WorkflowHub

https://workflowhub.org

Carole Goble University of Manchester, UK / ELIXIR-UK
Johan Gustafsson Australian BioCommons, University of Melbourne, Australia

WCI Webinar, 16 July 2025

Where can I find workflows?

Platforms / community repositories

Galaxy PROJECT

IWC -Intergalactic Workflow Commission

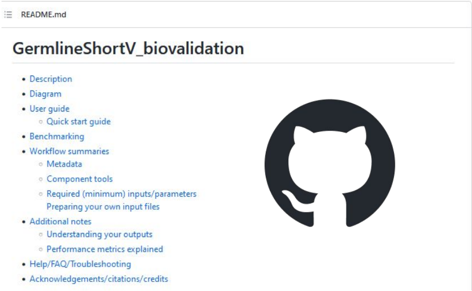
nf-core

A global community effort to collect a curated set of open-source analysis pipelines built using Nextflow.

Snakemake

Snakemake workflow catalog

GitHub



README.md

GermlineShortV_biovalidation

- Description
- Diagram
- User guide
  - Quick start guide
- Benchmarking
- Workflow summaries
  - Metadata
  - Component tools
  - Required (minimum) inputs/parameters
    Preparing your own input files
- Additional notes
  - Understanding your outputs
  - Performance metrics explained
- Help/FAQ/Troubleshooting
- Acknowledgements/citations/credits

Search engines

Publications

Murigneux, V., Roberts, L.W., Forde, B.M. *et al.* MicroPIPE: validating an end-to-end workflow for high-quality complete bacterial genome construction. *BMC Genomics* 22, 474 (2021). https://doi.org/10.1186/s12864-021-07767-z

Lott, M. J., Wright, B. R., Neaves, L. E., Frankham, G. J., Dennison, S., Eldridge, M. D. B., Potter, S., Alquezar-Planas, D. E., Hogg, C. J., Belov, K., & Johnson, R. N. (2022). Future-proofing the koala: Synergising genomic and environmental data for effective species management. Molecular Ecology, 31, 3035– 3055. https://doi.org/10.1111/mec.16446

Data Repositories

zenodo

Overcome a distributed, fragmented and variable world…

Use a registry

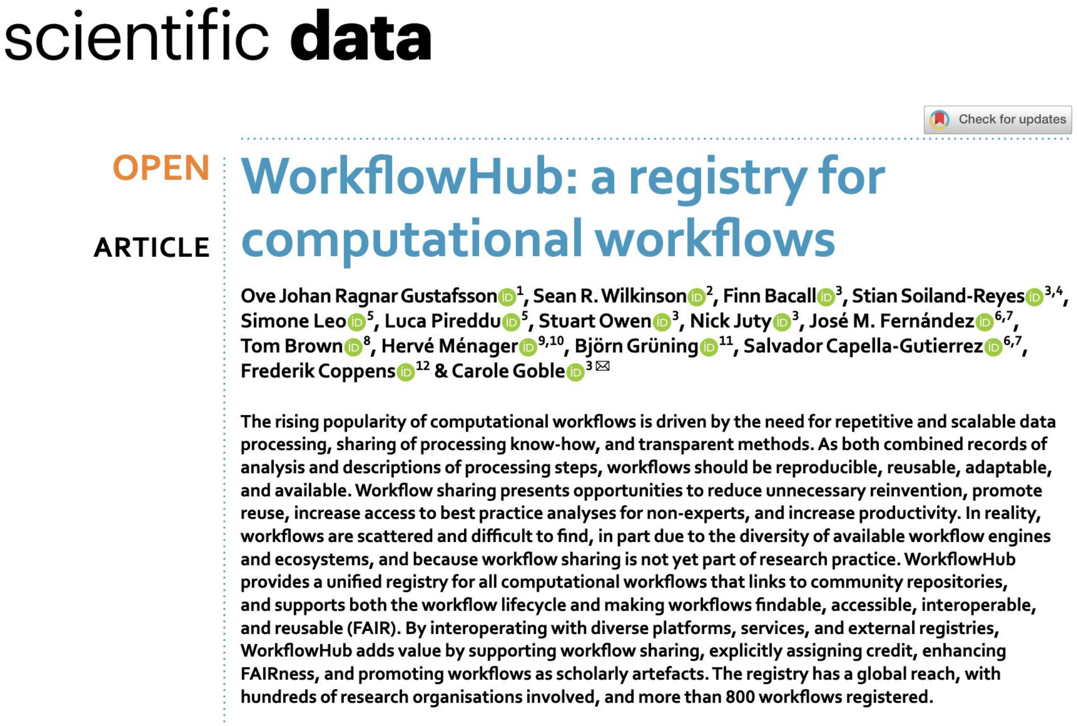
WorkflowHub

http://workflowhub.eu
http://workflowhub.org

- ✓ Central
- ✓ Integrated
- ✓ Searchable
- ✓ Standardised
- ✓ Citable
- ✓ Interoperable
- ✓ Rich metadata

WorkflowHub: a registry for computational workflows

Gustafsson, O.J.R., Wilkinson, S.R., Bacall, F. *et al.* WorkflowHub: a registry for computational workflows. *Sci Data* 12, 837 (2025). https://doi.org/10.1038/s41597-025-04786-3



scientific **data**

Check for updates

OPEN

ARTICLE

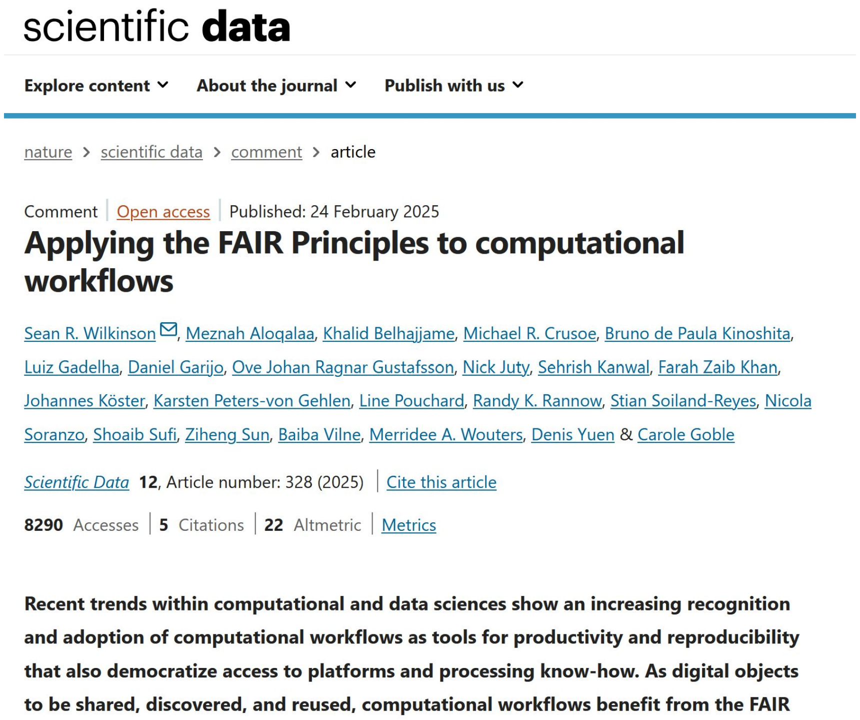
WorkflowHub: a registry for computational workflows

Ove Johan Ragnar Gustafsson[1], **Sean R. Wilkinson**[2], **Finn Bacall**[3], **Stian Soiland-Reyes**[3,4], **Simone Leo**[5], **Luca Pireddu**[5], **Stuart Owen**[3], **Nick Juty**[3], **José M. Fernández**[6,7], **Tom Brown**[8], **Hervé Ménager**[9,10], **Björn Grüning**[11], **Salvador Capella-Gutierrez**[6,7], **Frederik Coppens**[12] & **Carole Goble**[3] ✉

The rising popularity of computational workflows is driven by the need for repetitive and scalable data processing, sharing of processing know-how, and transparent methods. As both combined records of analysis and descriptions of processing steps, workflows should be reproducible, reusable, adaptable, and available. Workflow sharing presents opportunities to reduce unnecessary reinvention, promote reuse, increase access to best practice analyses for non-experts, and increase productivity. In reality, workflows are scattered and difficult to find, in part due to the diversity of available workflow engines and ecosystems, and because workflow sharing is not yet part of research practice. WorkflowHub provides a unified registry for all computational workflows that links to community repositories, and supports both the workflow lifecycle and making workflows findable, accessible, interoperable, and reusable (FAIR). By interoperating with diverse platforms, services, and external registries, WorkflowHub adds value by supporting workflow sharing, explicitly assigning credit, enhancing FAIRness, and promoting workflows as scholarly artefacts. The registry has a global reach, with hundreds of research organisations involved, and more than 800 workflows registered.

A first step for making workflows FAIR

Wilkinson, S.R., Aloqalaa, M., Belhajjame, K. et al. Applying the FAIR Principles to computational workflows.
Sci Data 12, 328 (2025). https://doi.org/10.1038/s41597-025-04451-9



scientific **data**

Explore content ⌄ **About the journal** ⌄ **Publish with us** ⌄

nature › scientific data › comment › article

Comment | Open access | Published: 24 February 2025

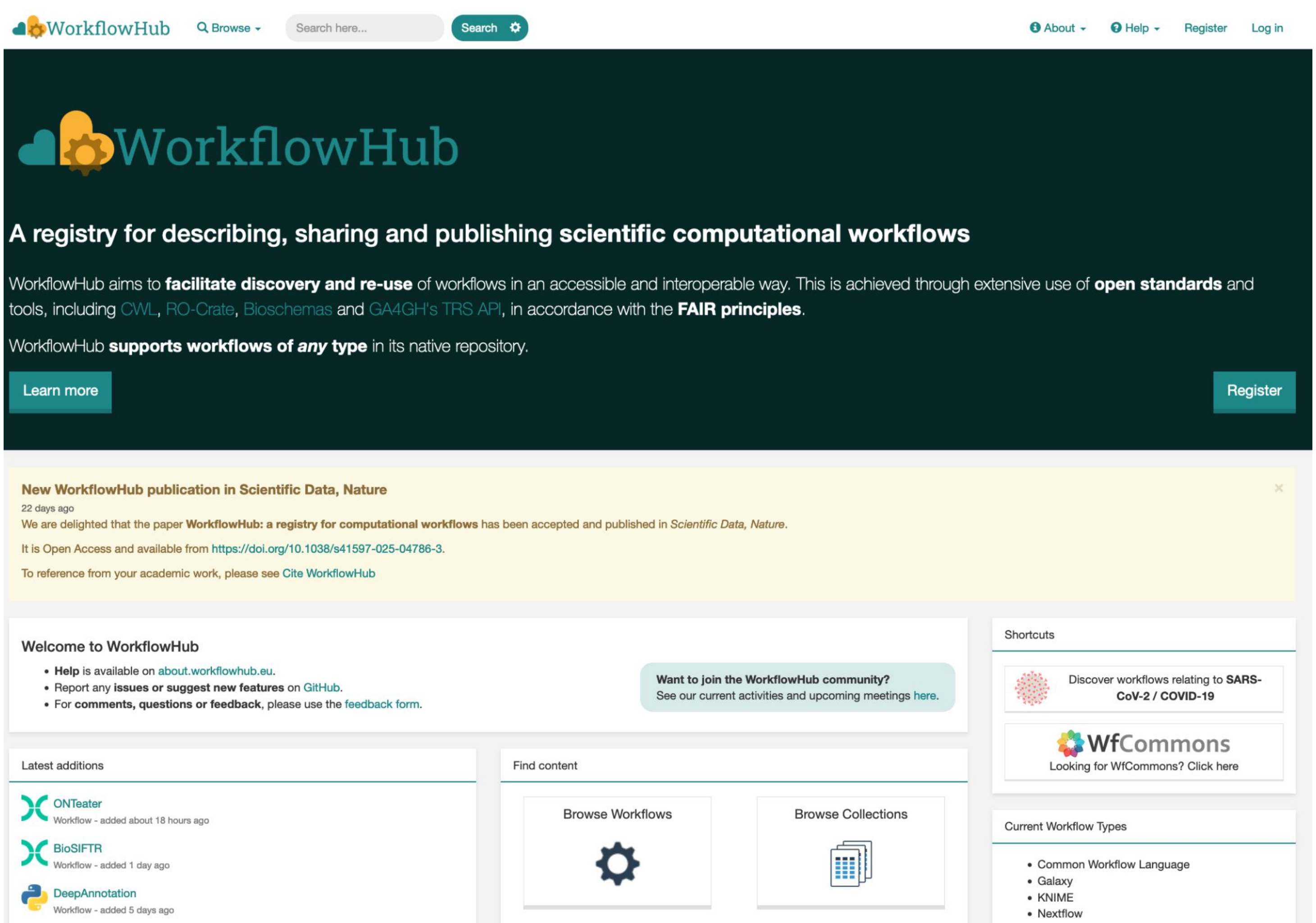
Applying the FAIR Principles to computational workflows

