# Visible & reusable workflows

Ove Johan Ragnar Gustafsson, Georgina Samaha, Ziad Al Bkhetan, Paula Andrea Martinez, Finn Bacall, Carole Goble, Steven Manos, Nigel Ward, Jeff Christiansen

# Acknowledgement of Country

We acknowledge the Turrbal and Jagera/Yuggera people as Traditional Owners and custodians of the lands on which we meet today and pay our respects to their Ancestors and their descendants.

Meanjin (Brisbane), Australia

# Acknowledgements

Pawsey

THE UNIVERSITY OF MELBOURNE

Galaxy AUSTRALIA

THE UNIVERSITY OF SYDNEY

WorkflowHub

NCI AUSTRALIA

THREATENED SPECIES INITIATIVE

BioCommons ‘Bring Your Own Data’ Expansion Project
https://www.biocommons.org.au/byo-data-platform-expansion

AUSTRALIAN ACCESS FEDERATION

Melbourne Bioinformatics

QCIF

ARDC
Australian Research Data Commons

elixir

EDAM ontology
EDAM is an ontology of bioscientific data analysis and data management

NCRIS
National Research Infrastructure for Australia
An Australian Government Initiative

aarnet
Australia's Academic and Research Network

elixir bio.tools

BIOPLATFORMS AUSTRALIA

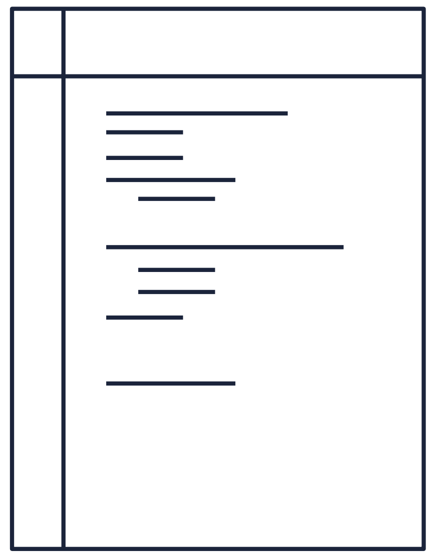
and a special thanks to all our research collaborators!

# From code snippets to workflows at scale

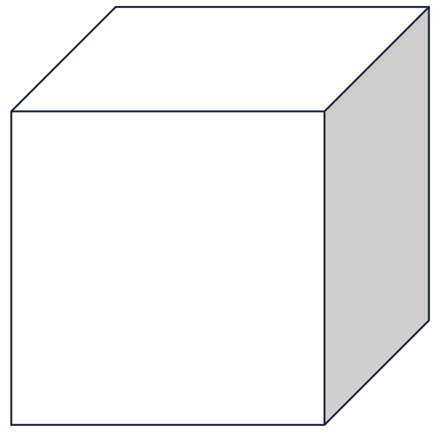
Workflows exist on a spectrum ranging from

- **heterogeneous composites of code** through to
- **production level pipelines**

.sh

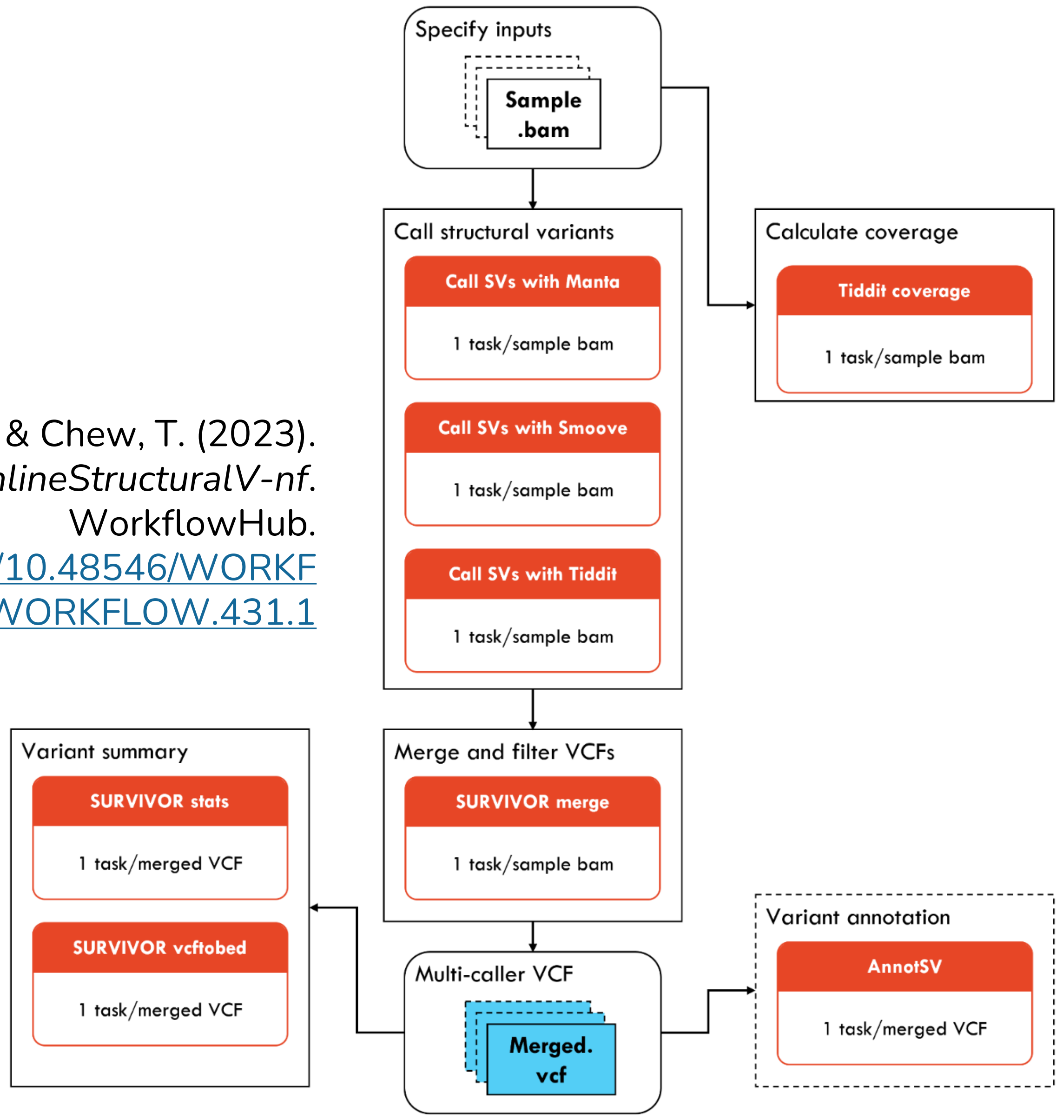


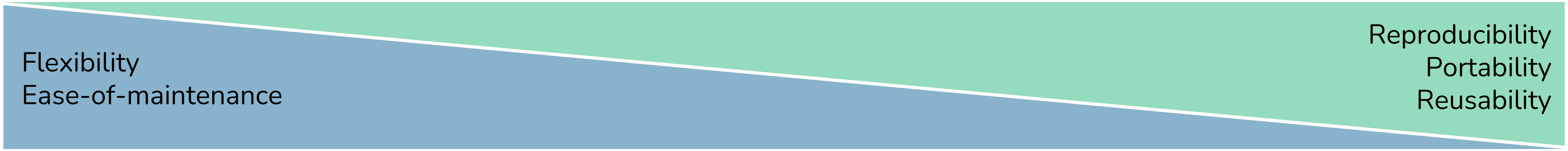
GitHub



WMS

Samaha, G., & Chew, T. (2023). *GermlineStructuralV-nf*. WorkflowHub. https://doi.org/10.48546/WORKFLOWHUB.WORKFLOW.431.1



