow Language (CWL) Workflow Description, v1.1 (commonwl.org)

Ontologie EDAM. EDAM: Ontology of bioscientific data analysis and data management (edamontology.org)

Métamodèle

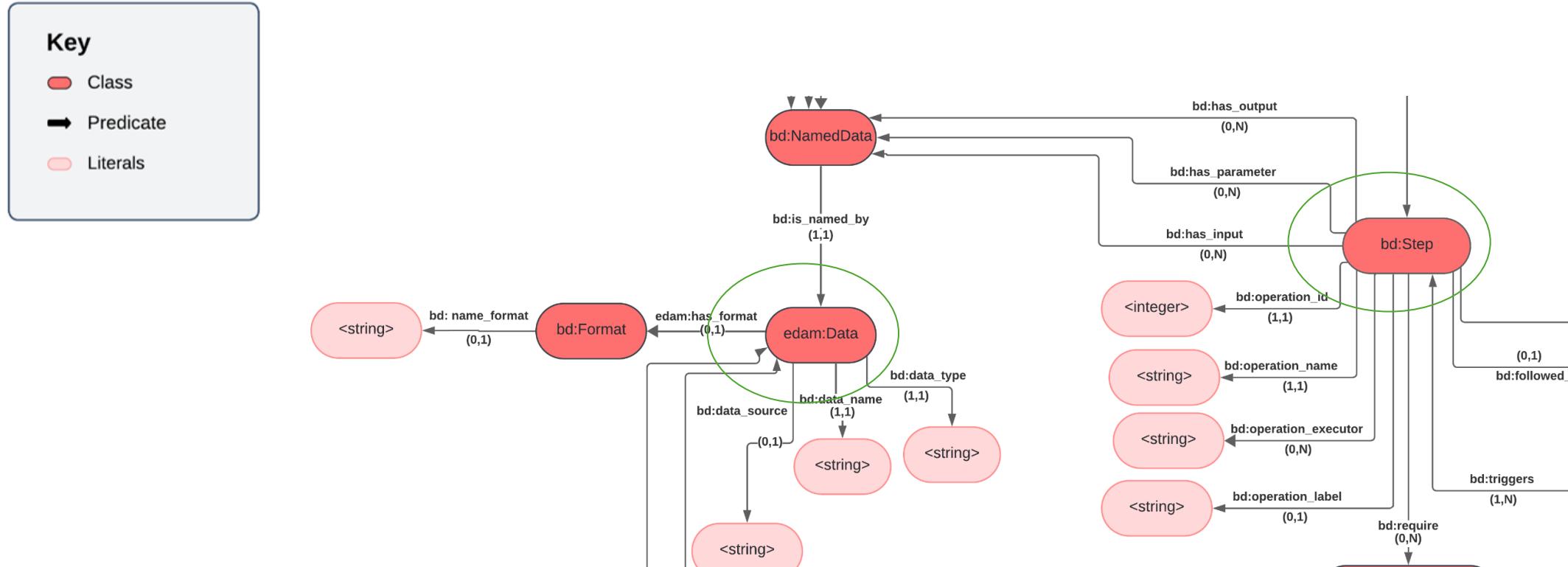
Key

- Class (Red oval)
- Predicate (Black arrow)
- Literals (Pink rounded rectangle)