Sean R. Wilkinson ✉, Meznah Aloqalaa, Khalid Belhajjame, Michael R. Crusoe, Bruno de Paula Kinoshita, Luiz Gadelha, Daniel Garijo, Ove Johan Ragnar Gustafsson, Nick Juty, Sehrish Kanwal, Farah Zaib Khan, Johannes Köster, Karsten Peters-von Gehlen, Line Pouchard, Randy K. Rannow, Stian Soiland-Reyes, Nicola Soranzo, Shoaib Sufi, Ziheng Sun, Baiba Vilne, Merridee A. Wouters, Denis Yuen & Carole Goble

Scientific Data **12**, Article number: 328 (2025) | Cite this article

8290 Accesses | **5** Citations | **22** Altmetric | Metrics

Recent trends within computational and data sciences show an increasing recognition and adoption of computational workflows as tools for productivity and reproducibility that also democratize access to platforms and processing know-how. As digital objects to be shared, discovered, and reused, computational workflows benefit from the FAIR



A registry for describing, sharing and publishing scientific computational workflows

WorkflowHub aims to **facilitate discovery and re-use** of workflows in an accessible and interoperable way. This is achieved through extensive use of **open standards** and tools, including CWL, RO-Crate, Bioschemas and GA4GH's TRS API, in accordance with the **FAIR principles**.

WorkflowHub **supports workflows of *any* type** in its native repository.

New WorkflowHub publication in Scientific Data, Nature

22 days ago

We are delighted that the paper **WorkflowHub: a registry for computational workflows** has been accepted and published in *Scientific Data, Nature*.

It is Open Access and available from https://doi.org/10.1038/s41597-025-04786-3.

To reference from your academic work, please see Cite WorkflowHub

Welcome to WorkflowHub

- **Help** is available on about.workflowhub.eu.
- Report any **issues or suggest new features** on GitHub.
- For **comments, questions or feedback**, please use the feedback form.

Want to join the WorkflowHub community?
See our current activities and upcoming meetings here.

Shortcuts

Discover workflows relating to **SARS-CoV-2 / COVID-19**

WfCommons
Looking for WfCommons? Click here

Latest additions

ONTeater
Workflow - added about 18 hours ago

BioSIFTR
Workflow - added 1 day ago

DeepAnnotation
Workflow - added 5 days ago

Find content

Current Workflow Types

- Common Workflow Language
- Galaxy
- KNIME
- Nextflow

1,337
Workflows

26
System types

1,117
Users

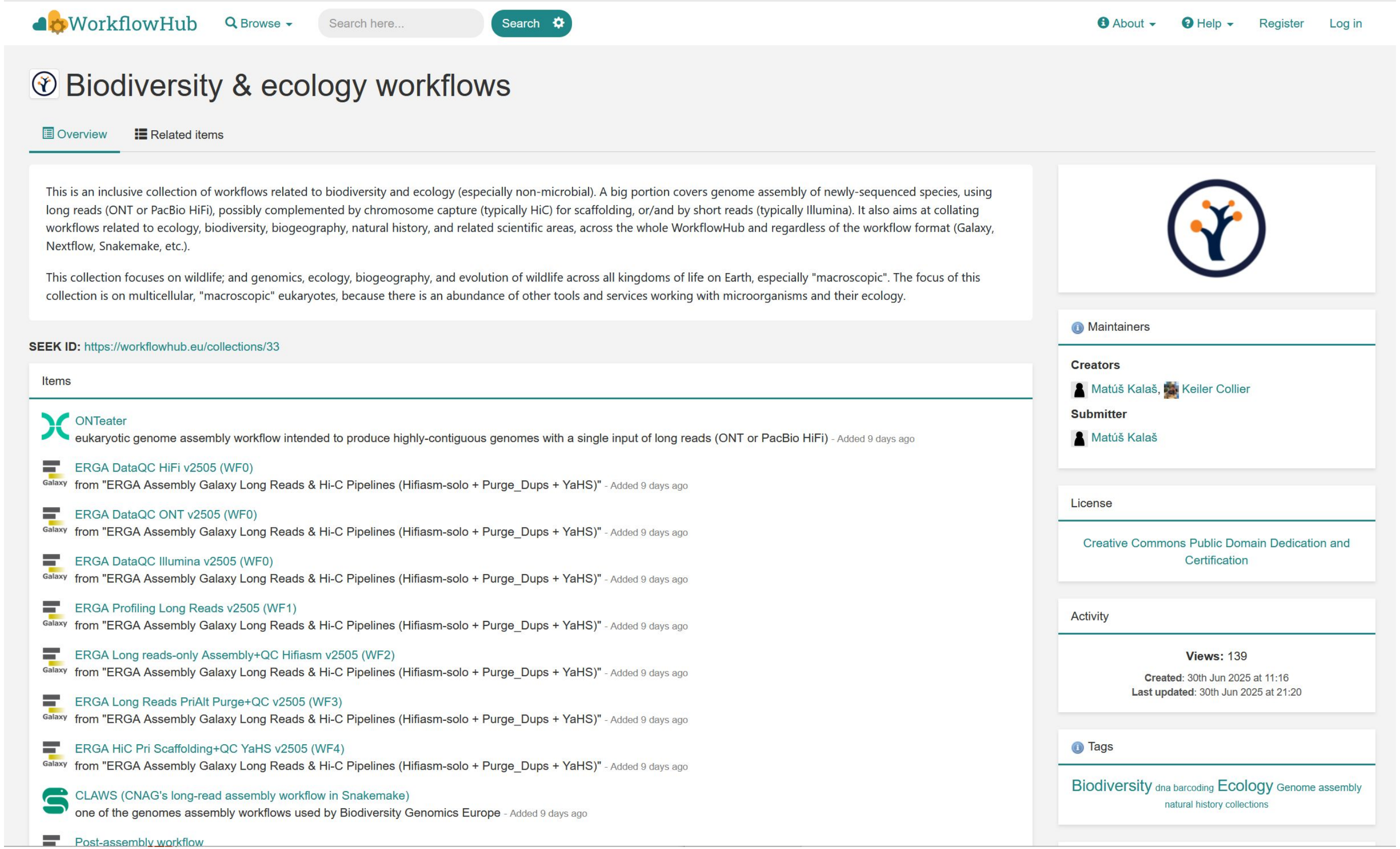
355
Teams

33
Collections

3,669,916
Views

728,597
Downloads

Any discipline, any language, any maturity



Biodiversity & ecology workflows

This is an inclusive collection of workflows related to biodiversity and ecology (especially non-microbial). A big portion covers genome assembly of newly-sequenced species, using long reads (ONT or PacBio HiFi), possibly complemented by chromosome capture (typically HiC) for scaffolding, or/and by short reads (typically Illumina). It also aims at collating workflows related to ecology, biodiversity, biogeography, natural history, and related scientific areas, across the whole WorkflowHub and regardless of the workflow format (Galaxy, Nextflow, Snakemake, etc.).

This collection focuses on wildlife; and genomics, ecology, biogeography, and evolution of wildlife across all kingdoms of life on Earth, especially "macroscopic". The focus of this collection is on multicellular, "macroscopic" eukaryotes, because there is an abundance of other tools and services working with microorganisms and their ecology.

SEEK ID: https://workflowhub.eu/collections/33

Items

ONTeater
eukaryotic genome assembly workflow intended to produce highly-contiguous genomes with a single input of long reads (ONT or PacBio HiFi) - Added 9 days ago

ERGA DataQC HiFi v2505 (WF0)
from "ERGA Assembly Galaxy Long Reads & Hi-C Pipelines (Hifiasm-solo + Purge_Dups + YaHS)" - Added 9 days ago

ERGA DataQC ONT v2505 (WF0)
from "ERGA Assembly Galaxy Long Reads & Hi-C Pipelines (Hifiasm-solo + Purge_Dups + YaHS)" - Added 9 days ago

ERGA DataQC Illumina v2505 (WF0)
from "ERGA Assembly Galaxy Long Reads & Hi-C Pipelines (Hifiasm-solo + Purge_Dups + YaHS)" - Added 9 days ago

ERGA Profiling Long Reads v2505 (WF1)
from "ERGA Assembly Galaxy Long Reads & Hi-C Pipelines (Hifiasm-solo + Purge_Dups + YaHS)" - Added 9 days ago

ERGA Long reads-only Assembly+QC Hifiasm v2505 (WF2)
from "ERGA Assembly Galaxy Long Reads & Hi-C Pipelines (Hifiasm-solo + Purge_Dups + YaHS)" - Added 9 days ago

ERGA Long Reads PriAlt Purge+QC v2505 (WF3)
from "ERGA Assembly Galaxy Long Reads & Hi-C Pipelines (Hifiasm-solo + Purge_Dups + YaHS)" - Added 9 days ago

ERGA HiC Pri Scaffolding+QC YaHS v2505 (WF4)
from "ERGA Assembly Galaxy Long Reads & Hi-C Pipelines (Hifiasm-solo + Purge_Dups + YaHS)" - Added 9 days ago

CLAWS (CNAG's long-read assembly workflow in Snakemake)
one of the genomes assembly workflows used by Biodiversity Genomics Europe - Added 9 days ago

Post-assembly workflow

Maintainers

Creators
Matúš Kalaš, Keiler Collier

Submitter
Matúš Kalaš

License

Creative Commons Public Domain Dedication and Certification

Activity

Views: 139
Created: 30th Jun 2025 at 11:16
Last updated: 30th Jun 2025 at 21:20

Tags

Biodiversity dna barcoding Ecology Genome assembly natural history collections

WorkflowHub acts as a Hub for workflows

How?

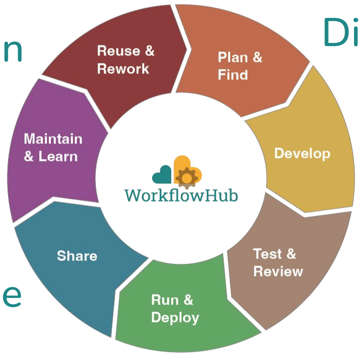
Supporting the workflow life cycle

Ease of access

Welcoming all workflows

Attribution and citation

- ✓ Author credit, citation and workflow attribution
- ✓ DOIs & ORCID, integration with DataCite



Discovery and understanding

Development and versioning

- ✓ System agnostic
- ✓ Native repository support
- ✓ Git integration
- ✓ Versions, any stage of development

Annotating workflow purpose

- ✓ Links to docs, test data, reference data
- ✓ Multiple artefacts types
- ✓ Metadata extraction & standards
- ✓ Bio.tools integration

Execution / reuse

WorkflowHub figure by Carole Goble and Xenia Perez-Sitja

Gustafsson, O.J.R., Wilkinson, S.R., Bacall, F. *et al.* WorkflowHub: a registry for computational workflows. *Sci Data* 12, 837 (2025). https://doi.org/10.1038/s41597-025-04786-3

Supporting the workflow life cycle

Using global and community standards

Standardised machine processable metadata

For reproducibility, metadata matters!