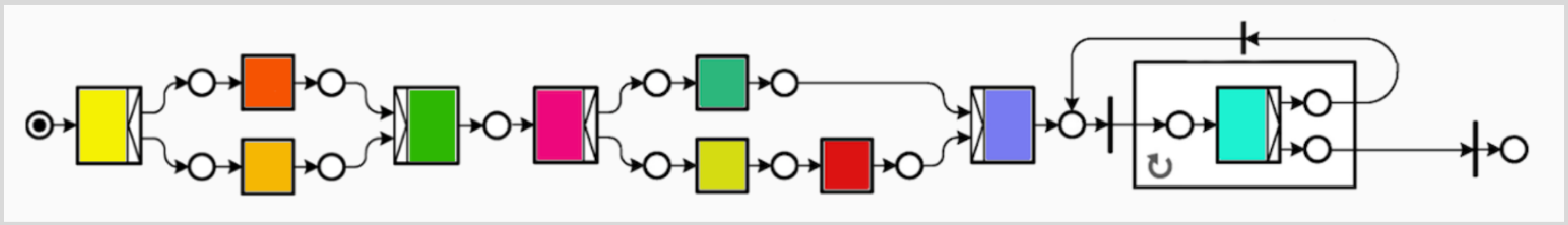


Flexibility
Ease-of-maintenance

Reproducibility
Portability
Reusability

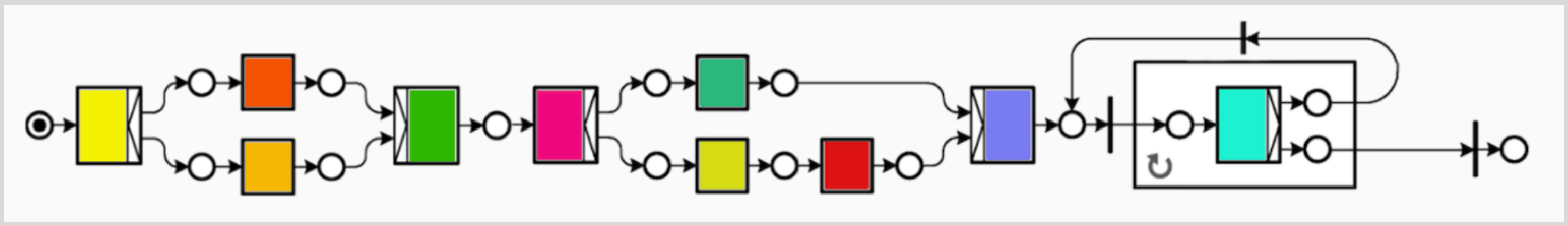
# We typically only see the tip of the iceberg

**What a user might see**



# We typically only see the tip of the iceberg

**What a user might see**



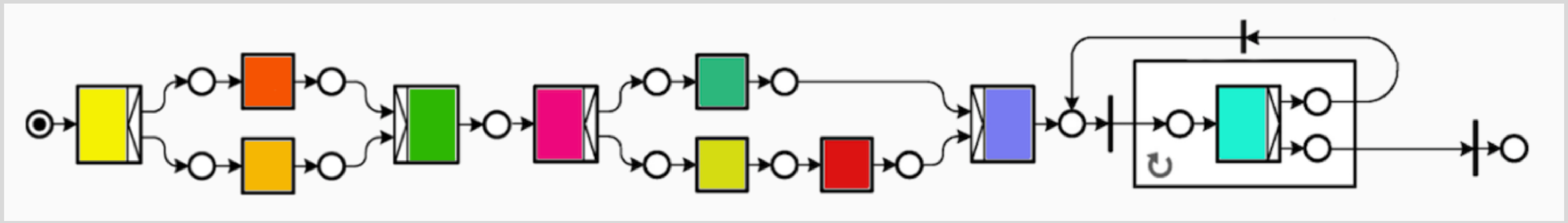
**The reality**

- ✓ Workflow creation
- ✓ Managing dependencies
- ✓ Infrastructure optimisation
- ✓ Testing
- ✓ Maintenance
- ✓ Workflow management system conversions

Australian BioCommons

# We typically only see the tip of the iceberg

**What a user might see**



**The reality**

- ✓ Workflow creation
- ✓ Managing dependencies
- ✓ Infrastructure optimisation
- ✓ Testing
- ✓ Maintenance
- ✓ Workflow management system conversions

**Requiring**

- ✓ Time
- ✓ Domain specific expertise
- ✓ Coding / workflow language expertise
- ✓ Knowledge of best practice

# Visibility and reusability

Australian
BioCommons

# Visibility and reusability

## Best practice

1. **Version control** using a popular repository platform (e.g. GitHub)
2. **Open source** your workflow from day 1
3. **License** for reuse by others
4. **Document** your workflow such that it can be understood and reused
5. **Follow gold standard examples** from your domain (e.g. genomics) or workflow community (Galaxy IWC, nf-core)
6. **Make your workflow modular and portable** by using containers and workflow languages
7. **ORCID** - having and using an ORCID - connecting to visibility services
8. **Register your workflow** (e.g. WorkflowHub / Dockstore) - connecting to scholarly knowledge graph, surfacing metadata for search engines

# Visibility and reusability

## FAIR

1. **Common workflow description** that is independent of a platform (e.g. CWL, WDL)
2. **Standard machine-readable metadata**
3. **Common metadata** for the workflow, its tools & input parameters (Bioschemas, CFF, EDAM)
4. **Digital Object**: for example one for packaging workflows + metadata + associated data objects + logs (Workflow RO-crate)

# Visibility and reusability

## FAIR

1. **Common workflow description** that is independent of a platform (e.g. CWL, WDL)
2. **Standard machine-readable metadata**
3. **Common metadata** for the workflow, its tools & input parameters (Bioschemas, CFF, EDAM)
4. **Digital Object**: for example one for packaging workflows + metadata + associated data objects + logs (Workflow RO-crate)

# Visibility and reusability

## FAIR

1. **Common workflow description** that is independent of a platform (e.g. CWL, WDL)
2. **Standard machine-readable metadata**
3. **Common metadata** for the workflow, its tools & input parameters (Bioschemas, CFF, EDAM)
4. **Digital Object**: for example one for packaging workflows + metadata + associated data objects + logs (Workflow RO-crate)

## Creators

Visibility of experts

# Visibility and reusability

## FAIR

1. **Common workflow description** that is independent of a platform (e.g. CWL, WDL)
2. **Standard machine-readable metadata**
3. **Common metadata** for the workflow, its tools & input parameters (Bioschemas, CFF, EDAM)
4. **Digital Object**: for example one for packaging workflows + metadata + associated data objects + logs (Workflow RO-crate)