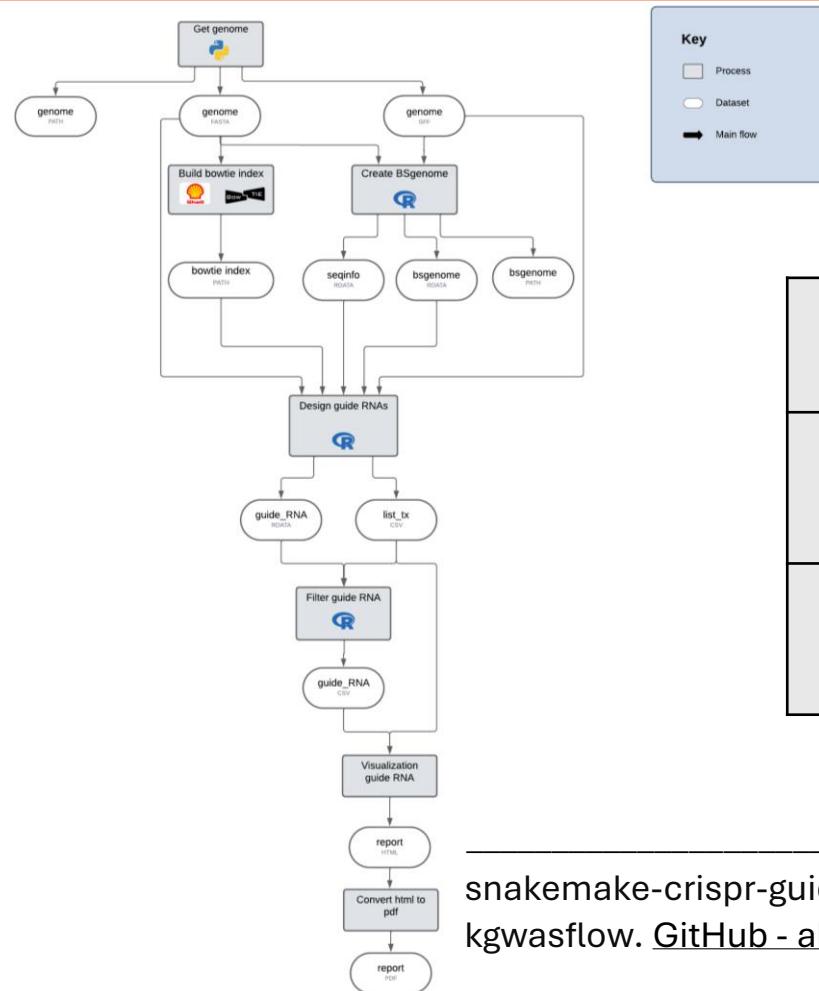


Description technique du métamodèle : 300 triplets

Métamodèle



Exemples de workflows



Workflow	# étapes	# input/output	# ligne de code
snakemake-crispr-guides	7	12	887
kgwasflow	18	44	1572



triplets métamodèle
336
(229)

snakemake-crispr-guides. <https://github.com/MPUSP/snakemake-crisprguides/tree/main#workflow-overview>.

kgwasflow. GitHub - akcorut/kGWASflow: kGWASflow is a Snakemake workflow for performing k-mers-based GWAS.

Besoins et langages utilisées

Besoins

- Lister les workflows disponibles
- Représentation du *dataflow*
- Affichage des métadonnées
- Représentation du *controlflow*



Langage	# lignes de code
HTML	119
CSS	238
JavaScript	645

HTML. HTML (HyperText Markup Language) | MDN (mozilla.org)

CSS. CSS : Feuilles de style en cascade | MDN (mozilla.org)

JavaScript. JavaScript | MDN (mozilla.org)