EDAM

Common metadata about the workflow, tools & parameters

https://bioschemas.org/
https://edamontology.org/
https://citation-file-format.github.io/

Bioschemas

CITATION.cff

Workflow Run RO-Crate

FAIR Digital Object format packaging workflows, metadata, companion data objects, logs

https://www.researchobject.org/ro-crate/
https://github.com/ResearchObject/workflow-run-crate

Workflow RO-Crate

COMMON WORKFLOW LANGUAGE

Common workflow description independent of platform

https://www.commonwl.org

Abstract CWL

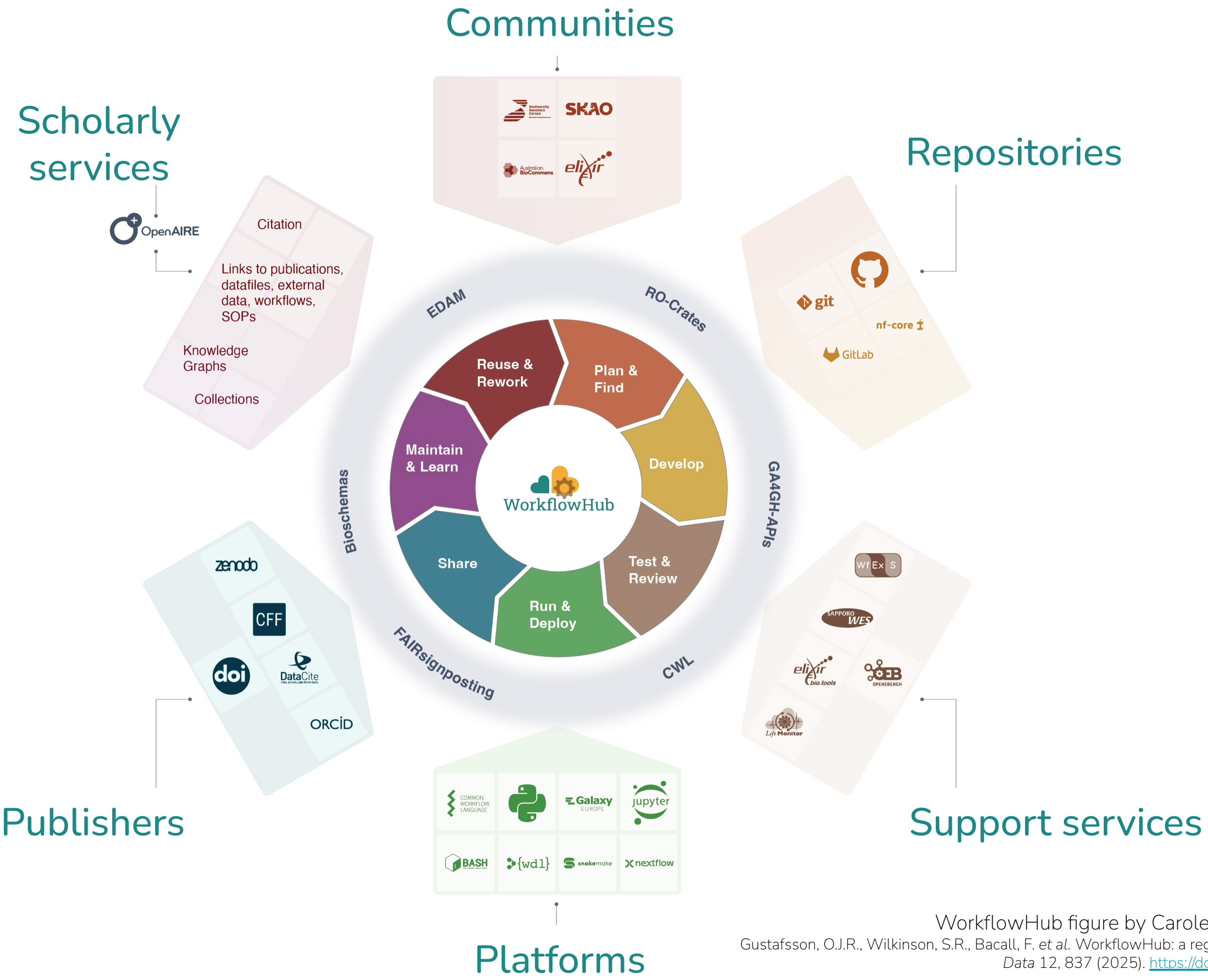
WorkflowHub figure by Carole Goble and Xenia Perez-Sitja

Gustafsson, O.J.R., Wilkinson, S.R., Bacall, F. *et al.* WorkflowHub: a registry for computational workflows. *Sci Data* 12, 837 (2025). https://doi.org/10.1038/s41597-025-04786-3

Supporting the workflow life cycle

Using global and community standards

Integrating with the global ecosystem



WorkflowHub figure by Carole Goble and Xenia Perez-Sitja

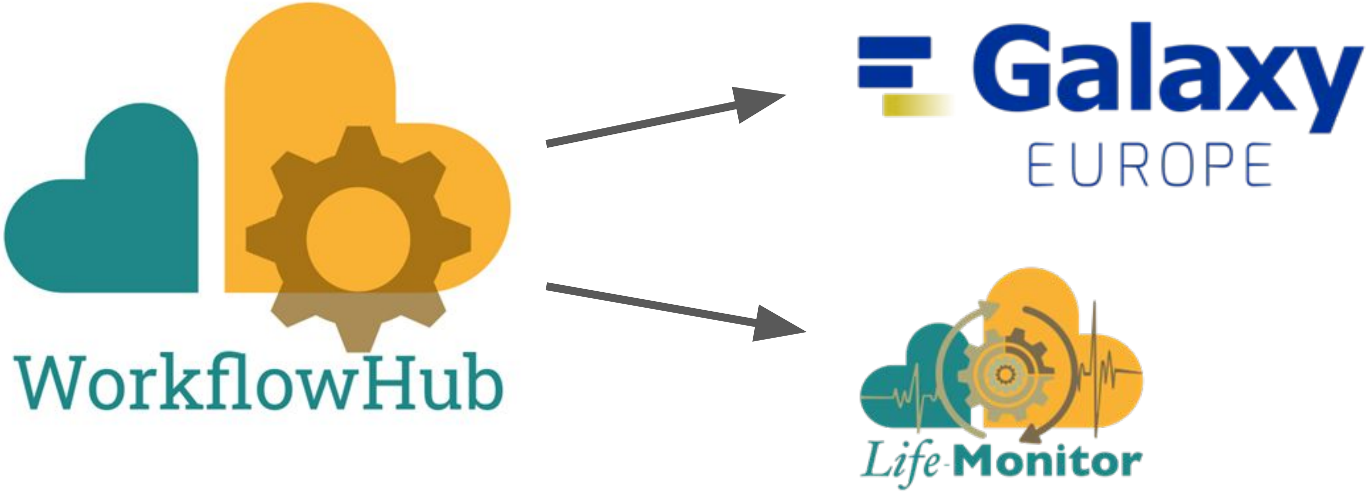
Gustafsson, O.J.R., Wilkinson, S.R., Bacall, F. *et al.* WorkflowHub: a registry for computational workflows. *Sci Data* 12, 837 (2025). https://doi.org/10.1038/s41597-025-04786-3

Supporting the workflow life cycle

Using global and community standards

Integrating with the global ecosystem

Ecosystem execution & monitoring



Global Alliance for Genomics & Health
Collaborate. Innovate. Accelerate.

TRS API

Workflow RO-Crate

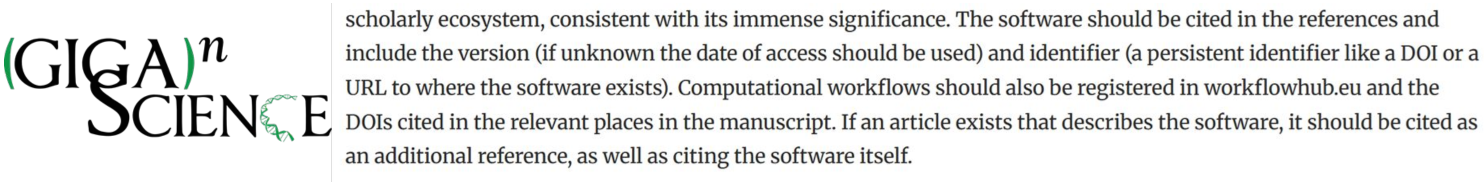
"Webby" FAIR Digital Object format

Supporting the workflow life cycle

Using global and community standards

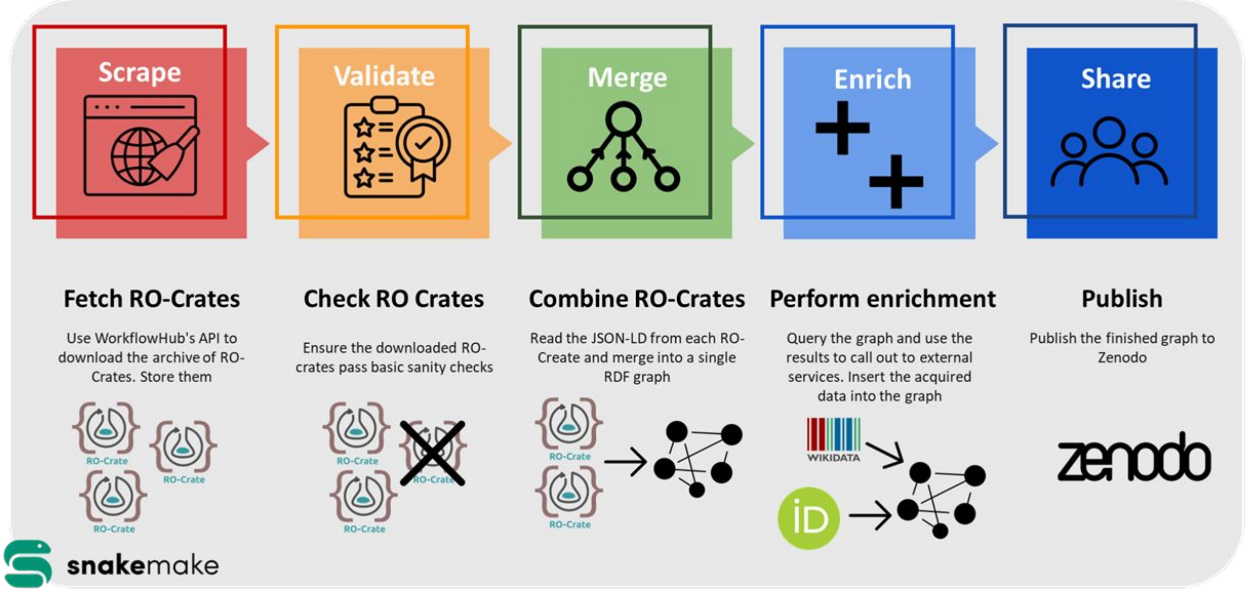
Integrating with the global ecosystem

Scholarly publication



scholarly ecosystem, consistent with its immense significance. The software should be cited in the references and include the version (if unknown the date of access should be used) and identifier (a persistent identifier like a DOI or a URL to where the software exists). Computational workflows should also be registered in workflowhub.eu and the DOIs cited in the relevant places in the manuscript. If an article exists that describes the software, it should be cited as an additional reference, as well as citing the software itself.

Knowledge Graph



Fetch RO-Crates

Use WorkflowHub's API to download the archive of RO-Crates. Store them

Check RO Crates

Ensure the downloaded RO-crates pass basic sanity checks

Combine RO-Crates

Read the JSON-LD from each RO-Create and merge into a single RDF graph

Perform enrichment

Query the graph and use the results to call out to external services. Insert the acquired data into the graph

Publish

Publish the finished graph to Zenodo

Supporting the workflow life cycle

Using global and community standards

Integrating with the global ecosystem

Faithfully representing workflows

WorkflowHub doesn't replace community and dev repositories, it works with them (+ versioning!)