## Creators

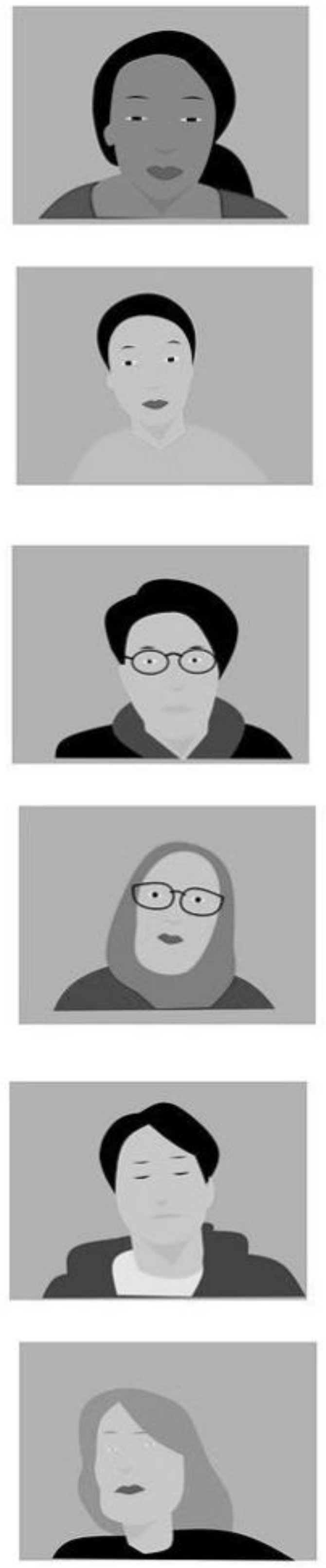
### Visibility of experts

Funding
Career opportunities

# Visibility and reusability

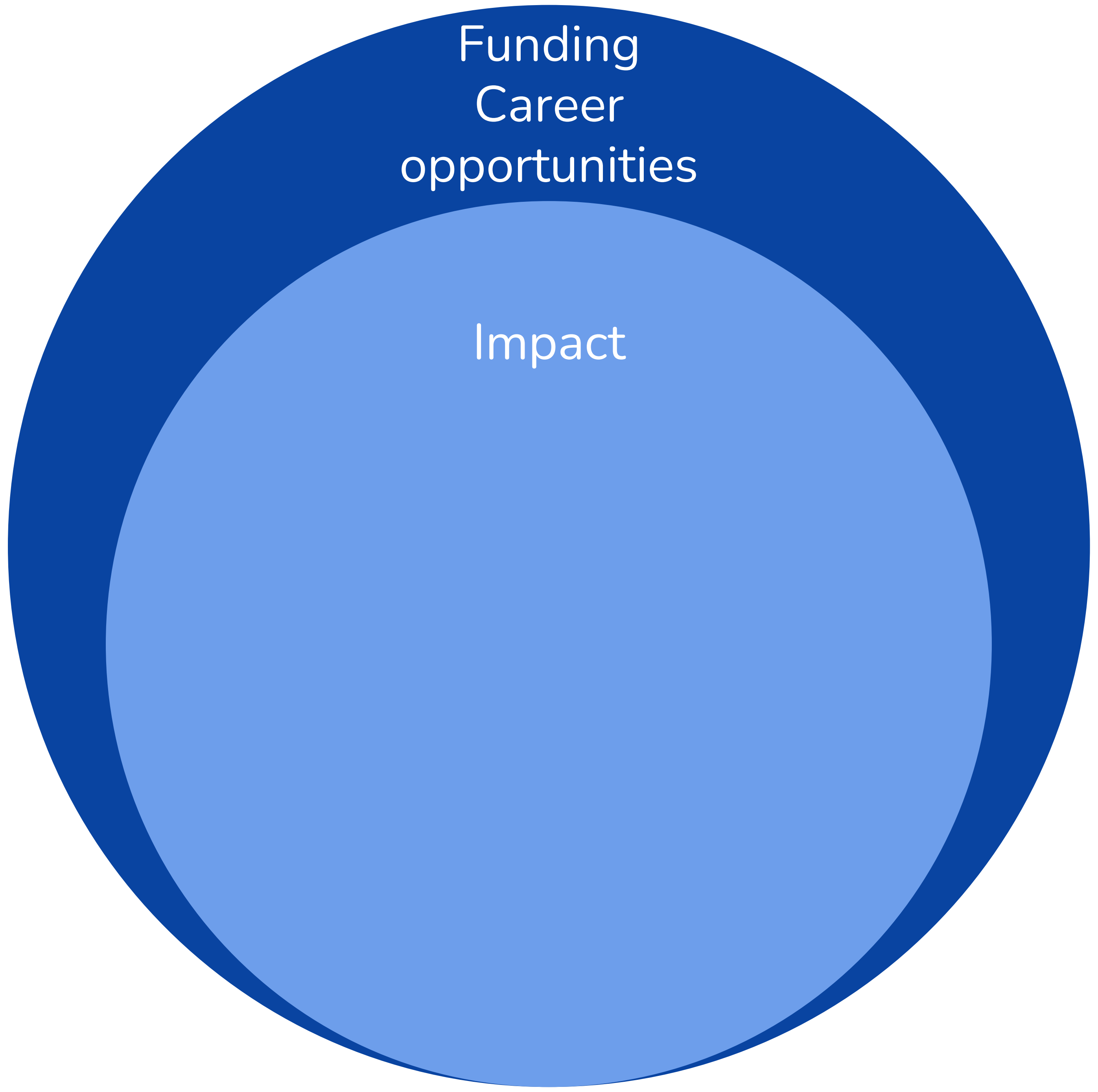
## FAIR

1. **Common workflow description** that is independent of a platform (e.g. CWL, WDL)
2. **Standard machine-readable metadata**
3. **Common metadata** for the workflow, its tools & input parameters (Bioschemas, CFF, EDAM)
4. **Digital Object**: for example one for packaging workflows + metadata + associated data objects + logs (Workflow RO-crate)



## Creators

### Visibility of experts



Funding
Career
opportunities

Impact

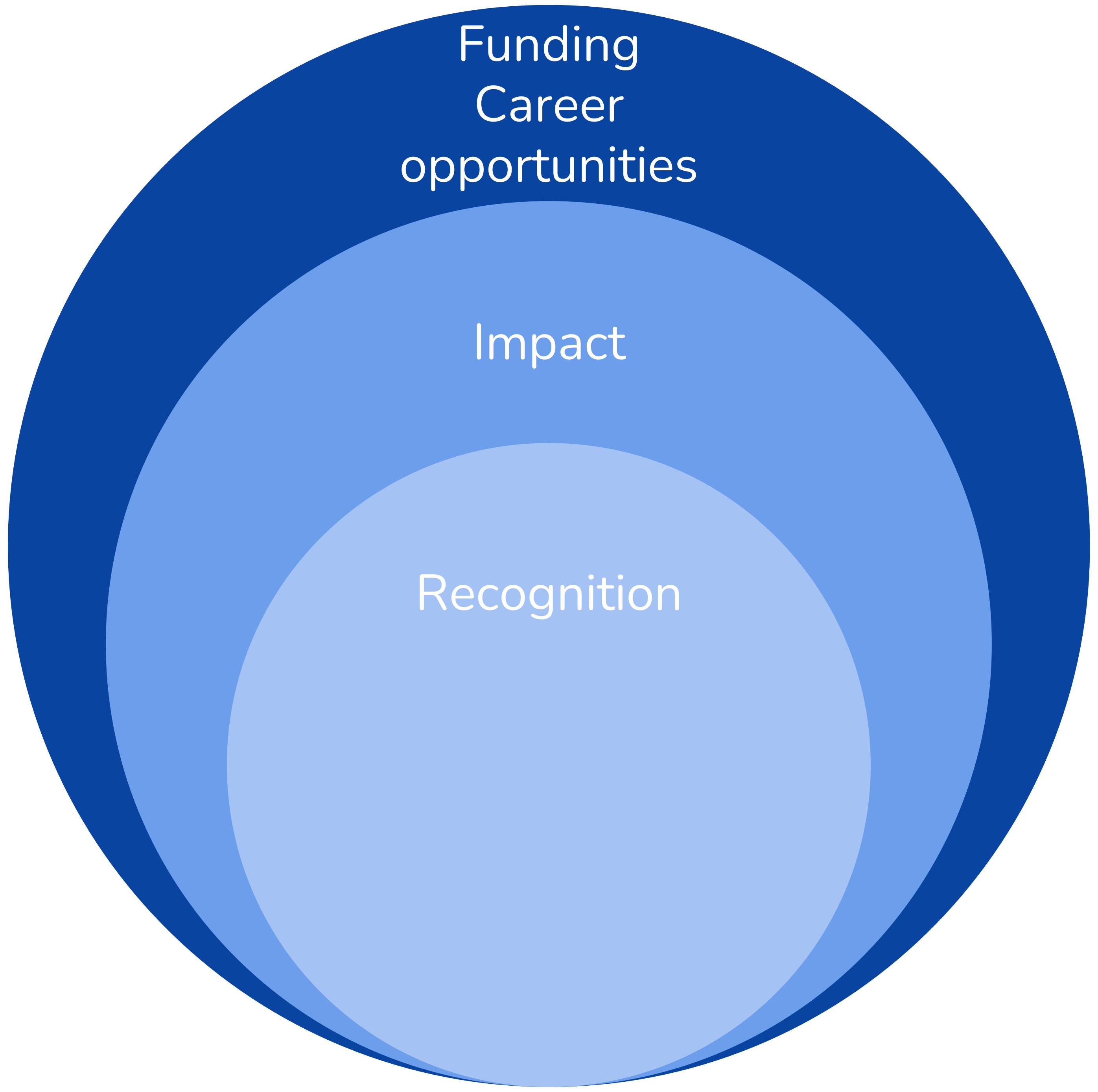
# Visibility and reusability

## FAIR

1. **Common workflow description** that is independent of a platform (e.g. CWL, WDL)
2. **Standard machine-readable metadata**
3. **Common metadata** for the workflow, its tools & input parameters (Bioschemas, CFF, EDAM)
4. **Digital Object**: for example one for packaging workflows + metadata + associated data objects + logs (Workflow RO-crate)