Liste des workflows de la base de données

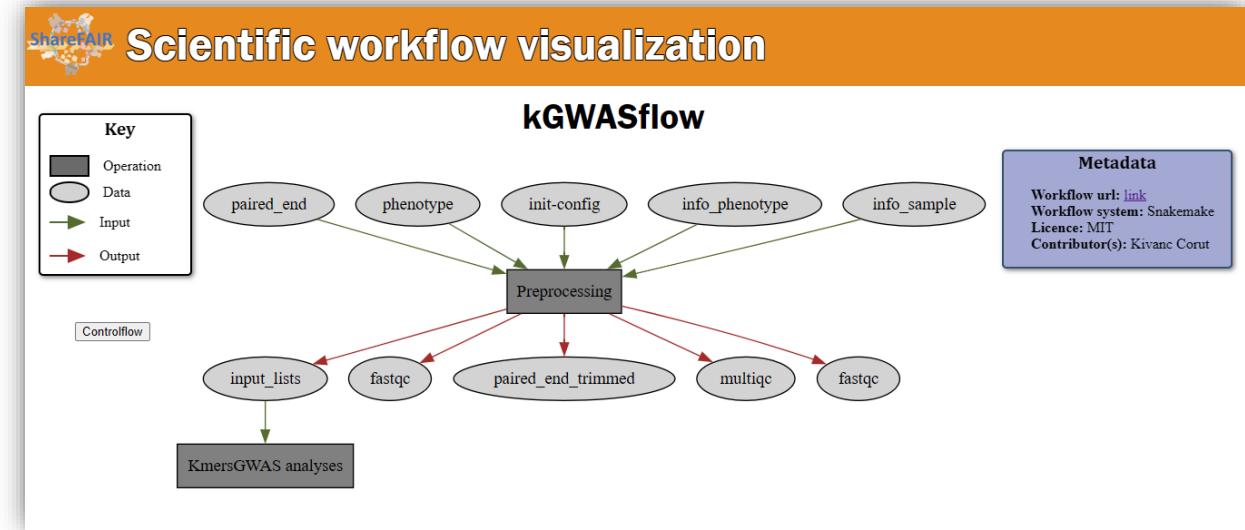
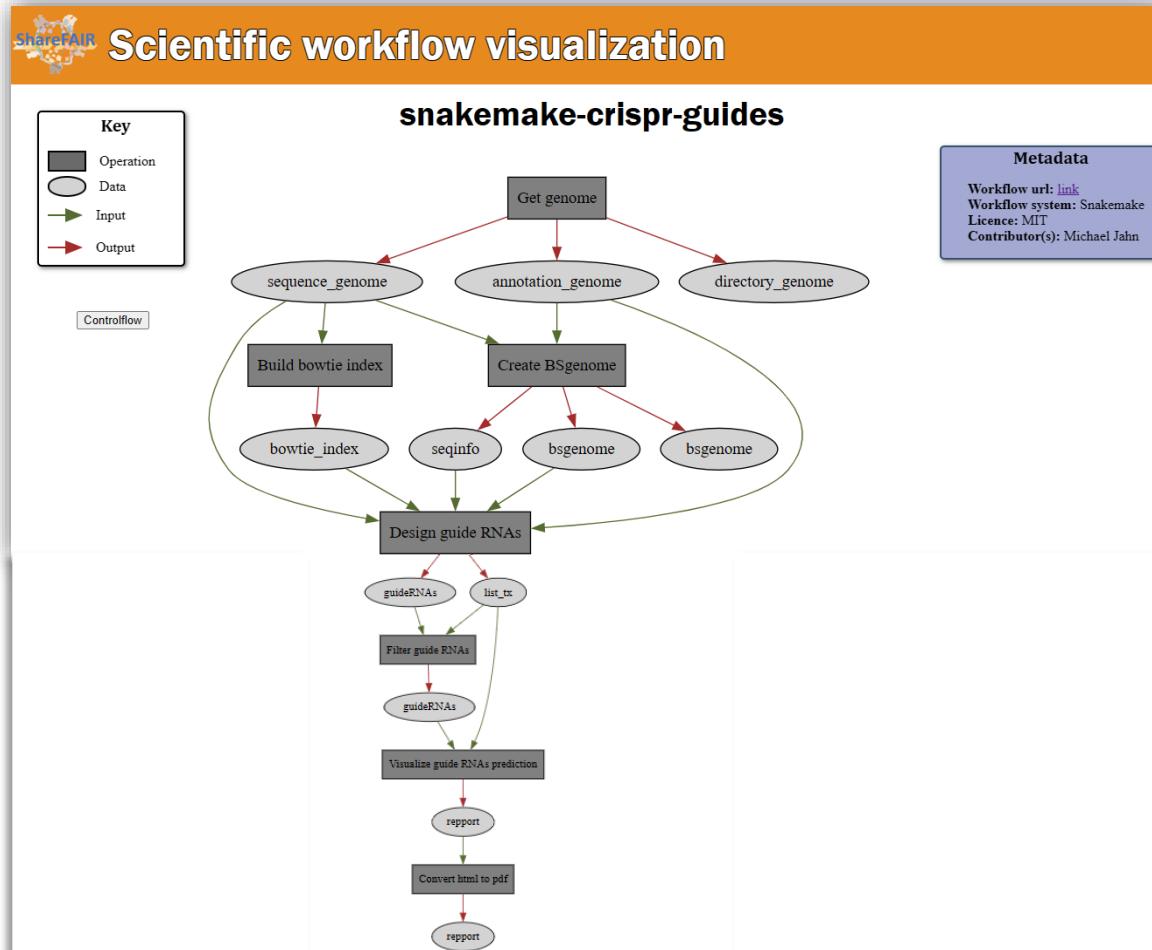