GitHub link

Overview, files, and related items

Complete author list (in correct order)

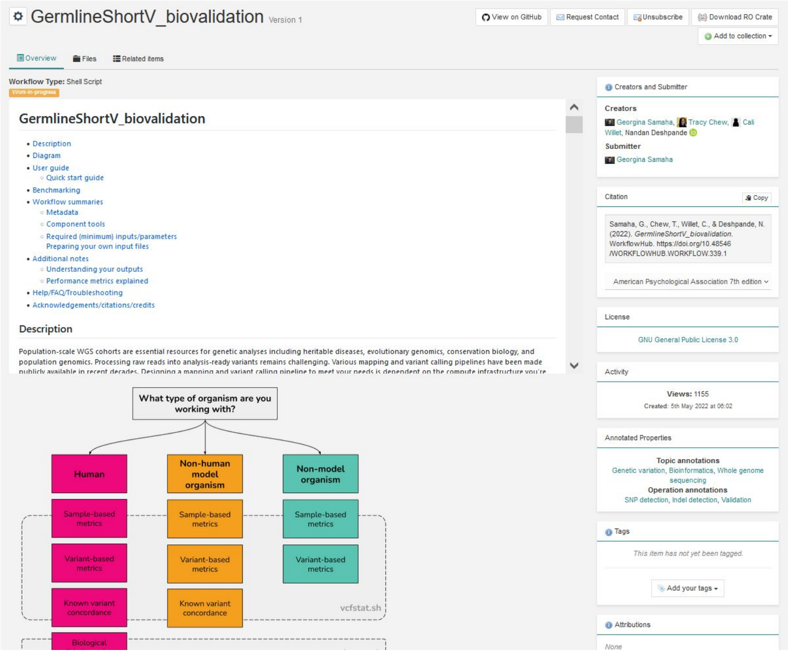
Structured documentation imported from GitHub

DOI

License

Diagram

Annotations



GermlineShortV_biovalidation Version 1

View on GitHub | Request Contact | Unsubscribe | Download RO Crate | Add to collection

Overview | Files | Related items

Workflow Type: Shell Script

Work-in-progress

GermlineShortV_biovalidation

- Description
- Diagram
- User guide
  - Quick start guide
- Benchmarking
- Workflow summaries
  - Metadata
  - Component tools
  - Required (minimum) inputs/parameters
    Preparing your own input files
- Additional notes
  - Understanding your outputs
  - Performance metrics explained
- Help/FAQ/Troubleshooting
- Acknowledgements/citations/credits

Description

Population-scale WGS cohorts are essential resources for genetic analyses including heritable diseases, evolutionary genomics, conservation biology, and population genomics. Processing raw reads into analysis-ready variants remains challenging. Various mapping and variant calling pipelines have been made publicly available in recent decades. Designing a mapping and variant calling pipeline to meet your needs is dependent on the compute infrastructure you're

Creators and Submitter

Creators

Georgina Samaha, Tracy Chew, Cali Willet, Nandan Deshpande

Submitter

Georgina Samaha

Citation | Copy

Samaha, G., Chew, T., Willet, C., & Deshpande, N. (2022). *GermlineShortV_biovalidation*. WorkflowHub. https://doi.org/10.48546/WORKFLOWHUB.WORKFLOW.339.1

American Psychological Association 7th edition

License

GNU General Public License 3.0

Activity

Views: 1155

Created: 5th May 2022 at 06:02

Annotated Properties

Topic annotations

Genetic variation, Bioinformatics, Whole genome sequencing

Operation annotations

SNP detection, Indel detection, Validation

Tags

This item has not yet been tagged.

Add your tags

Attributions

None

Samaha, G., Chew, T., Willet, C., & Deshpande, N. (2022). *GermlineShortV_biovalidation*. WorkflowHub. https://doi.org/10.48546/WORKFLOWHUB.WORKFLOW.339.1

Supporting the workflow life cycle

Using global and community standards

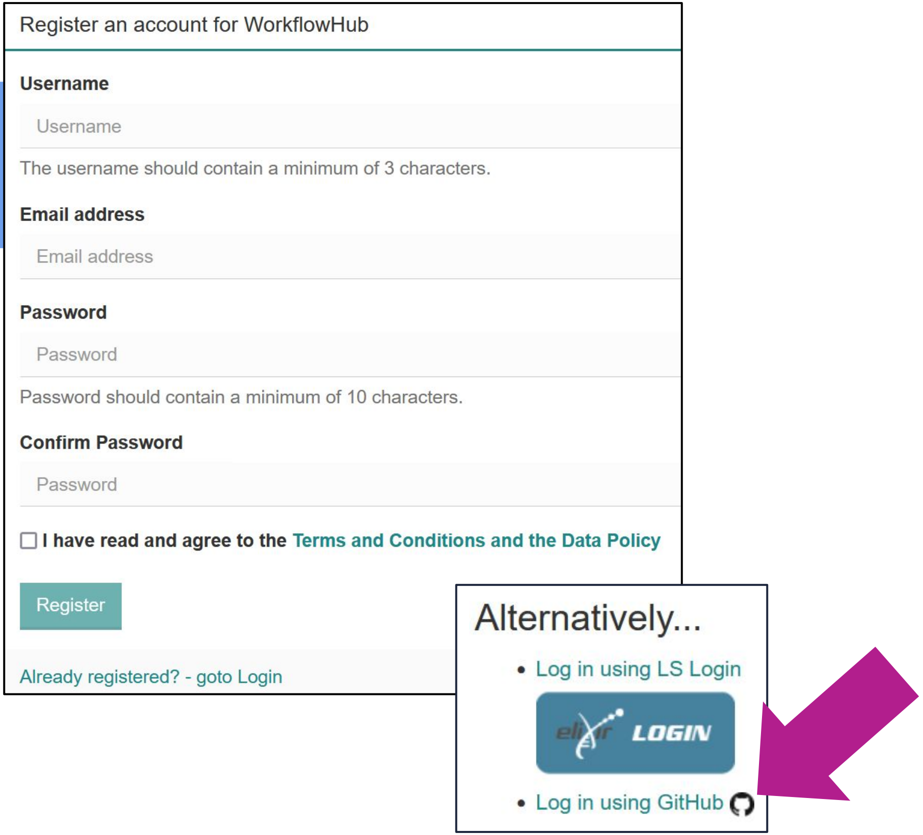
Integrating with the global ecosystem

Faithfully representing workflows

Streamlined processes

Register

1. Register yourself on WorkflowHub



Register an account for WorkflowHub

Username

Username

The username should contain a minimum of 3 characters.

Email address

Email address

Password

Password

Password should contain a minimum of 10 characters.

Confirm Password

Password

☐ **I have read and agree to the Terms and Conditions and the Data Policy**

Register

Already registered? - goto Login

Alternatively...

- Log in using LS Login
- Log in using GitHub

Create a team, or join an existing team

Supporting the workflow life cycle

Using global and community standards

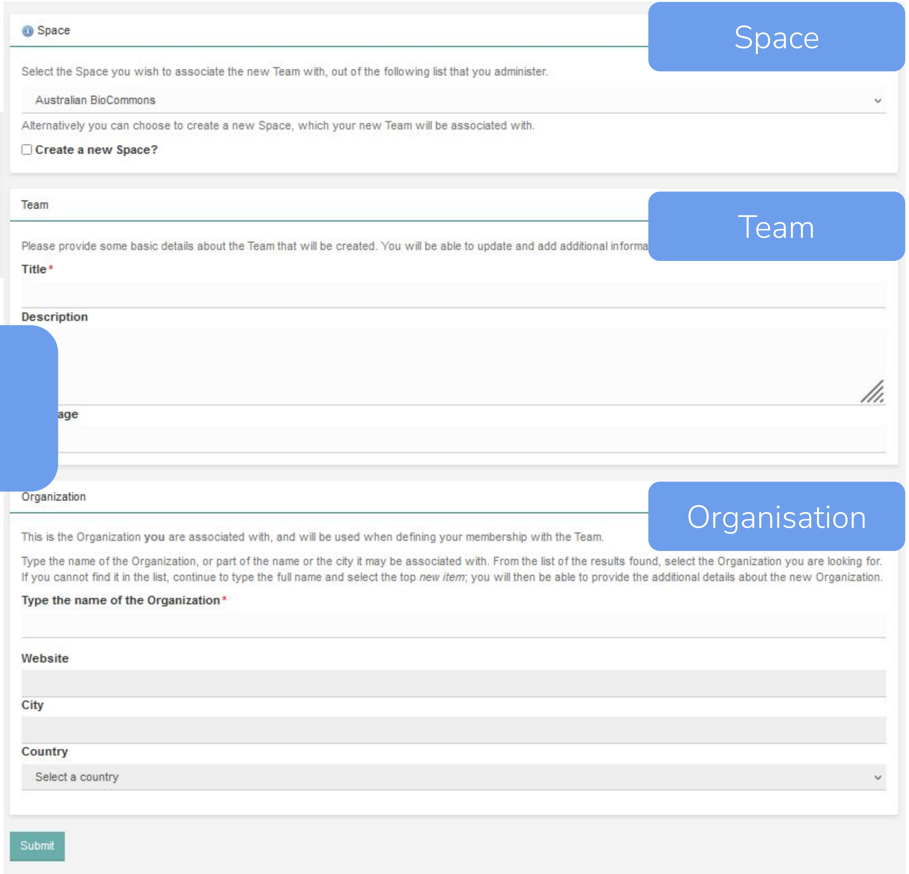
Integrating with the global ecosystem

Faithfully representing workflows

Streamlined processes

1. Register yourself on WorkflowHub

2. Create a team, or join an existing team



Space

Select the Space you wish to associate the new Team with, out of the following list that you administer.

Australian BioCommons

Alternatively you can choose to create a new Space, which your new Team will be associated with.

☐ **Create a new Space?**

Team

Please provide some basic details about the Team that will be created. You will be able to update and add additional informa

Title *

Description

age

Organization

This is the Organization you are associated with, and will be used when defining your membership with the Team.

Type the name of the Organization, or part of the name or the city it may be associated with. From the list of the results found, select the Organization you are looking for. If you cannot find it in the list, continue to type the full name and select the top *new item*; you will then be able to provide the additional details about the new Organization.

Type the name of the Organization *

Website

City

Country

Select a country

Supporting the workflow life cycle

Using global and community standards

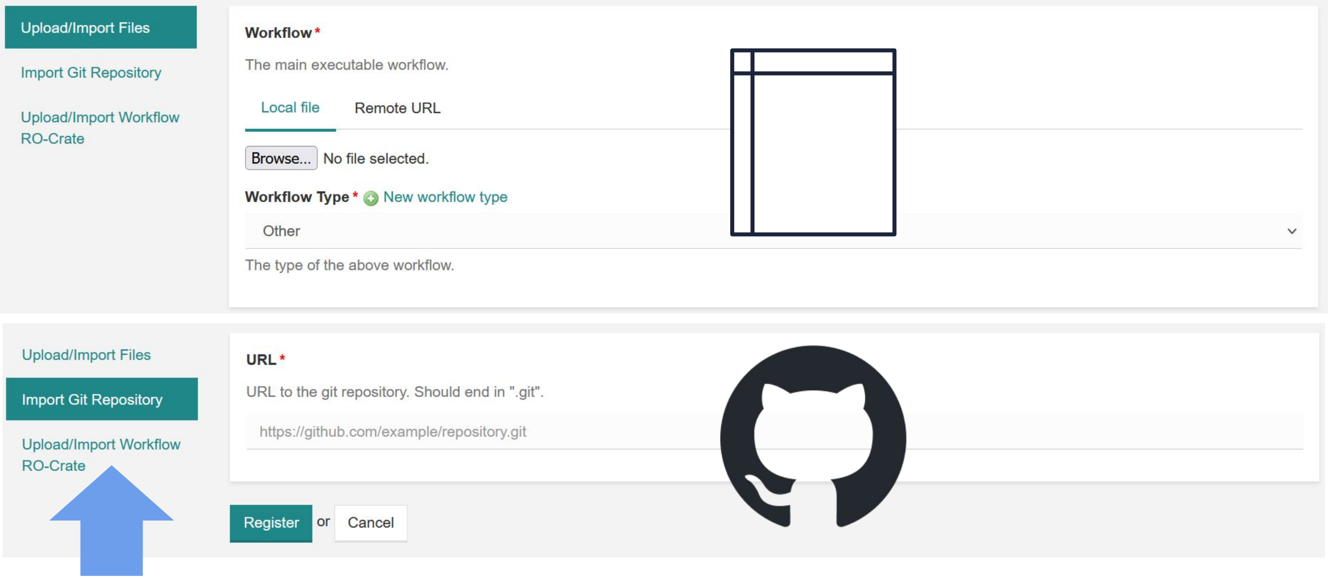
Integrating with the global ecosystem

Faithfully representing workflows

Streamlined processes

Upload workflow

3. Upload workflow



Workflow *

The main executable workflow.

No file selected.