## Creators

### Visibility of experts



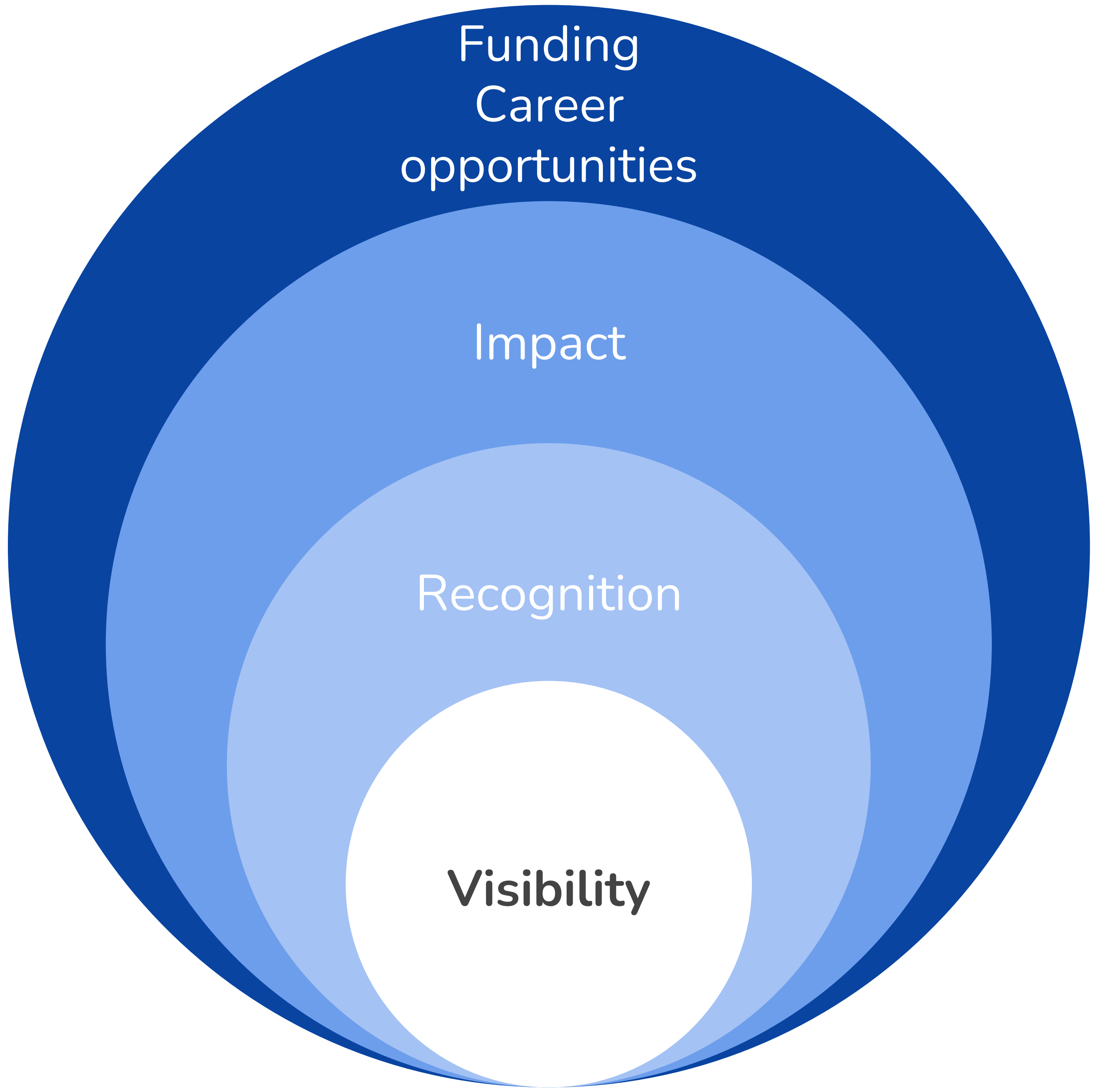
# Visibility and reusability

## FAIR

1. **Common workflow description** that is independent of a platform (e.g. CWL, WDL)
2. **Standard machine-readable metadata**
3. **Common metadata** for the workflow, its tools & input parameters (Bioschemas, CFF, EDAM)
4. **Digital Object**: for example one for packaging workflows + metadata + associated data objects + logs (Workflow RO-crate)

## Creators

### Visibility of experts



Funding
Career
opportunities

Impact

Recognition

Visibility

First challenge to address is **visibility**
Second is **reusability**

First challenge to address is **visibility**
Second is **reusability**

Workflow developers need solutions to be **straightforward**, **best practice** and **widely accepted**

First challenge to address is **visibility**
Second is **reusability**

Workflow developers need solutions to be
**straightforward**, **best practice** and **widely accepted**

