

# EuroScienceGateway Annual Project Meeting

24th October 2024, Berlin



Funded by  
the European Union



# Report WP2

1. WorkflowHub integrations
2. Workflow Knowledge Graph
3. Reproducible FAIR Digital Objects for workflows
4. FAIR Computational Workflows
5. Workflow Publisher forum



# Report WP2

WorkflowHub integrations

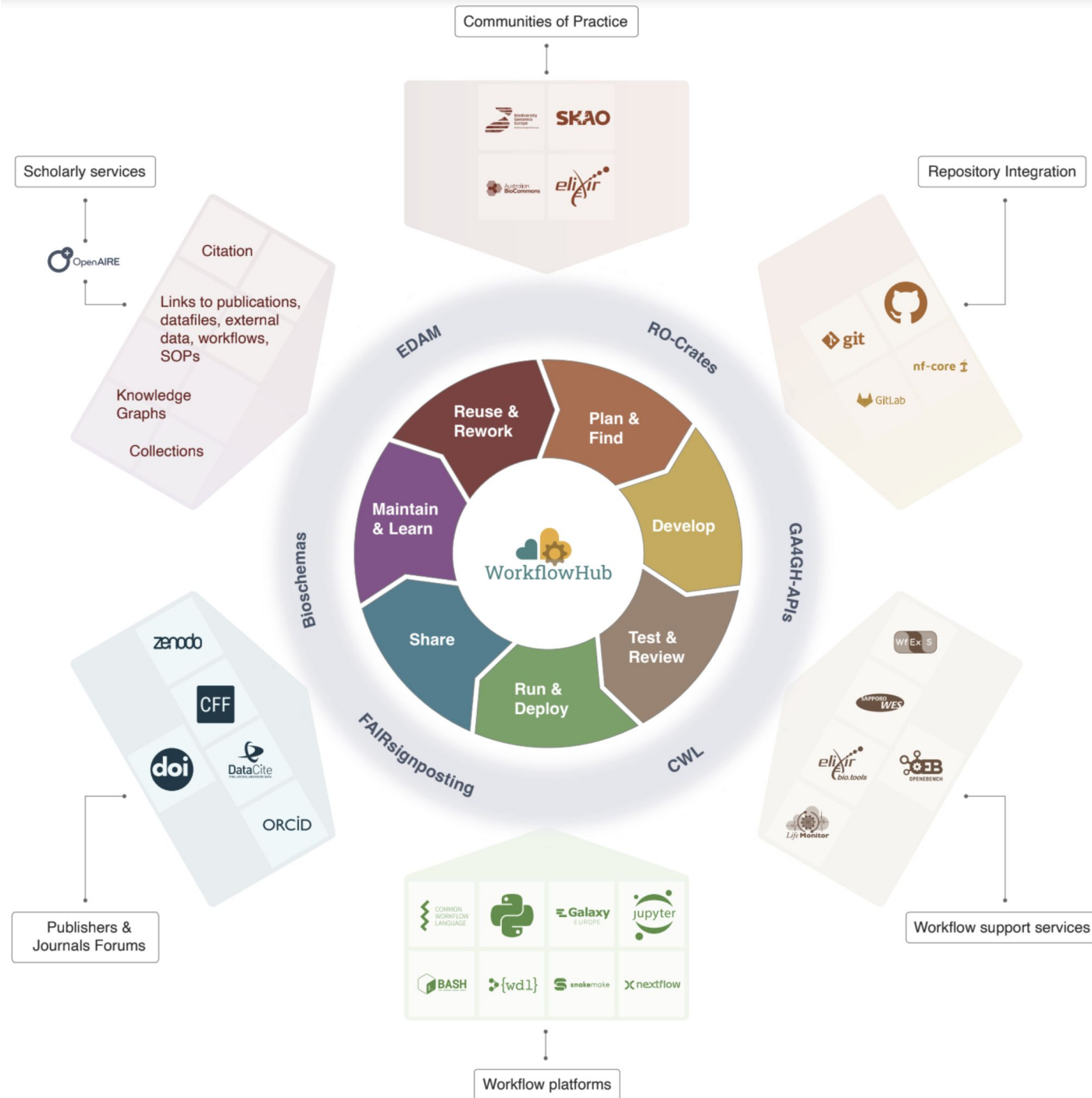
Intergalactic Workflow Commission (IWC)