Scientific workflow visualization

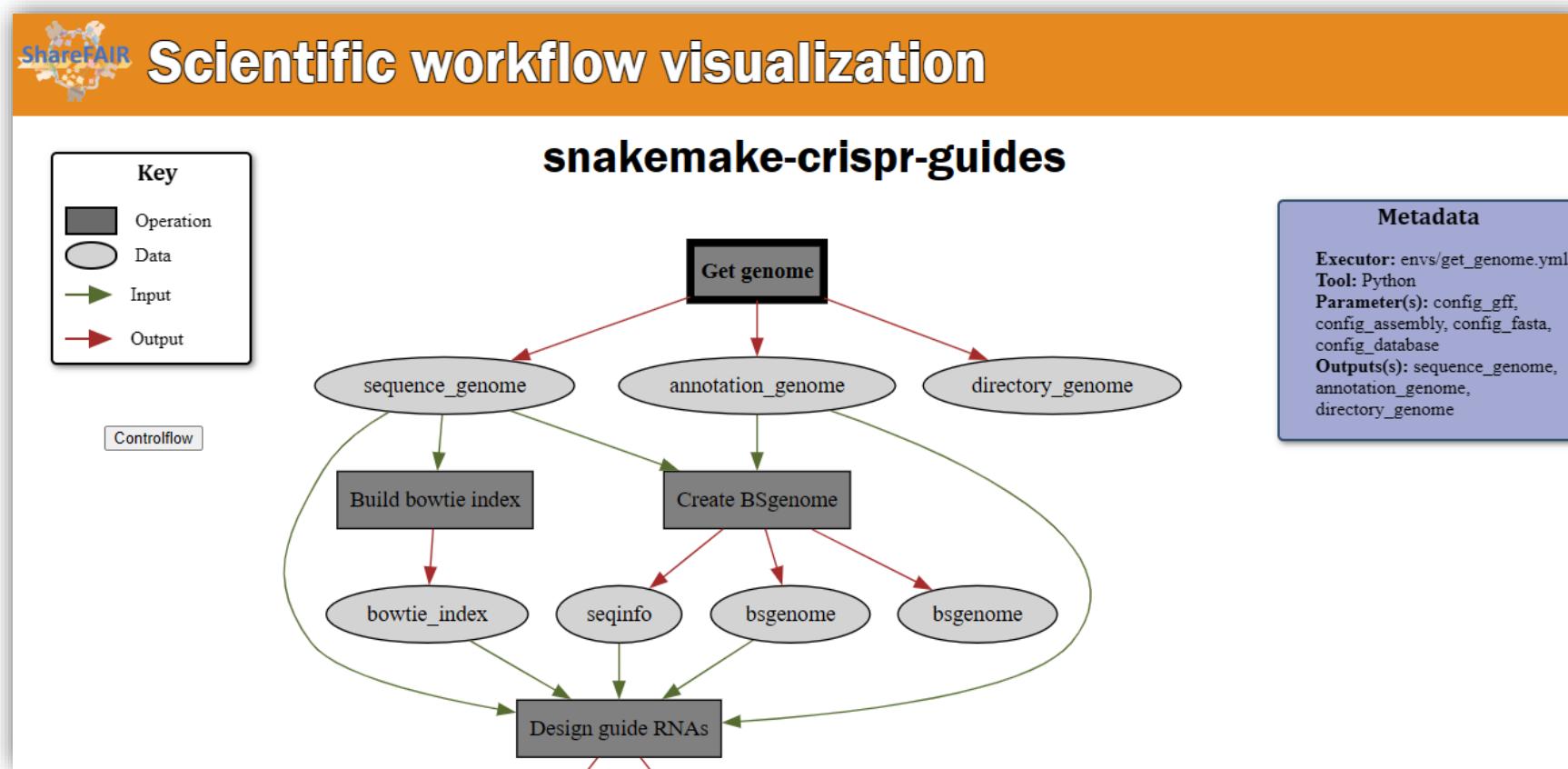
Search...