Workflow Type *

Other

The type of the above workflow.

URL *

URL to the git repository. Should end in ".git".

https://github.com/example/repository.git

Workflow RO-Crate

https://www.researchobject.org/ro-crate/
https://w3id.org/workflowhub/workflow-ro-crate/

Supporting the workflow life cycle

Using global and community standards

Integrating with the global ecosystem

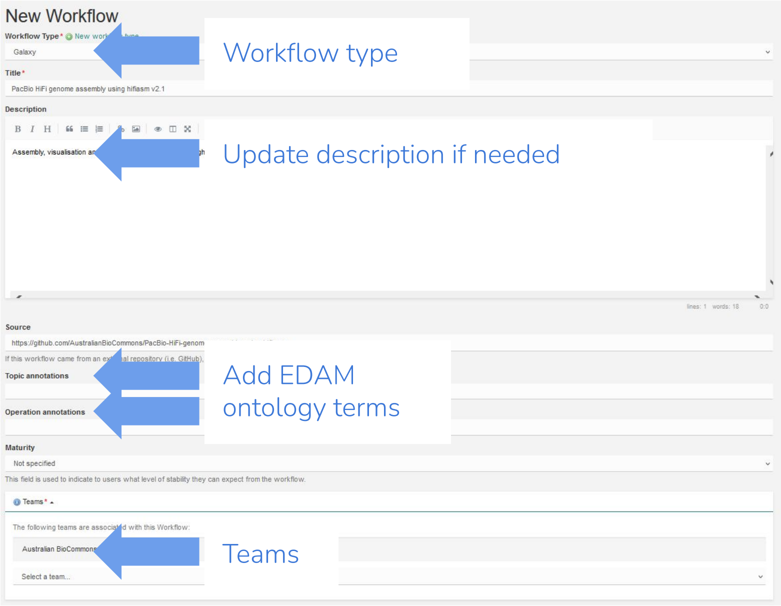
Faithfully representing workflows

Streamlined processes

Complete metadata

3. Upload workflow

4. Complete metadata



New Workflow

Workflow Type* New wor

Galaxy

Title*

PacBio HiFi genome assembly using hifiasm v2.1

Description

Assembly, visualisation an

lines: 1 words: 18 0:0

Source

https://github.com/AustralianBioCommons/PacBio-HiFi-genom

If this workflow came from an ex al repository (i.e. GitHub),

Topic annotations

Operation annotations

Maturity

Not specified

This field is used to indicate to users what level of stability they can expect from the workflow.

Teams* ▴

The following teams are associat d with this Workflow:

Australian BioCommons

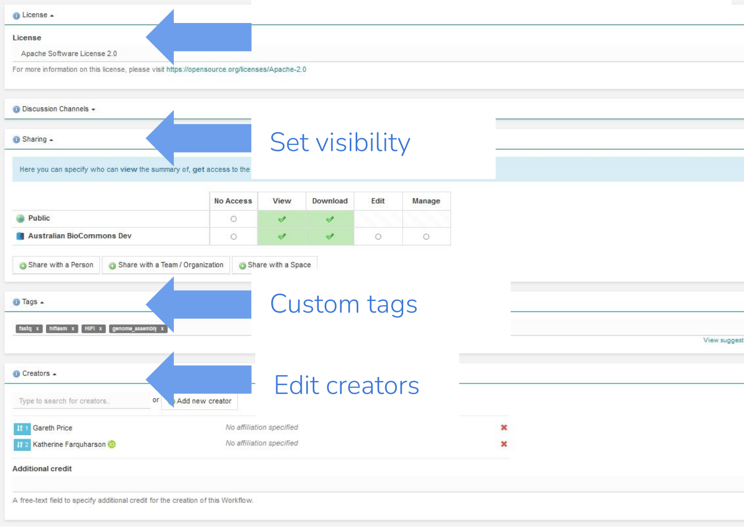
Select a team...

Workflow type

Update description if needed

Add EDAM ontology terms

Teams



Add license (auto if license in GitHub)

License ▴

License

Apache Software License 2.0

For more information on this license, please visit https://opensource.org/licenses/Apache-2.0

Discussion Channels ▾

Sharing ▴

Here you can specify who can **view** the summary of, **get** access to the

	No Access	View	Download	Edit	Manage
Public	○	✓	✓		
Australian BioCommons Dev	○	✓	✓	○	○

Share with a Person

Share with a Team / Organization

Share with a Space

Tags ▴

fastq x hifiasm x HiFi x genome_assembly x

View suggest

Creators ▴

Type to search for creators.. or Add new creator

1 Gareth Price — *No affiliation specified*

2 Katherine Farquharson — *No affiliation specified*

Additional credit

A free-text field to specify additional credit for the creation of this Workflow.

Set visibility

Custom tags

Edit creators

Supporting the workflow life cycle

Using global and community standards

Integrating with the global ecosystem

Faithfully representing workflows

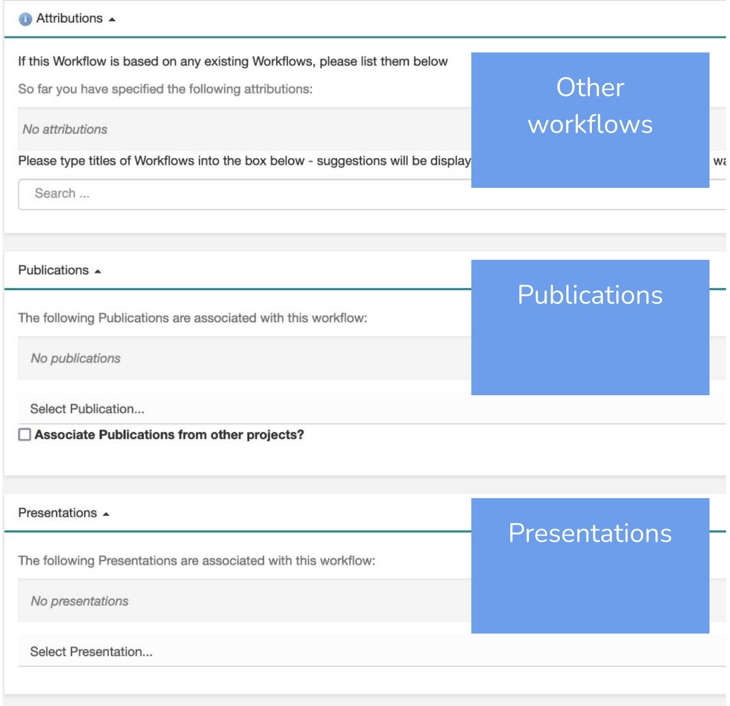
Streamlined processes

Add context: SOPs, data, tools, workflows

3. Upload workflow

4. Complete metadata

5. Add context



Attributions ▴

If this Workflow is based on any existing Workflows, please list them below

So far you have specified the following attributions:

No attributions

Please type titles of Workflows into the box below - suggestions will be display wa

Search ...

Other workflows

Publications ▴

The following Publications are associated with this workflow:

No publications

Select Publication...

☐ **Associate Publications from other projects?**

Publications

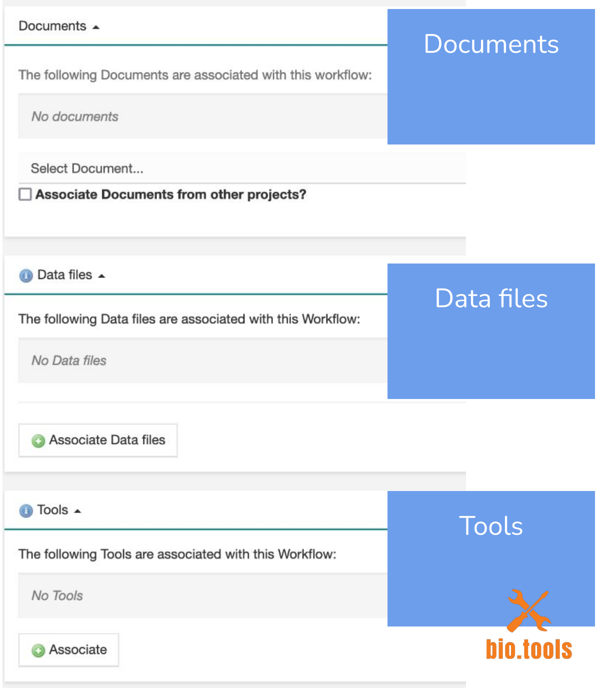
Presentations ▴

The following Presentations are associated with this workflow:

No presentations

Select Presentation...

Presentations



Documents ▴

The following Documents are associated with this workflow:

No documents

Select Document...

☐ **Associate Documents from other projects?**

Documents

Data files ▴

The following Data files are associated with this Workflow:

No Data files

Associate Data files

Data files

Tools ▴

The following Tools are associated with this Workflow:

No Tools

Associate

Tools

bio.tools

Supporting the workflow life cycle

Using global and community standards

Integrating with the global ecosystem

Faithfully representing workflows

Streamlined processes

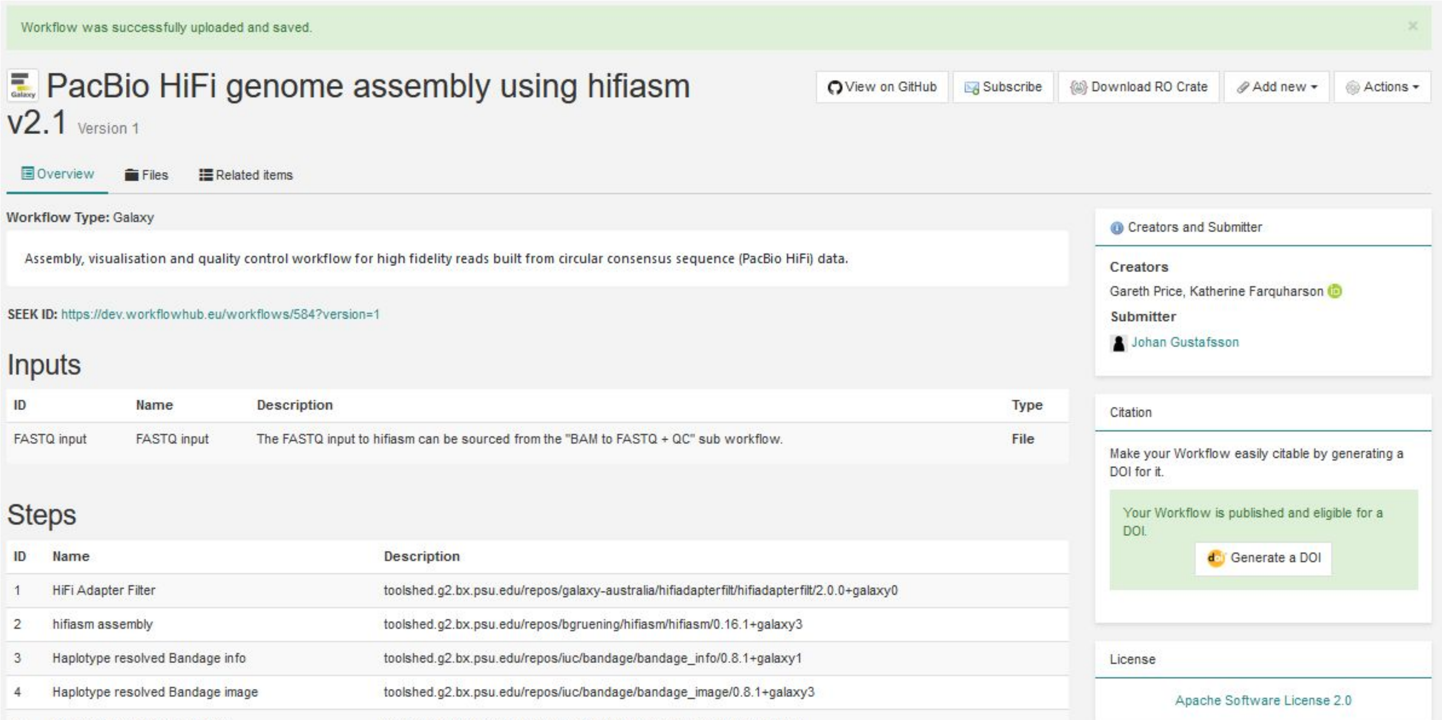
Review entry

3. Upload workflow

4. Complete metadata

5. Add context

6. Review entry



Workflow was successfully uploaded and saved.