Nextflow nf-core

WorkflowHub bot / GitHub

← Dockstore

Preprint: [10.48550/arXiv.2410.06941](https://arxiv.org/abs/10.48550/arXiv.2410.06941)



# Report WP2

## WorkflowHub knowledge graph

First knowledge graph from WorkflowHub

RDF from all RO-Crates

→ QA improvements, example queries

→ OpenAIRE integration EOSC-side

→ ELIXIR Research Software ecosystem

SPARQL Endpoint

Content Type (SELECT)

/workflowhub/query

JSON

```

1 PREFIX schemas: <https://schema.org/>
2 PREFIX bioschemas: <https://bioschemas.org/>
3 PREFIX owl: <http://www.w3.org/2002/07/owl#>
4 PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
5 PREFIX schema: <http://schema.org/>
6
7 SELECT DISTINCT ?wf ?name ?description
8 WHERE {
9   ?wf a bioschemas:ComputationalWorkflow
10  OPTIONAL { ?wf schema:name ?name }
11  OPTIONAL { ?wf schema:description ?description }
12 }
13 order by ?name
14

```

Table Response 1512 results in 0.292 seconds

wf	name	description
1<arcp://uuid,cbfed3b9-db0b-5635-bae6-793d376fd020/workflow/Snakefile>		
2<arcp://uuid,2e9150aa-9527-5cd5-846c-23f7dc971895/workflow_metagenomic...>	(Hybrid) Metagenomics workflow	### Workflow (hybrid) metagenomic assembl
3<arcp://uuid,11fc1582-7ae0-577b-8b59-3409711490bb/vpipe.snake>	(old) SARS-COV2 version of the V-Pipe workflow	A version of V-pipe (analysis of next generatio
4<arcp://uuid,b4f66c43-0402-5c08-a7c7-dc35f234de7a/Galaxy-Workflow-0_Vie...>	0: View complete virus identification	Non-functional workflow to get a global view
5<arcp://uuid,a9d87bcf-4c18-5ae1-b6d8-893112e88c65/16S_biodiversity_BIOM...>	16S_biodiversity_BIOM	This is a Galaxy workflow that uses to convert
6<arcp://uuid,f1a49697-92ea-5924-b170-1620646b3080/16S_biodiversity_for_n...>	16S_biodiversity_for_nonoverlap_paired_end	
7<arcp://uuid,14a24432-de25-5142-89bc-3e3056ef8fa9/16S_biodiversity_for_ov...>	16S_biodiversity_for_overlap_paired_end	MetaDEGalaxy: Galaxy workflow for differenti
8<arcp://uuid,4c3c9e85-7c7d-5333-9638-35b4d1a484bc/Galaxy-Workflow-1_Pl...>	1: Plant virus detection with kraken2 (PE)	Metagenomic dataset taxonomic classification
9<arcp://uuid,2914bff8-2bf0-56d2-a0a9-7a502e0f4384/Galaxy-Workflow-1_Pl...>	1: Plant virus detection with kraken2 (SE)	Metagenomic dataset taxonomic classification
10<arcp://uuid,a95c157c-dae2-50e6-bb6a-5ad4104056e2/Galaxy-Workflow-2_Pl...>	2: Plant virus confirmation	Mapping against all plant virus then make conl

## Report WP2

Reproducible FAIR Digital Objects for workflows

FDO profile defined in D2.1

<https://doi.org/10.5281/zenodo.13225792>

Workflow Run Crate paper in PLOS One

<https://doi.org/10.1371/journal.pone.0309210>

Assisting WRROC support in Snakemake, Nextflow,  
LEXUS, PyDolt

WfExS and Galaxy have added *rerun* functionality

next for Galaxy/WfEx: Further tool/container information

### Implementations & examples

Several workflow engines have implemented the [WRROC profiles](#) or are planning to do so:

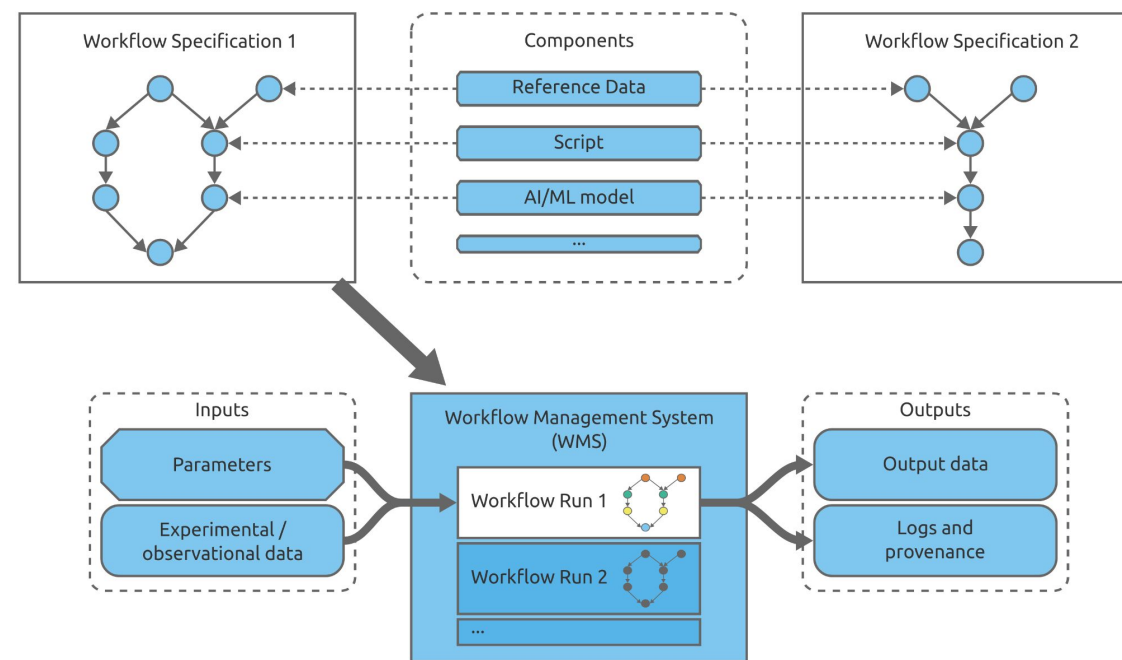
Implementation	Profile	Version URL/DOI	Example
<a href="#">runcrate</a>	Provenance	<a href="#">runcrate 0.5.0 or later</a>	<a href="#">10.5281/zenodo.7774351</a>
<a href="#">Galaxy</a>	Workflow	<a href="#">Galaxy 23.1.1 or later</a>	<a href="#">10.5281/zenodo.7785861</a>
<a href="#">COMPSs</a>	Workflow	<a href="#">compss 3.2 or later</a>	<a href="#">10.5281/zenodo.7788030</a>
<a href="#">StreamFlow</a>	Provenance	<a href="#">Streamflow 0.2.0.dev10</a>	<a href="#">10.5281/zenodo.7911906</a>
<a href="#">WfExS</a>	Workflow	<a href="#">WfExS 0.10.1 or later</a>	<a href="#">10.5281/zenodo.10091550</a>
<a href="#">Sapporo</a>	Workflow	<a href="#">sapporo-service 1.5.1 or later</a>	<a href="#">10.5281/zenodo.10134581</a>
<a href="#">Autosubmit</a>	Workflow	<a href="#">Autosubmit v4.0.100 or later</a>	<a href="#">10.5281/zenodo.8144612</a>
<a href="#">Nextflow</a>	Provenance	<a href="#">(nf-prov in development)</a>	<a href="#">example</a>

# Report WP2

## FAIR Computational Workflows

First draft of principles to be published in *Scientific Data*

Preprint: <https://doi.org/10.48550/arXiv.2410.03490>



# Report WP2

## Workflow Publisher Forum

Representatives from:

1. Elsevier
2. GigaScience
3. PLoS
4. Taylor & Francis
5. TIER2

Workshop planned for ~~2024 Q3~~ 2025 Q1

<https://galaxyproject.org/news/2024-08-03-workflow-publisher-forum/>



# Report WP2

Next period

Refresh anything EOSC:

Deliverable D2.2 Publishing workflow enriched FDOs to EOSC – including PID graph?

EOSC EU Node integrations e.g. Notebook, storage

WorkflowHub as EOSC service under ELIXIR node

Knowledge graph QA and enrichments

Tool/Container provenance – how were they made?

LLM to assist metadata curation in WorkflowHub

Reconnect with ESG work packages and Galaxy team

WorkflowHub onboarding events w/ELIXIR Tools



# EuroScienceGateway – Thank you!