Workflow name	Keywords
kGWASflow	association-mapping, gwas-tools, gwas-pipeline, bioinformatics, structural-variation, conda, kmers, genomics, bioinformatics-tool, gwas, workflow, sciworkflows, snakemake, pipeline, bioinformatics-pipeline, ngs
snakemake-crispr-guides	crispr-design, crispr, workflow, snakemake, r-markdown, guide-rna-library, python3, bioinformatics-pipeline

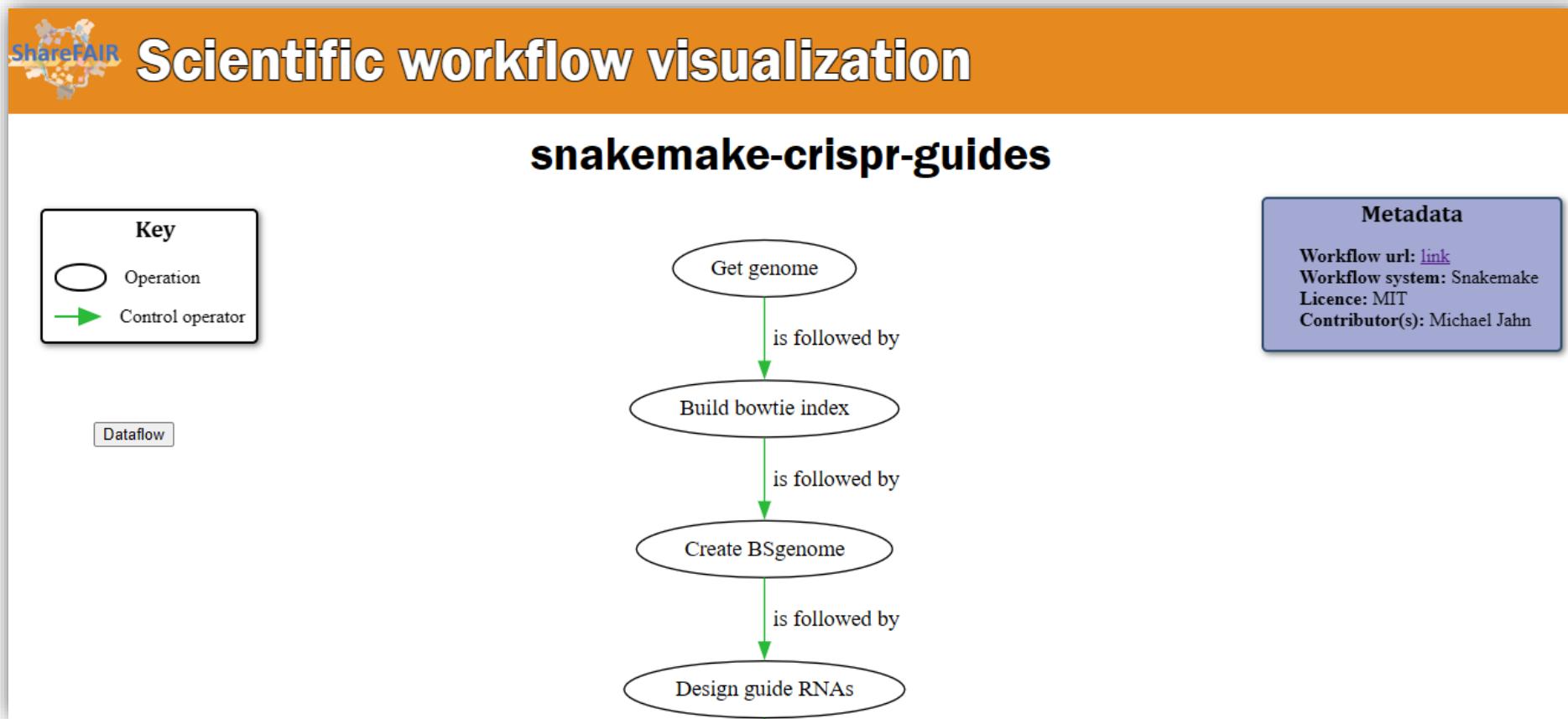
Représentation du *dataflow*



Affichage des métadonnées d'un nœud



Représentation du *controlflow*



Travaux réalisés

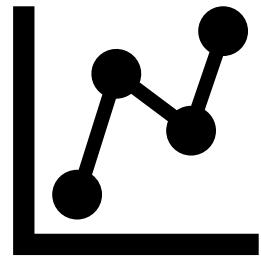


Modélisation des workflows en base de données

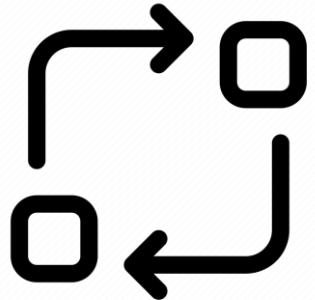


Développement d'une interface web de visualisation des workflows