PacBio HiFi genome assembly using hifiasm v2.1 Version 1

Workflow Type: Galaxy

Assembly, visualisation and quality control workflow for high fidelity reads built from circular consensus sequence (PacBio HiFi) data.

SEEK ID: https://dev.workflowhub.eu/workflows/584?version=1

Inputs

ID	Name	Description	Type
FASTQ input	FASTQ input	The FASTQ input to hifiasm can be sourced from the "BAM to FASTQ + QC" sub workflow.	**File**

Steps

ID	Name	Description
1	HiFi Adapter Filter	toolshed.g2.bx.psu.edu/repos/galaxy-australia/hifiadapterfilt/hifiadapterfilt/2.0.0+galaxy0
2	hifiasm assembly	toolshed.g2.bx.psu.edu/repos/bgruening/hifiasm/hifiasm/0.16.1+galaxy3
3	Haplotype resolved Bandage info	toolshed.g2.bx.psu.edu/repos/iuc/bandage/bandage_info/0.8.1+galaxy1
4	Haplotype resolved Bandage image	toolshed.g2.bx.psu.edu/repos/iuc/bandage/bandage_image/0.8.1+galaxy3

Creators and Submitter

Creators

Gareth Price, Katherine Farquharson

Submitter

Johan Gustafsson

Citation

Make your Workflow easily citable by generating a DOI for it.

Your Workflow is published and eligible for a DOI.

License

Apache Software License 2.0

Supporting the workflow life cycle

Using global and community standards

Integrating with the global ecosystem

Faithfully representing workflows

Streamlined processes

Fit-for-purpose data model

A data model that reflects the real-world collaborations that create workflows

Australian BioCommons

Overview | Related items

The Australian BioCommons enhances digital life science research through world class collaborative distributed infrastructure. It aims to ensure that Australian life science research remains globally competitive, through sustained strategic leadership, research community engagement, digital service provision, training and support.

Web page: https://www.biocommons.org.au/

Funding details:

Core funding for the Australian BioCommons comes from the National Collaborative Research Infrastructure Strategy (NCRIS) via Bioplatforms Australia, which is subcontracted to The University of Melbourne as the lead agent. This core funding is amplified through coinvestment from BioCommons partners https://www.biocommons.org.au/funding

QCIF Bioinformatics

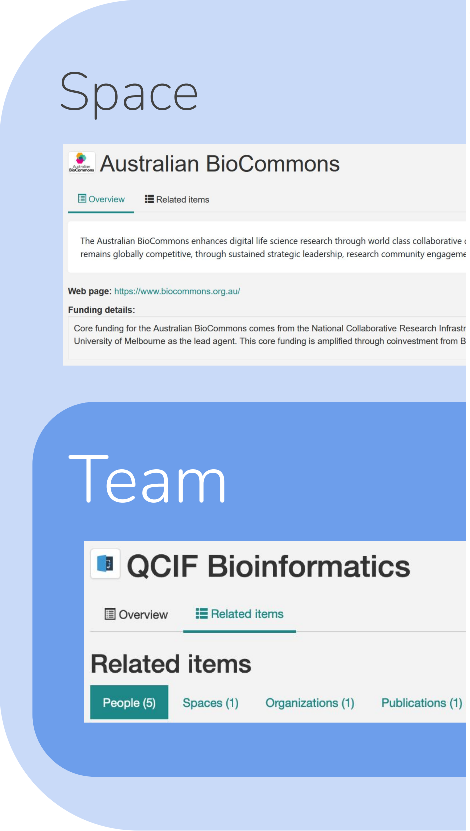
Overview | Related items

Related items

People (5) | Spaces (1) | Organizations (1) | Publications (1) | Documents (1) | Workflows (20) | Collections (2)

A data model that reflects the real-world collaborations that create workflows

- Supporting the workflow life cycle
- Using global and community standards
- Integrating with the global ecosystem
- Faithfully representing workflows
- Streamlined processes
- **Fit-for-purpose data model**


Space

Australian BioCommons

The Australian BioCommons enhances digital life science research through world class collaborative
remains globally competitive, through sustained strategic leadership, research community engageme

Web page: https://www.biocommons.org.au/

Funding details:

Core funding for the Australian BioCommons comes from the National Collaborative Research Infrastr
University of Melbourne as the lead agent. This core funding is amplified through coinvestment from B

Team

QCIF Bioinformatics

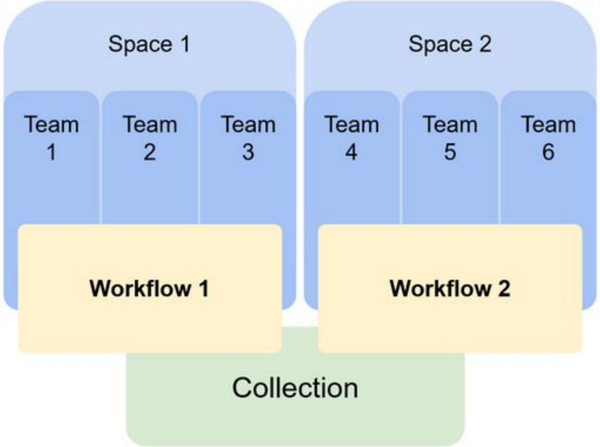
Related items


- Australian BioCommons
  - Australian BioCommons
  - QCIF Bioinformatics
    - MicroPIPE: validating an end-to-end workflow for high-quality complete bacterial genome construction.
    - 16S Microbial Analysis with mother - howto
    - microPIPE: a pipeline for high-quality bacterial genome construction using ONT and Illumina sequencing
    - 16S_biodiversity_for_overlap_paired_end
    - 16S_biodiversity_for_nonoverlap_paired_end
    - scRNAseq Single Sample Processing STARSolo
    - scRNAseq_CellQC
    - scRNAseq_QCtoBasicProcessing
    - scRNAseq: Load counts matrix
    - scRNAseq: Count and Load with starSOLO
    - scRNAseq Single Sample Processing Counts Matrix
    - Analyses of shotgun metagenomics data with MetaPhlAn2
    - scRNAseq: Count and Load with Cell Ranger
    - scRNAseq Single Sample Processing Cell Ranger
    - Workflow 1: Further Quality Control [16S Microbial Analysis With Mothur]
    - Workflow 2: Data Cleaning And Chimera Removal [16S Microbial Analysis With Mothur]
    - Workflow 3: Classification [Galaxy Training: 16S Microbial Analysis With Mothur]
    - Workflow 5: OTU Clustering [16S Microbial Analysis With Mothur]
    - Workflow 6: Alpha Diversity [16S Microbial Analysis With Mothur]
    - Workflow 7 : Beta Diversity [16S Microbial Analysis With Mothur]
    - Taxonomy classification using Kraken2 and Bracken
    - ONTViSc (ONT-based Viral Screening for Biosecurity)
    - scRNAseq processing in galaxy
    - 16S Microbial Analysis with mothur (on Galaxy Australia)

- Supporting the workflow life cycle
- Using global and community standards
- Integrating with the global ecosystem
- Faithfully representing workflows
- Streamlined processes
- **Fit-for-purpose data model**

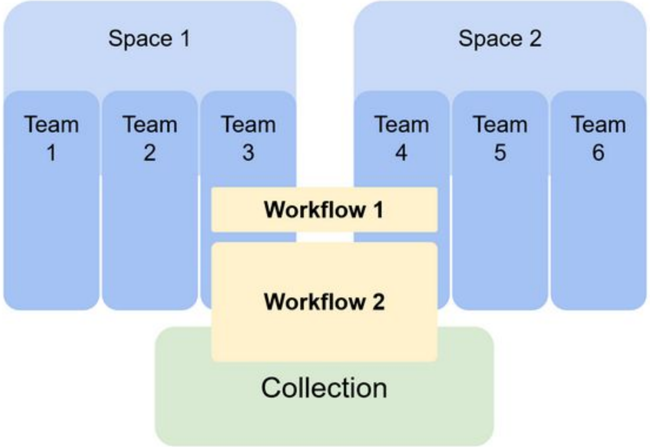
Note: these structures are flexible

Scenario A



- 3 *Teams* from space one, and 3 *Teams* from space two have created a *workflow* each
- These are both part of the same *collection*
- *Spaces* are not sharing a *workflow*, but do contribute to the same *Collection*

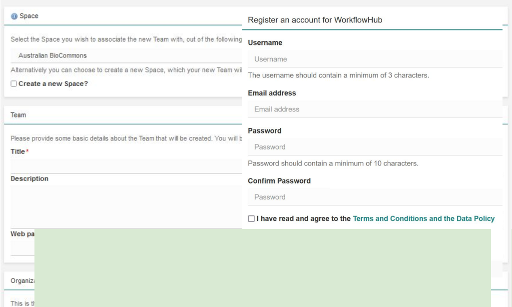
Scenario B



- Team #3 from space one, and team #4 from space two have collaborated to create 2x *workflows*
- *Workflow* 2 is contributed to a collection
- *Spaces* are sharing *workflows* via their component *Teams*, but do not directly share the *workflows*

- Supporting the workflow life cycle
- Using global and community standards
- Integrating with the global ecosystem
- Faithfully representing workflows
- Streamlined processes
- Fit-for-purpose data model
- **User support**

Wizards and streamlined processes



Space

Select the Space you wish to associate the new Team with, out of the following

Australian BioCommons

Alternatively you can choose to create a new Space, which your new Team wil

☐ **Create a new Space?**

Team

Please provide some basic details about the Team that will be created. You will b

Title*

Description

Web pa

Organiz

This is t

Register an account for WorkflowHub

Username

Username

The username should contain a minimum of 3 characters.

Email address

Email address

Password

Password

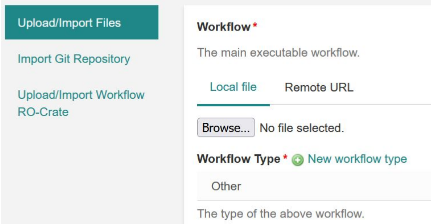
Password should contain a minimum of 10 characters.

Confirm Password

Password

☐ **I have read and agree to the Terms and Conditions and the Data Policy**

Set up



Workflow*

The main executable workflow.

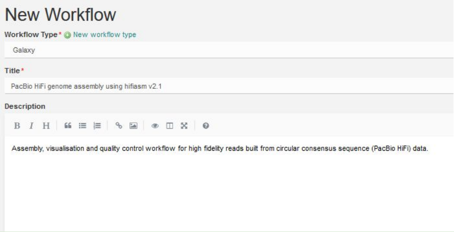
No file selected.

Workflow Type* New workflow type

Other

The type of the above workflow.

Workflow upload



New Workflow

Workflow Type* New workflow type

Galaxy

Title*

PacBio HiFi genome assembly using hifiasm v2.1

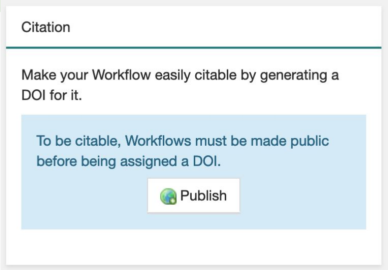
Description

Assembly, visualisation and quality control workflow for high fidelity reads built from circular consensus sequence (PacBio HiFi) data.

Metadata entry

DOI minting

Via DataCite

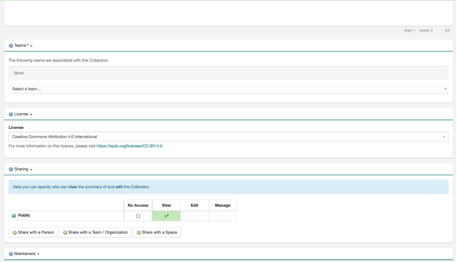


Citation

Make your Workflow easily citable by generating a DOI for it.

To be citable, Workflows must be made public before being assigned a DOI.

Creating Collections



lines: 1 words: 0 0:0

Teams* ▴

The following teams are associated with this Collection:

None

Select a team...