Infrastructure providers need this to be
**best practice** & **sustainable**

Australian BioCommons

Galaxy

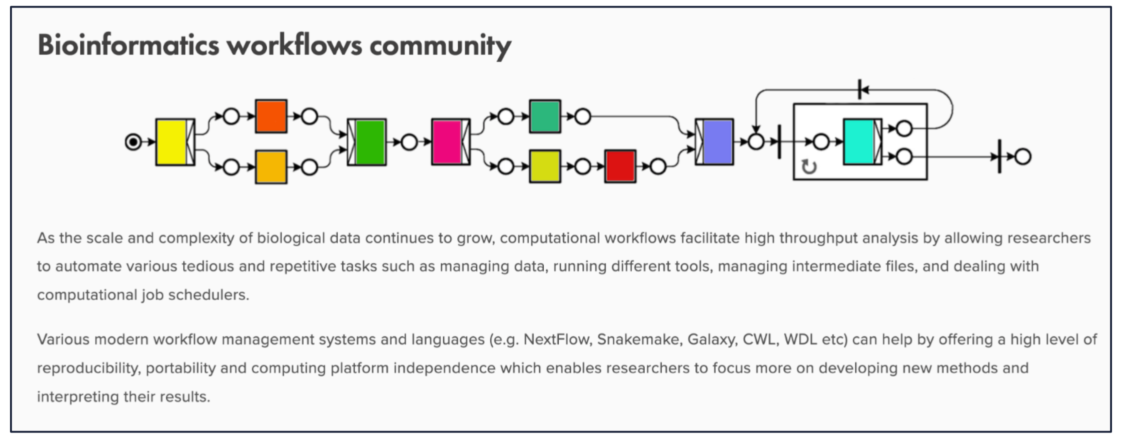
Nextflow

# Real world example: Galaxy / Nextflow

Prioritise

Communities

Data formats

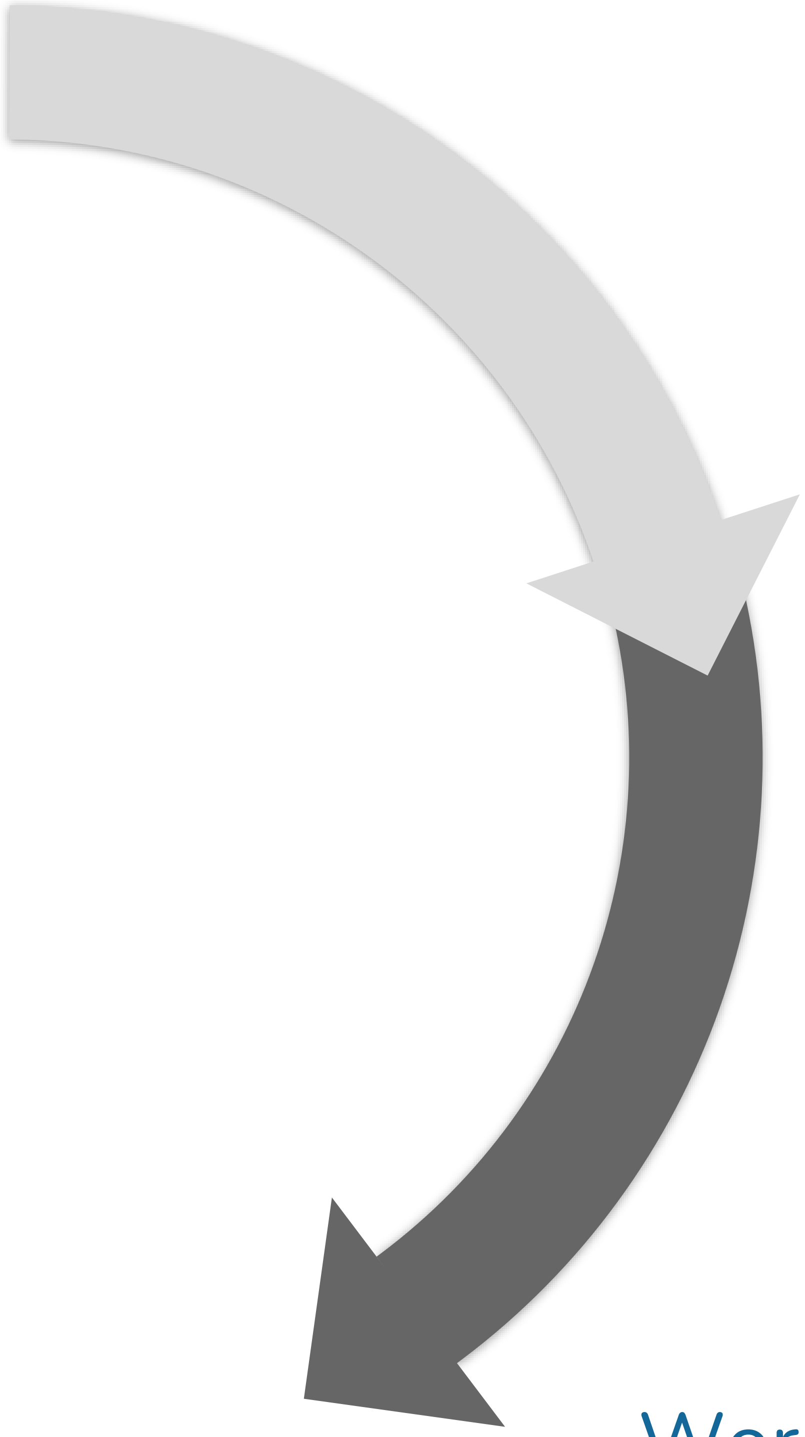


**Bioinformatics workflows community**

As the scale and complexity of biological data continues to grow, computational workflows facilitate high throughput analysis by allowing researchers to automate various tedious and repetitive tasks such as managing data, running different tools, managing intermediate files, and dealing with computational job schedulers.

Various modern workflow management systems and languages (e.g. NextFlow, Snakemake, Galaxy, CWL, WDL etc) can help by offering a high level of reproducibility, portability and computing platform independence which enables researchers to focus more on developing new methods and interpreting their results.

# Real world example: Galaxy / Nextflow

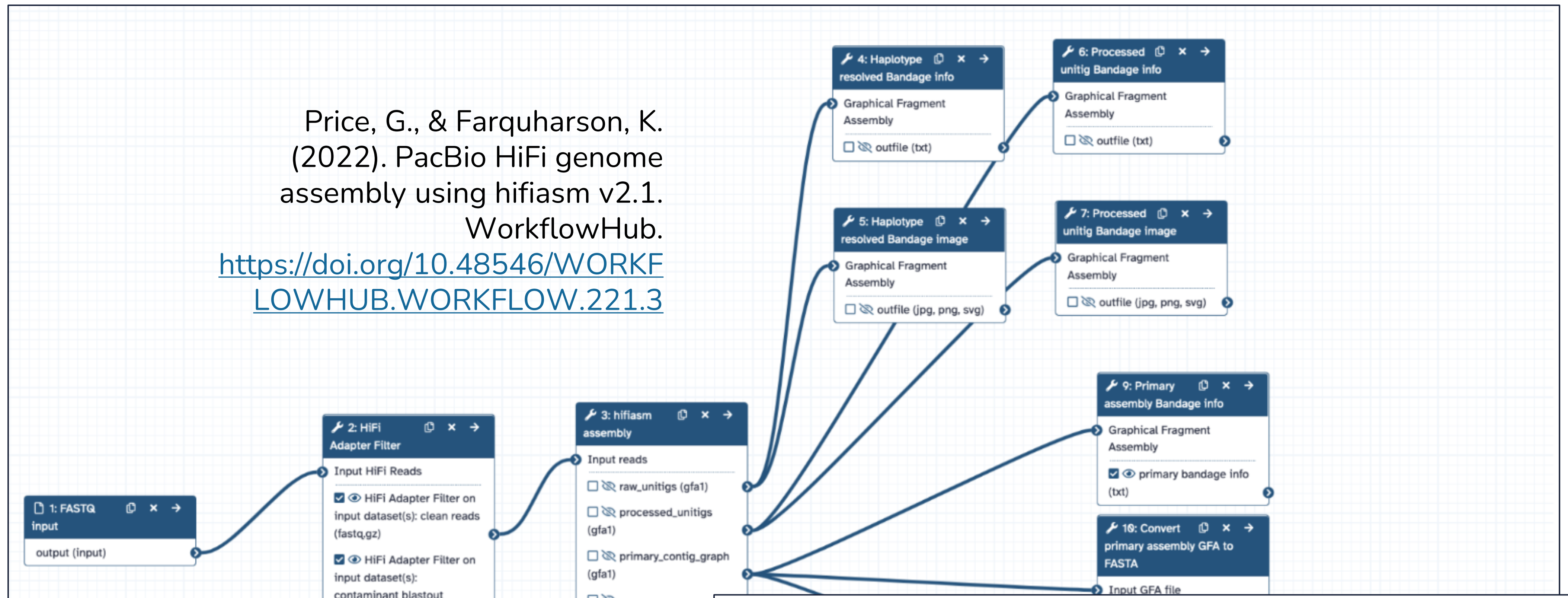


**Design / development**

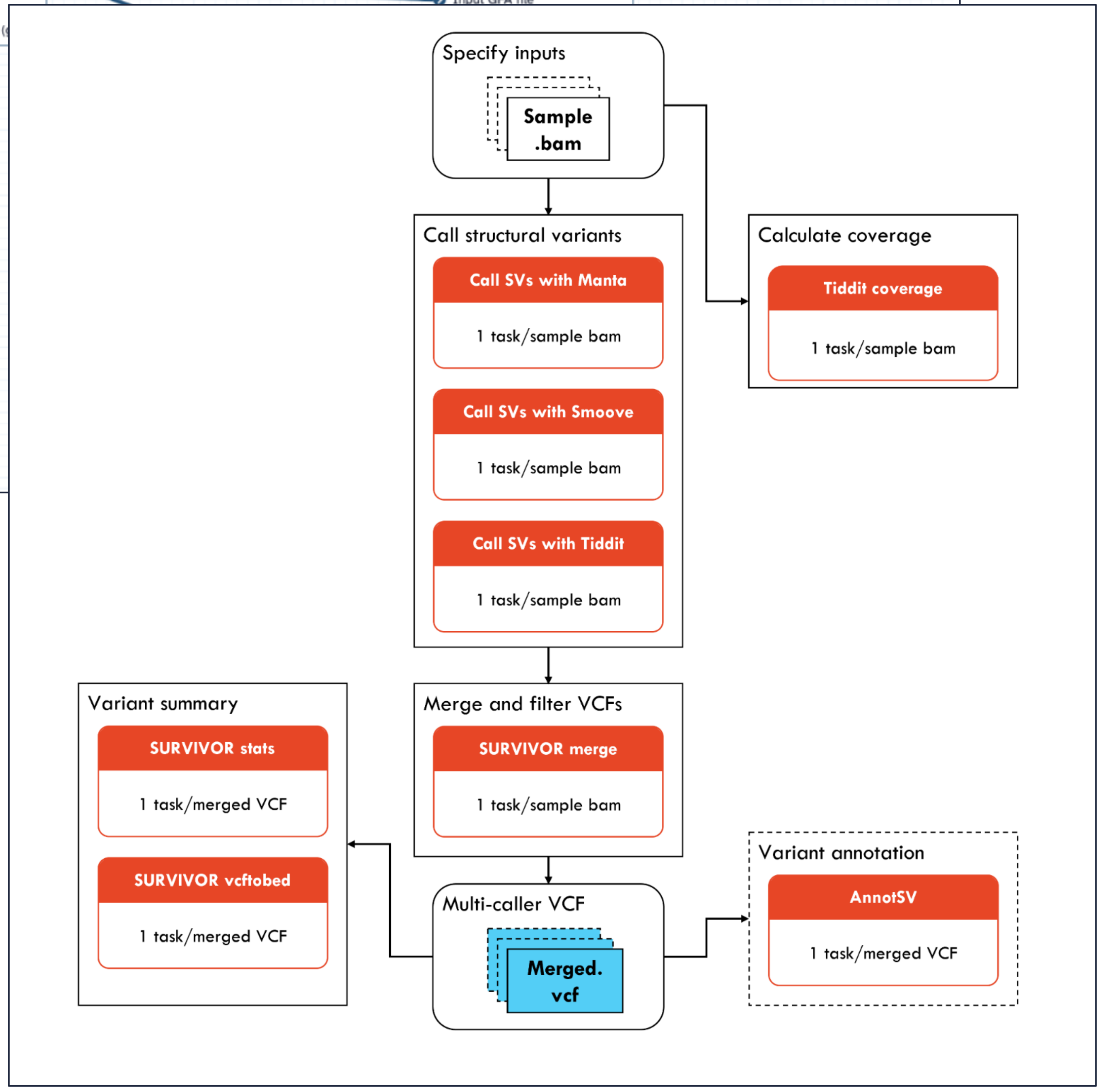
Workflow management system

Workflow management system

# Real world example: Galaxy / Nextflow



Price, G., & Farquharson, K. (2022). PacBio HiFi genome assembly using hifiasm v2.1. WorkflowHub. https://doi.org/10.48546/WORKFLOWHUB.WORKFLOW.221.3