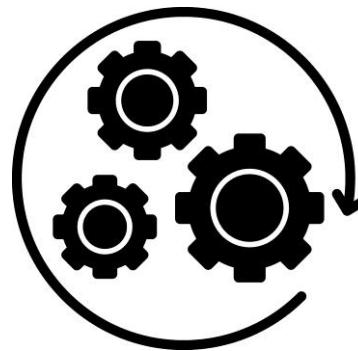
Perspectives



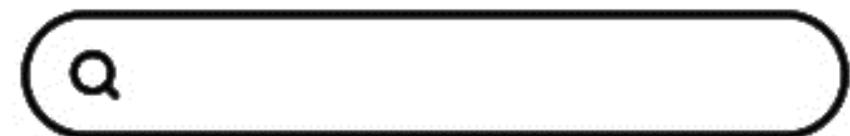
Test workflows



Controlflow

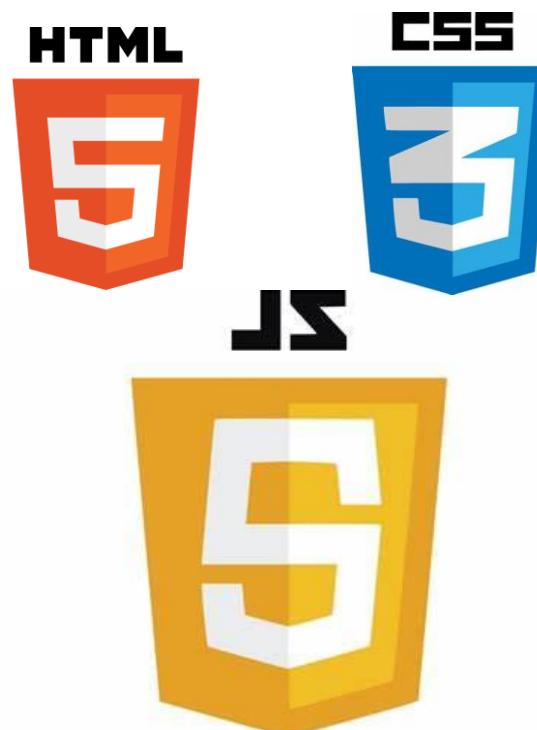


Automatisation



Langage d'interrogation

Retour d'expérience



Documenter !

Recherche de mots clés



Scientific workflow visualization

Q Rechercher

Nombre d'occurrences trouvées : 1

Workflow name	Keywords
kGWASflow	association-mapping, gwas-tools, gwas-pipeline, bioinformatics, structural-variation, conda, kmers, genomics, bioinformatics-tool, gwas, workflow, sciworkflows, snakemake, pipeline, bioinformatics-pipeline, ngs
snakemake-crispr-guides	crispr-design, crispr, workflow, snakemake, r-markdown, guide-rna-library, python3, bioinformatics-pipeline