License ▴

License

Creative Commons Attribution 4.0 International

For more information on this license, please visit https://spdx.org/licenses/CC-BY-4.0

Sharing ▴

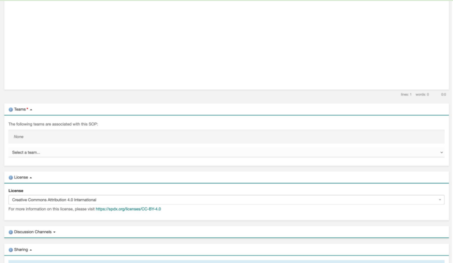
Here you can specify who can **view** the summary of and **edit** the Collection.

	No Access	View	Edit	Manage
Public	○	✓		

Maintainers ▾

Registering context documents

i.e. SOPs, publications



lines: 1 words: 0 0:0

Teams* ▴

The following teams are associated with this SOP:

None

Select a team...

License ▴

License

Creative Commons Attribution 4.0 International

For more information on this license, please visit https://spdx.org/licenses/CC-BY-4.0

Discussion Channels ▾

Sharing ▴

Supporting the workflow life cycle
Using global and community standards
Integrating with the global ecosystem
Faithfully representing workflows
Streamlined processes
Fit-for-purpose data model
User support

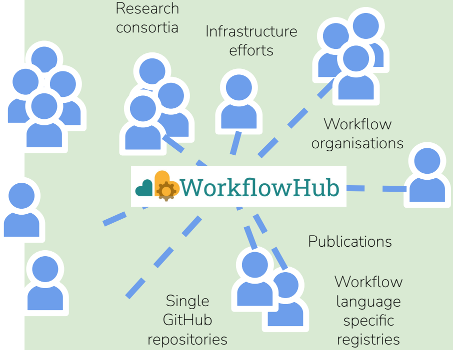
User engagement, and training

WorkflowHub Club

WorkflowHub Club coordination call

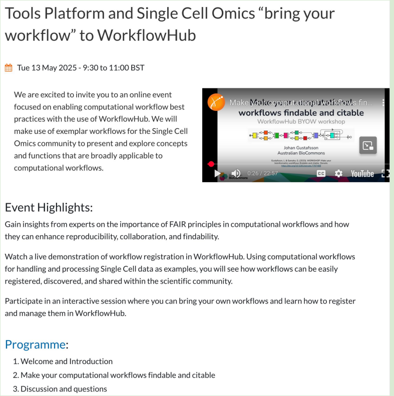
This is the rolling minutes from the WorkflowHub telcons. **Feel free** to join the mailing list, **suggest changes** to the agenda or to help **scribe** the minutes of the call!

Regular schedule is every other **Wednesday** 10:00 current UK time zone / 11:00 current European Time zone. When the clocks change we still stick to the 10:00/11:00 regardless. So winter it's GMT/CET and summer it's BST/CEST.



https://s.apache.org/workflowhub-minutes

Webinars / Workshops



Tools Platform and Single Cell Omics "bring your workflow" to WorkflowHub

Tue 13 May 2025 - 9:30 to 11:00 BST

We are excited to invite you to an online event focused on enabling computational workflow best practices with the use of WorkflowHub. We will make use of exemplar workflows for the Single Cell Omics community to present and explore concepts and functions that are broadly applicable to computational workflows.

Event Highlights:

Gain insights from experts on the importance of FAIR principles in computational workflows and how they can enhance reproducibility, collaboration, and findability.

Watch a live demonstration of workflow registration in WorkflowHub. Using computational workflows for handling and processing Single Cell data as examples, you will see how workflows can be easily registered, discovered, and shared within the scientific community.

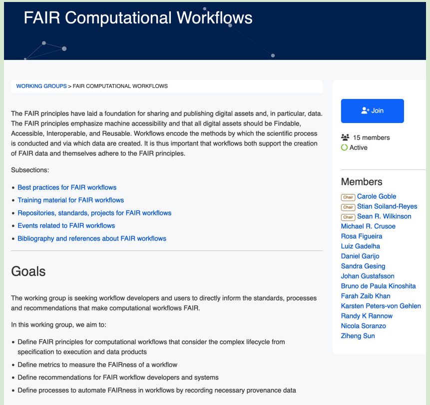
Participate in an interactive session where you can bring your own workflows and learn how to register and manage them in WorkflowHub.

Programme:

1. Welcome and Introduction
2. Make your computational workflows findable and citable
3. Discussion and questions

Gustafsson, O. J. R. (2025, June 4). Make your computational workflows findable and citable: WorkflowHub BYOW workshop. Zenodo. https://doi.org/10.5281/zenodo.15588745

FAIR workflows (WCI)



FAIR Computational Workflows

WORKING GROUPS > FAIR COMPUTATIONAL WORKFLOWS

The FAIR principles have laid a foundation for sharing and publishing digital assets and, in particular, data. The FAIR principles emphasize machine accessibility and that all digital assets should be Findable, Accessible, Interoperable, and Reusable. Workflows encode the methods by which the scientific process is conducted and via which data are created. It is thus important that workflows both support the creation of FAIR data and themselves adhere to the FAIR principles.

Subsections:

- Best practices for FAIR workflows
- Training material for FAIR workflows
- Repositories, standards, projects for FAIR workflows
- Events related to FAIR workflows
- Bibliography and references about FAIR workflows

Goals

The working group is seeking workflow developers and users to directly inform the standards, processes and recommendations that make computational workflows FAIR.

In this working group, we aim to:

- Define FAIR principles for computational workflows that consider the complex lifecycle from specification to execution and data products
- Define metrics to measure the FAIRness of a workflow
- Define recommendations for FAIR workflow developers and systems
- Define processes to automate FAIRness in workflows by recording necessary provenance data

15 members
Active

Members

Chair Carole Goble
Chair Stian Soiland-Reyes
Chair Sean R. Wilkinson
Michael R. Crusoe
Rosa Figueira
Luiz Gadelha
Daniel Garijo
Sandra Gesing
Johan Gustafsson
Bruno de Paula Kinoshita
Farah Zaib Khan
Karsten Peters-von Gehlen
Randy K Rannow
Nicola Soranzo
Ziheng Sun

Wilkinson, S.R., Aloqalaa, M., Belhajjame, K. et al. Applying the FAIR Principles to computational workflows. Sci Data 12, 328 (2025). https://doi.org/10.1038/s41597-025-04451-9

Use cases

nf-core

A global community effort to collect a curated set of open-source analysis pipelines built using Nextflow.

https://workflowhub.eu/projects/15

Galaxy Training Network

https://workflowhub.eu/projects/12

ERGA

European Reference Genome Atlas

https://workflowhub.eu/programmes/33

Biodiversity Genomics Europe

https://workflowhub.eu/programmes/25

BIODT

biodiversitydigitaltwin

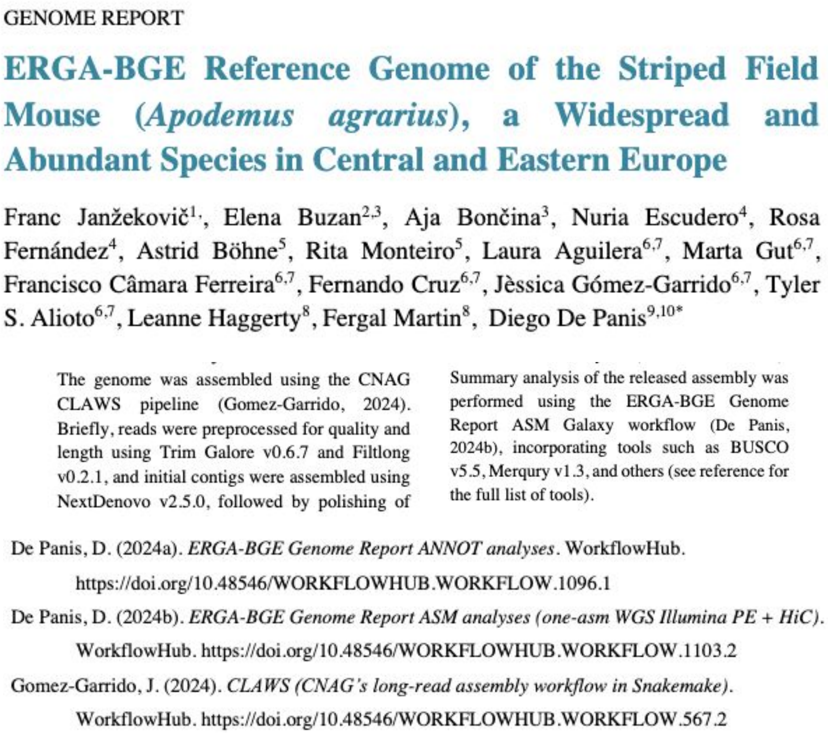
https://workflowhub.eu/programmes/22

Australian BioCommons

https://workflowhub.eu/programmes/8

ERGA-BGE

Cite workflows in methods section of genome reports/papers, credit workflow developers



GENOME REPORT

ERGA-BGE Reference Genome of the Striped Field Mouse (*Apodemus agrarius*), a Widespread and Abundant Species in Central and Eastern Europe

Franc Janžekovič[1], Elena Buzan[2,3], Aja Bončina[3], Nuria Escudero[4], Rosa Fernández[4], Astrid Böhne[5], Rita Monteiro[5], Laura Aguilera[6,7], Marta Gut[6,7], Francisco Câmara Ferreira[6,7], Fernando Cruz[6,7], Jèssica Gómez-Garrido[6,7], Tyler S. Alioto[6,7], Leanne Haggerty[8], Fergal Martin[8], Diego De Panis[9,10*]

The genome was assembled using the CNAG CLAWS pipeline (Gomez-Garrido, 2024). Briefly, reads were preprocessed for quality and length using Trim Galore v0.6.7 and Filtlong v0.2.1, and initial contigs were assembled using NextDenovo v2.5.0, followed by polishing of

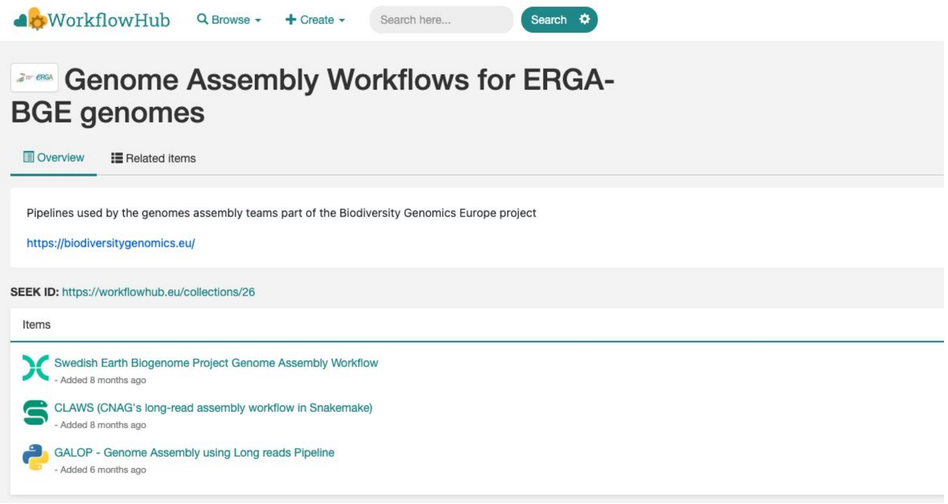
Summary analysis of the released assembly was performed using the ERGA-BGE Genome Report ASM Galaxy workflow (De Panis, 2024b), incorporating tools such as BUSCO v5.5, Merqury v1.3, and others (see reference for the full list of tools).