Samaha, G., & Chew, T. (2023). *GermlineStructuralV-nf*. WorkflowHub. https://doi.org/10.48546/WORKFLOWHUB.WORKFLOW.431.1

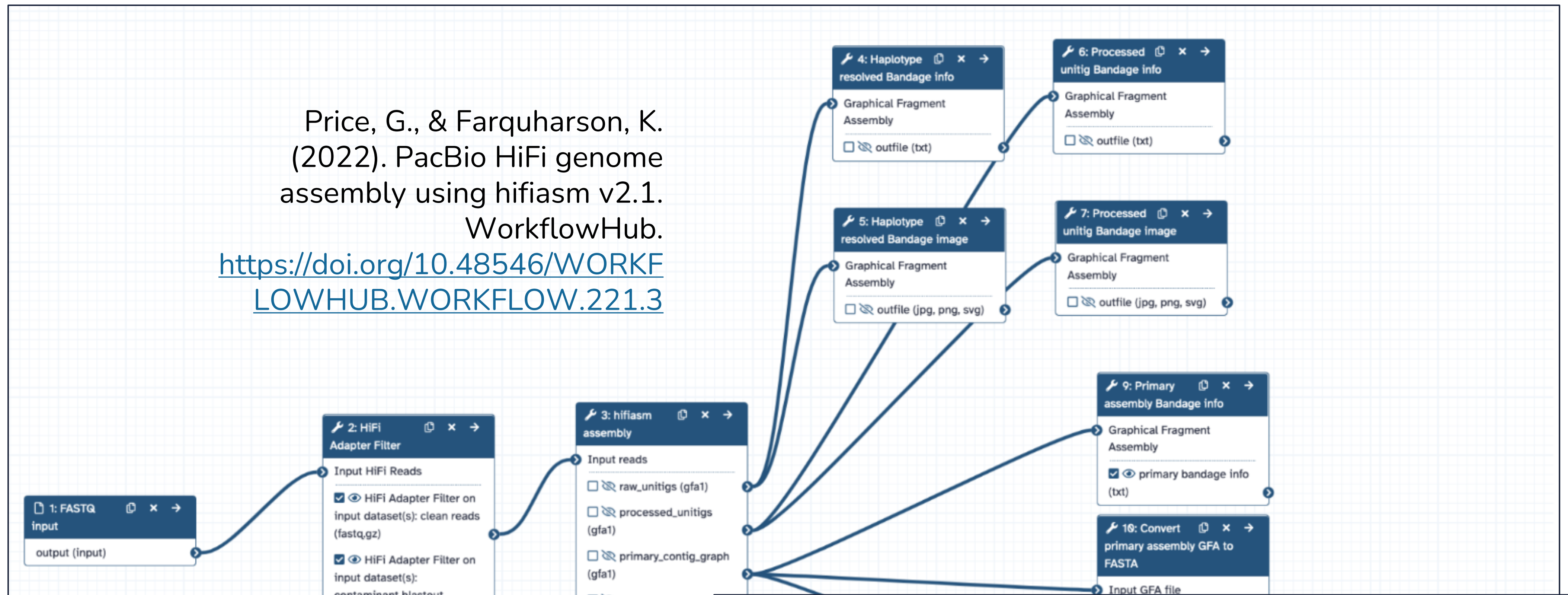
## Design / development

Workflow management system

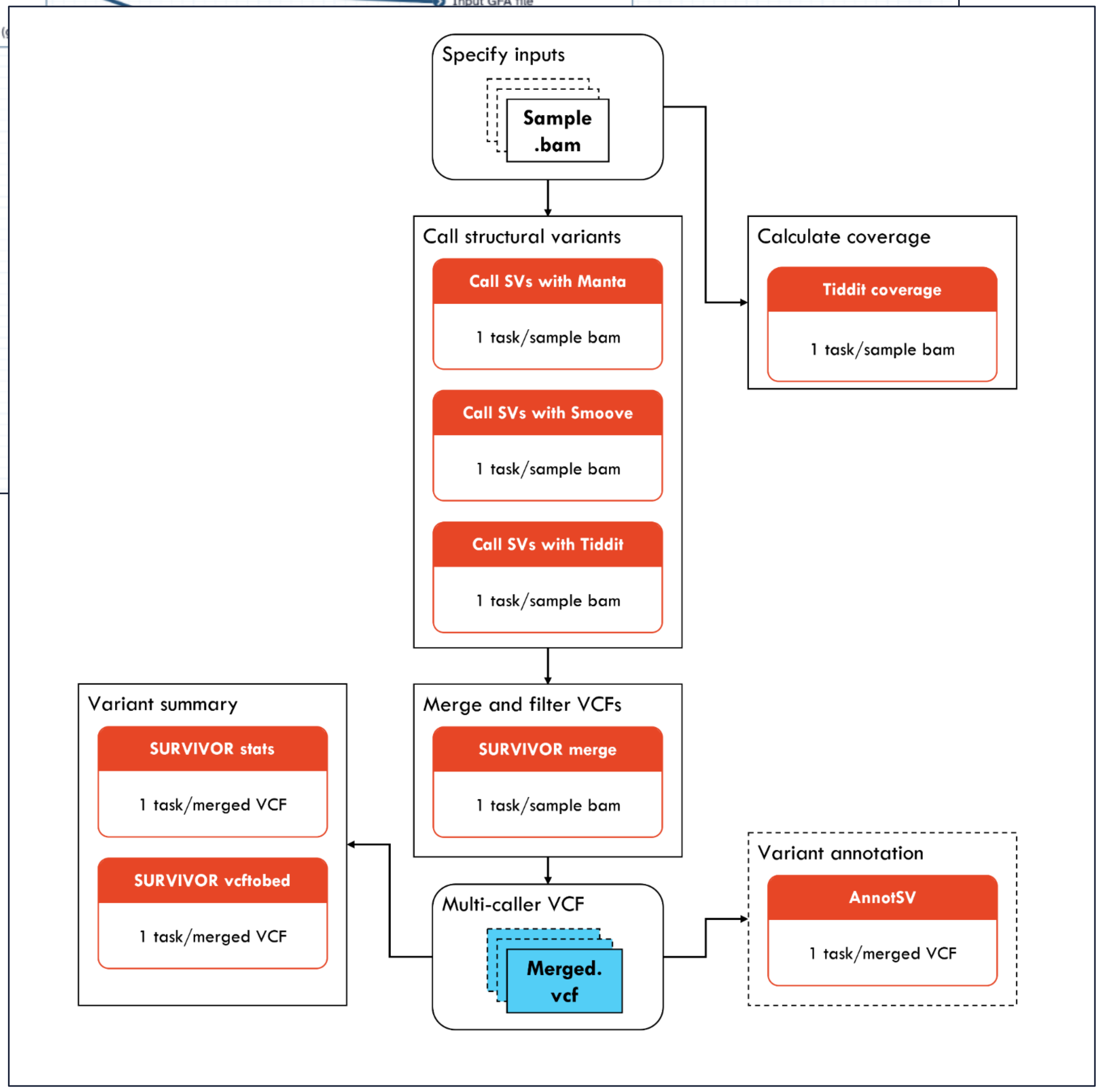
Workflow management system

Australian BioCommons

# Real world example: Galaxy / Nextflow



Price, G., & Farquharson, K. (2022). PacBio HiFi genome assembly using hifiasm v2.1. WorkflowHub. https://doi.org/10.48546/WORKFLOWHUB.WORKFLOW.221.3



Samaha, G., & Chew, T. (2023). *GermlineStructuralV-nf*. WorkflowHub. https://doi.org/10.48546/WORKFLOWHUB.WORKFLOW.431.1