De Panis, D. (2024a). *ERGA-BGE Genome Report ANNOT analyses*. WorkflowHub. https://doi.org/10.48546/WORKFLOWHUB.WORKFLOW.1096.1

De Panis, D. (2024b). *ERGA-BGE Genome Report ASM analyses (one-asm WGS Illumina PE + HiC)*. WorkflowHub. https://doi.org/10.48546/WORKFLOWHUB.WORKFLOW.1103.2

Gomez-Garrido, J. (2024). *CLAWS (CNAG's long-read assembly workflow in Snakemake)*. WorkflowHub. https://doi.org/10.48546/WORKFLOWHUB.WORKFLOW.567.2

https://doi.org/10.3897/arphapreprints.e154773



Genome Assembly Workflows for ERGA-BGE genomes

Pipelines used by the genomes assembly teams part of the Biodiversity Genomics Europe project

https://biodiversitygenomics.eu/

SEEK ID: https://workflowhub.eu/collections/26

Items

- Swedish Earth Biogenome Project Genome Assembly Workflow - Added 8 months ago
- CLAWS (CNAG's long-read assembly workflow in Snakemake) - Added 8 months ago
- GALOP - Genome Assembly using Long reads Pipeline - Added 6 months ago

https://workflowhub.eu/collections/26

Slide credit Tom Brown

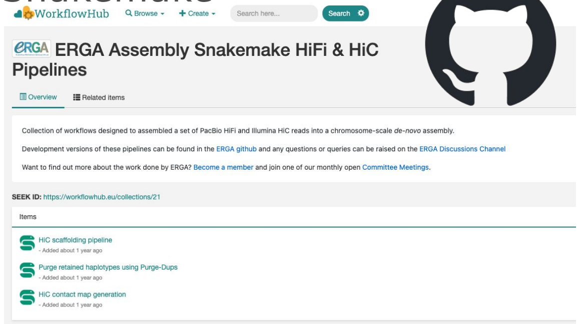
EUROPEAN REFERENCE GENOME ATLAS

Biodiversity Genomics Europe

ERGA-BGE

Collect workflows and best practices from the Reference Genome Community of Practice

Snakemake



ERGA Assembly Snakemake HiFi & HiC Pipelines

Collection of workflows designed to assembled a set of PacBio HiFi and Illumina HiC reads into a chromosome-scale *de-novo* assembly.

Development versions of these pipelines can be found in the ERGA github and any questions or queries can be raised on the ERGA Discussions Channel

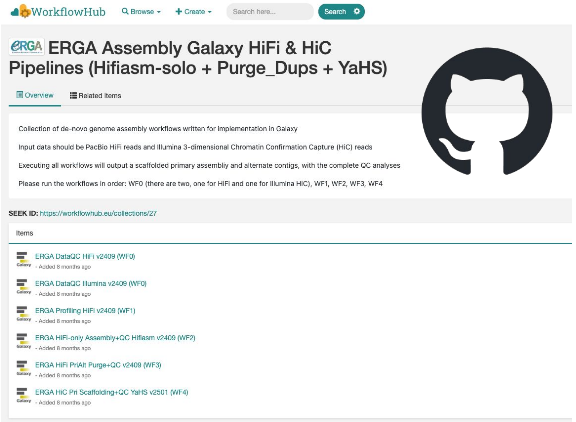
Want to find out more about the work done by ERGA? Become a member and join one of our monthly open Committee Meetings.

SEEK ID: https://workflowhub.eu/collections/21

Items

- HiC scaffolding pipeline - Added about 1 year ago
- Purge retained haplotypes using Purge-Dups - Added about 1 year ago
- HiC contact map generation - Added about 1 year ago

Galaxy EUROPE



ERGA Assembly Galaxy HiFi & HiC Pipelines (Hifiasm-solo + Purge_Dups + YaHS)

Collection of de-novo genome assembly workflows written for implementation in Galaxy

Input data should be PacBio HiFi reads and Illumina 3-dimensional Chromatin Confirmation Capture (HiC) reads

Executing all workflows will output a scaffolded primary assembliy and alternate contigs, with the complete QC analyses

Please run the workflows in order: WF0 (there are two, one for HiFi and one for Illumina HiC), WF1, WF2, WF3, WF4

SEEK ID: https://workflowhub.eu/collections/27

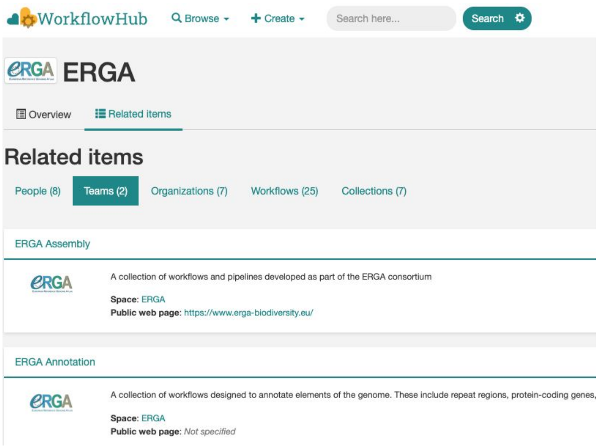
Items

- ERGA DataQC HiFi v2409 (WF0) - Added 8 months ago
- ERGA DataQC Illumina v2409 (WF0) - Added 8 months ago
- ERGA Profiling HiFi v2409 (WF1) - Added 8 months ago
- ERGA HiFi-only Assembly+QC Hifiasm v2409 (WF2) - Added 8 months ago
- ERGA HiFi PriAlt Purge+QC v2409 (WF3) - Added 8 months ago
- ERGA HiC Pri Scaffolding+QC YaHS v2501 (WF4) - Added 8 months ago

WM Language-agnostic pipelines for generation of high-quality genome assemblies and annotations

Collected from and developed with the community of researchers developing and implementing workflows for assembly and annotation of reference genomes

Training and capacity building



ERGA

Related items

ERGA Assembly

A collection of workflows and pipelines developed as part of the ERGA consortium

Space: ERGA
Public web page: https://www.erga-biodiversity.eu/

ERGA Annotation

A collection of workflows designed to annotate elements of the genome. These include repeat regions, protein-coding genes,

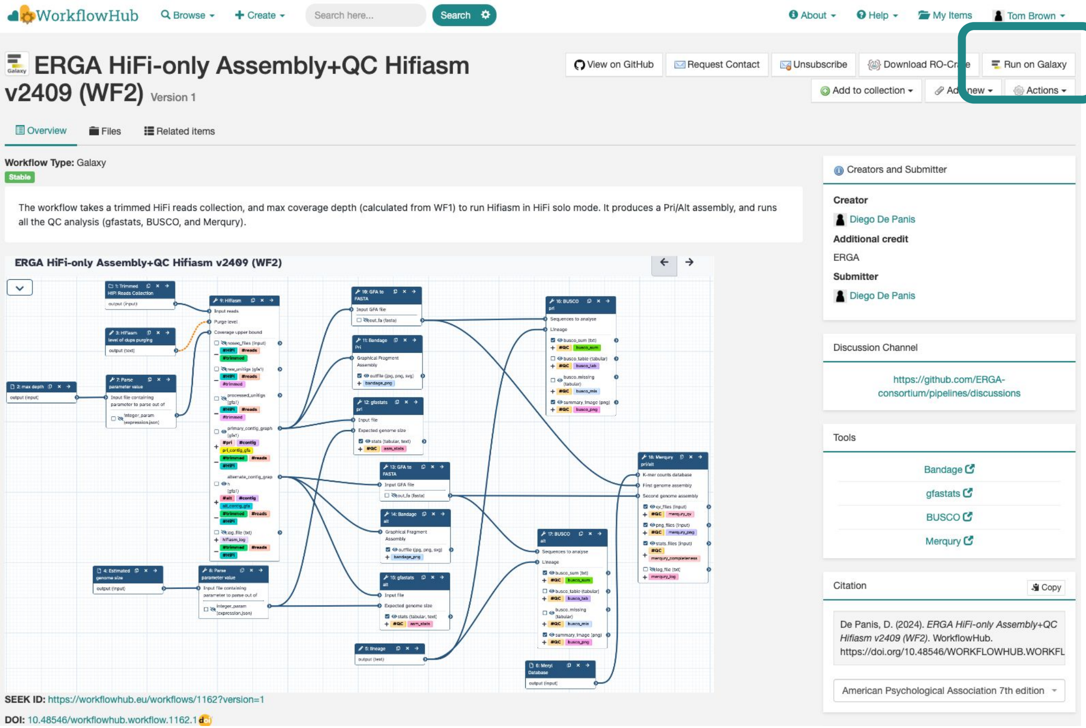
Space: ERGA
Public web page: *Not specified*

ERGA space in WorkflowHub

Slide credit Tom Brown

ERGA-BGE

Reuse existing datasets, curated workflows and the Galaxy workflow infrastructure: launch automatically via WorkflowHub, collect metrics



ERGA HiFi-only Assembly+QC Hifiasm v2409 (WF2) Version 1

Workflow Type: Galaxy

Stable

The workflow takes a trimmed HiFi reads collection, and max coverage depth (calculated from WF1) to run Hifiasm in HiFi solo mode. It produces a Pri/Alt assembly, and runs all the QC analysis (gfastats, BUSCO, and Merqury).

ERGA HiFi-only Assembly+QC Hifiasm v2409 (WF2)

SEEK ID: https://workflowhub.eu/workflows/1162?version=1

DOI: 10.48546/workflowhub.workflow.1162.1

Creators and Submitter

Creator
Diego De Panis

Additional credit
ERGA

Submitter
Diego De Panis

Discussion Channel

https://github.com/ERGA-consortium/pipelines/discussions

Tools

- Bandage
- gfastats
- BUSCO
- Merqury

Citation

De Panis, D. (2024). *ERGA HiFi-only Assembly+QC Hifiasm v2409 (WF2)*. WorkflowHub. https://doi.org/10.48546/WORKFLOWHUB.WORKFL

Automatically import workflow to Galaxy current dataset history ready for immediate execution

Slide credit Tom Brown

Some challenges and observations

Registration - getting it! Even for projects funded to do so

Automate from GitHub, nurture community champions, BYOW sessions

Metadata - getting it!

Automate from GitHub practices, partnerships with WfMS developers

Range of Workflow forms

From HPC where the machine matters, simulations where the run set up matters, to R and Python scripts

Registry / Repository

A record of active software held elsewhere vs file and forget archive

Bot Battering

Open comes with a down side - AI bot attacks

Some plans and prospects

Sustainability

Service for National and International Infrastructures: European (ELIXIR, Eurobioimaging..), European Open Science Cloud, Australian BioCommons...

Integrations and Discipline widening

Crosswalks with other registries (Lifewatch, MethodsHub, DockStore...)

More added-value features using the power of AI

Full fat FAIR workflow support, metadata acquisition and curation, workflow similarity tracking ...
Smoothing support

Embedded into Journal, Funder and Policy maker FAIR practice

FAIR Workflow management planning

Community champions network

Some Acknowledgements

WorkflowHub

The University of Manchester, UK


Finn
Bacall


Nick
Juty


Stuart
Owen


Alan
Williams


Frederik Coppens
VIB, Belgium

workflows community *initiative*


Sean Wilkinson
Oak Ridge National Labs, USA

Workflow RO-Crate


Simone
Leo
CR4S,
Sardinia

The University of Manchester, UK


Stian
Soiland-Reyes


Eli
Chadwick


Biodiversity Genomics Europe

ERGA European Reference Genome Atlas

Tom Brown, *Leibniz Institute for Zoo and Wildlife Research, Berlin, Germany*

Galaxy PROJECT

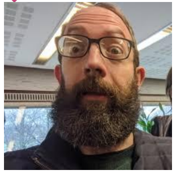

Björn Grüning,
Freiburg University,
Germany

nextflow


Phil Ewels, *Seqera*

Tim Booth, *U of Edinburgh, UK*

COMMON WORKFLOW LANGUAGE


Michael Crusoe,
CWL

Some Acknowledgements

https://workflowhub.org

BY-COVID
beyond covid

Workflow RO-Crate

Workflow Run RO-Crate

workflows community initiative

IBISBA

Life Monitor

Galaxy PROJECT

Biodiversity Genomics Europe

MGnify

EOSC-Life

nf-core

A global community effort to collect a curated set of open-source analysis pipelines built using Nextflow.

elixir

COMMON WORKFLOW LANGUAGE

EDAM

nextflow

THE UNIVERSITY OF MELBOURNE

Bioindustry4.0
Advanced digital solutions

Bioschemas

Australian BioCommons

zenodo

bio.tools

FAIRDOM

SEEK

BIOPLATFORMS AUSTRALIA

NCRIS
National Research Infrastructure for Australia
An Australian Government Initiative

BIODT
biodiversitydigitaltwin

AgroServ
Integrated SERvices supporting a sustainable AGROecological transition

TRE-FX

eosc | EuroScienceGateway