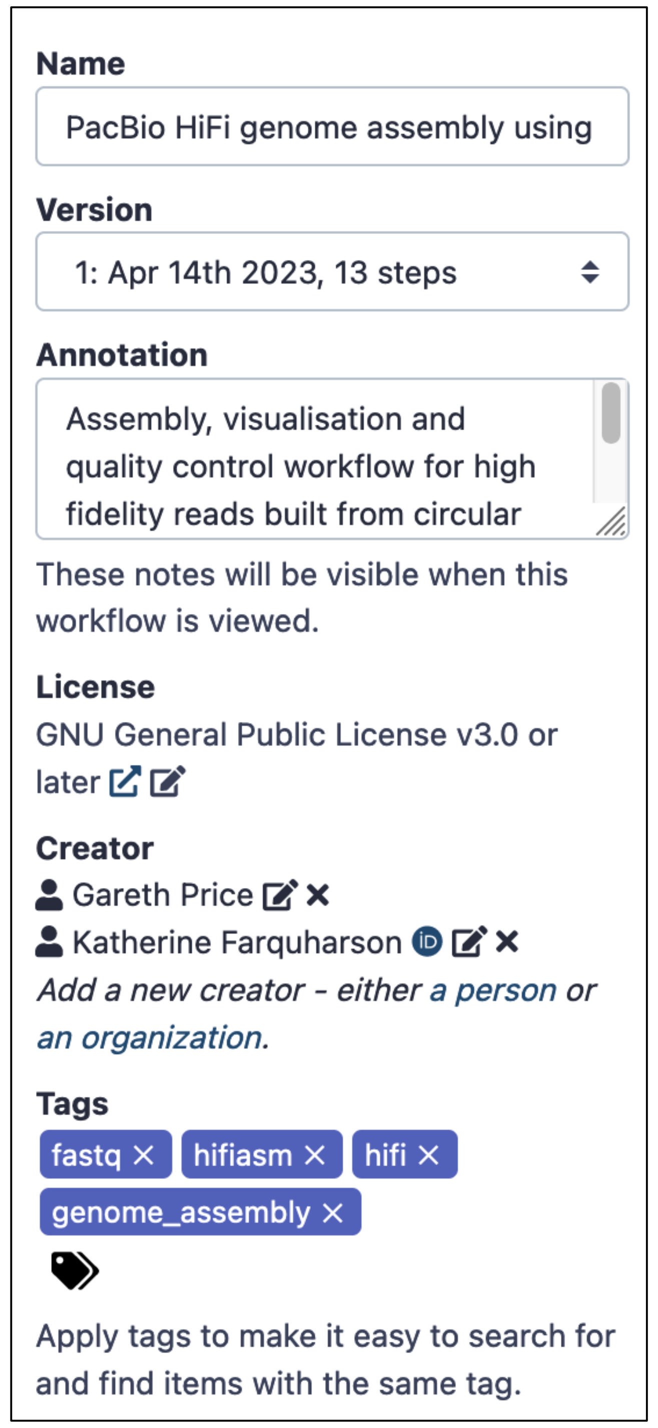
Galaxy workflow best practices

Nextflow & nf-core community guidelines & standards

## Design / development

Workflow management system

Workflow management system



**Name**

PacBio HiFi genome assembly using

**Version**

1: Apr 14th 2023, 13 steps

**Annotation**

Assembly, visualisation and quality control workflow for high fidelity reads built from circular

These notes will be visible when this workflow is viewed.

**License**

GNU General Public License v3.0 or later

**Creator**

Gareth Price

Katherine Farquharson

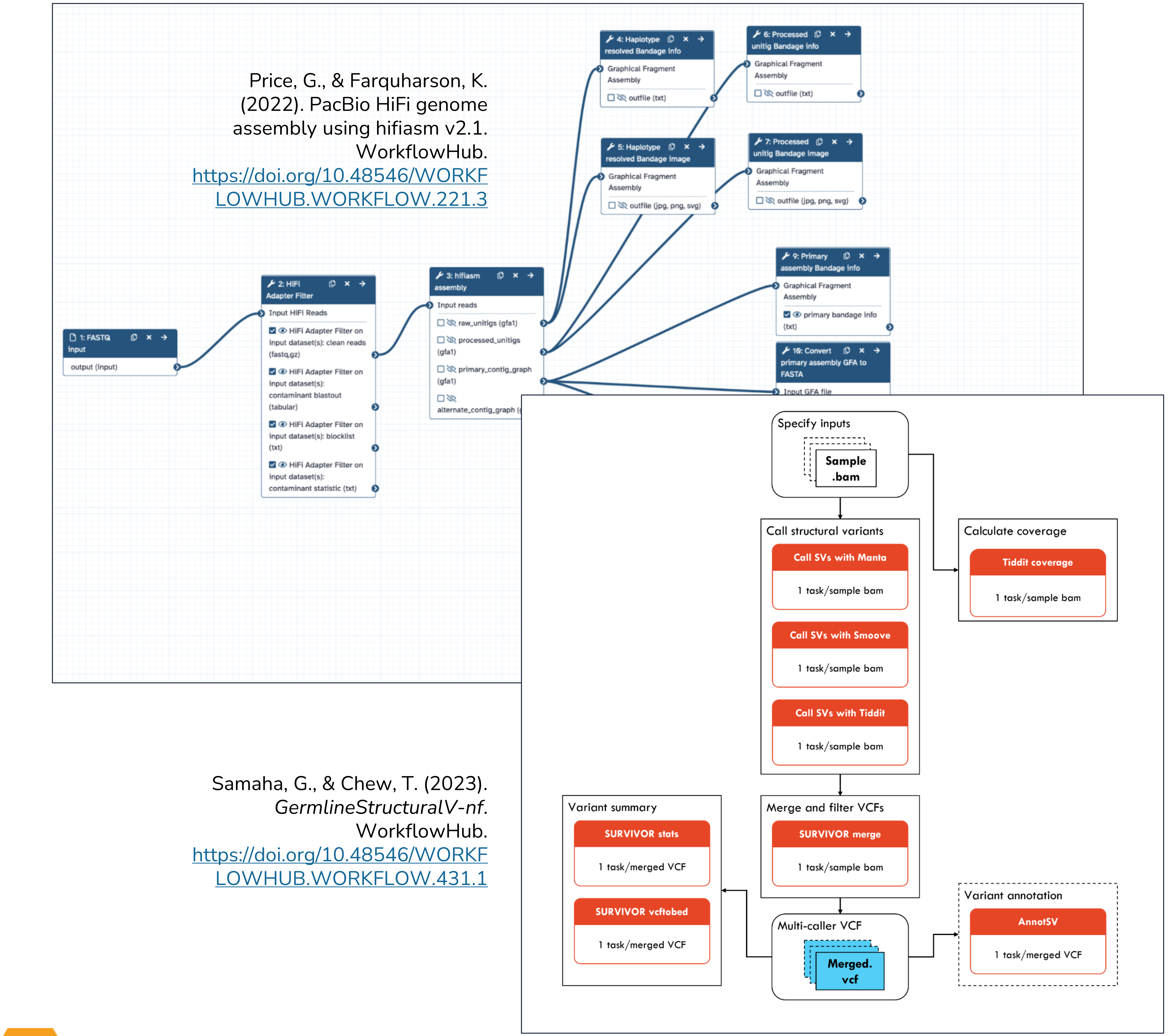
*Add a new creator - either a person or an organization.*

**Tags**

fastq ✕ hifiasm ✕ hifi ✕ genome_assembly ✕

Apply tags to make it easy to search for and find items with the same tag.

# Real world example: Galaxy / Nextflow

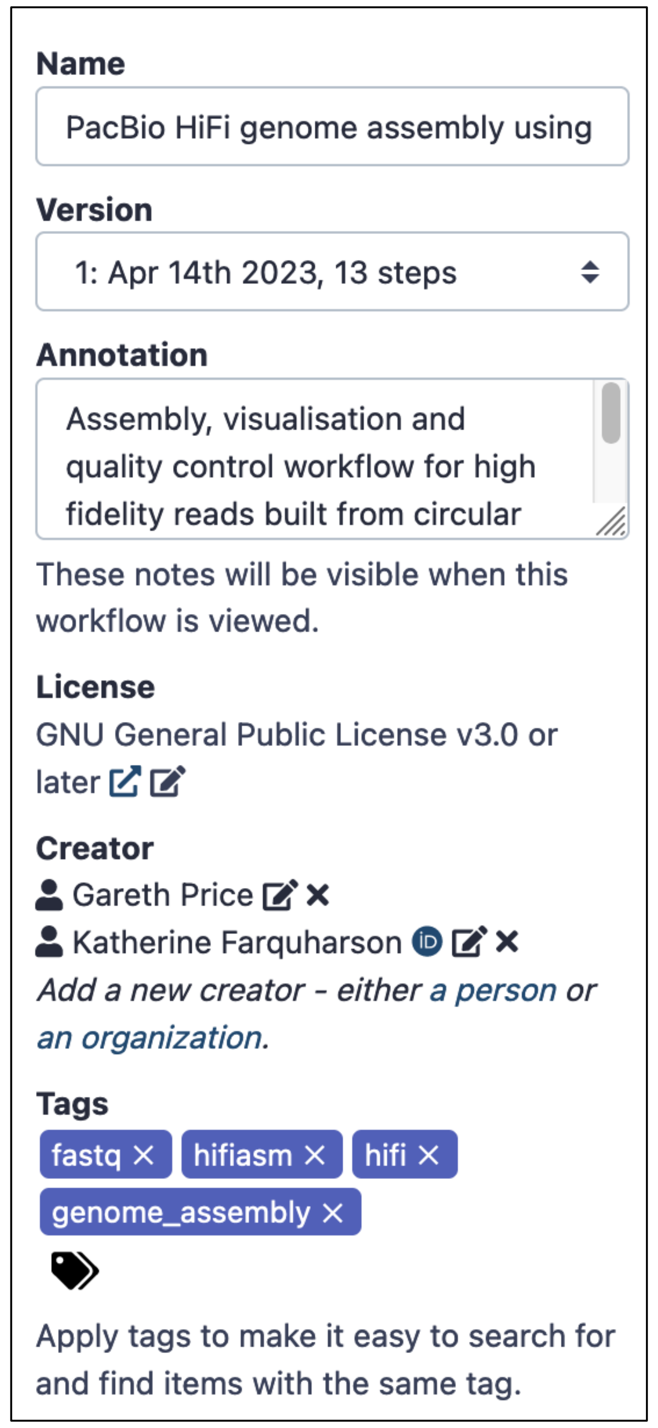


Price, G., & Farquharson, K. (2022). PacBio HiFi genome assembly using hifiasm v2.1. WorkflowHub. https://doi.org/10.48546/WORKFLOWHUB.WORKFLOW.221.3

Samaha, G., & Chew, T. (2023). *GermlineStructuralV-nf.* WorkflowHub. https://doi.org/10.48546/WORKFLOWHUB.WORKFLOW.431.1

Galaxy workflow best practices

Nextflow & nf-core community guidelines & standards



**Name**

PacBio HiFi genome assembly using

**Version**

1: Apr 14th 2023, 13 steps

**Annotation**

Assembly, visualisation and quality control workflow for high fidelity reads built from circular

These notes will be visible when this workflow is viewed.

**License**

GNU General Public License v3.0 or later

**Creator**

Gareth Price

Katherine Farquharson

*Add a new creator - either a person or an organization.*

**Tags**

fastq ✕ hifiasm ✕ hifi ✕ genome_assembly ✕

Apply tags to make it easy to search for and find items with the same tag.

## Design / development

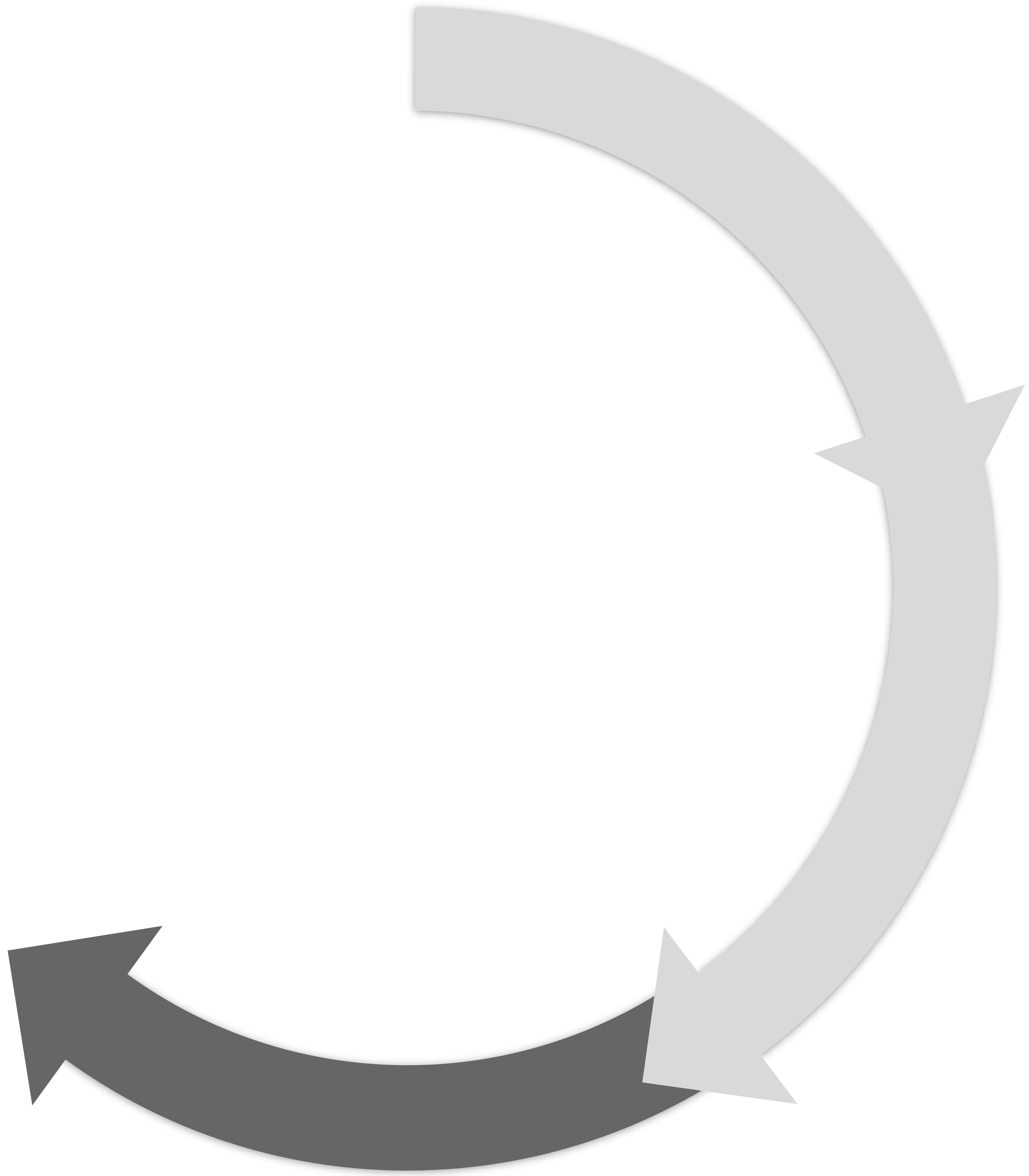
Workflow management system

Workflow management system

- ✓ Reproducibility
- ✓ Best practice
- ✓ Widely accepted
- ✓ Sustainable

Australian BioCommons

# Real world example: Galaxy / Nextflow



NCI AUSTRALIA

Pawsey

**Optimise**

Real world
test data

Galaxy AUSTRALIA

- ✓ Reproducibility
- ✓ Best practice
- ✓ Widely accepted
- ✓ Sustainable

# Real world example: Galaxy / Nextflow

Ross Crates, Brenton von Takach, Catherine M Young, et al. Genomic insights into a critically endangered island endemic songbird provide a roadmap for preventing extinction. *Authorea*. September 14, 2023.

NCI AUSTRALIA

**Optimise**

Real world test data

pawsey

Galaxy AUSTRALIA

- ✓ Reproducibility
- ✓ Best practice
- ✓ Widely accepted
- ✓ Sustainable

# Real world example: Galaxy / Nextflow

Ross Crates, Brenton von Takach, Catherine M Young, et al. Genomic insights into a critically endangered island endemic songbird provide a roadmap for preventing extinction. *Authorea*. September 14, 2023.

NCI AUSTRALIA

pawsey

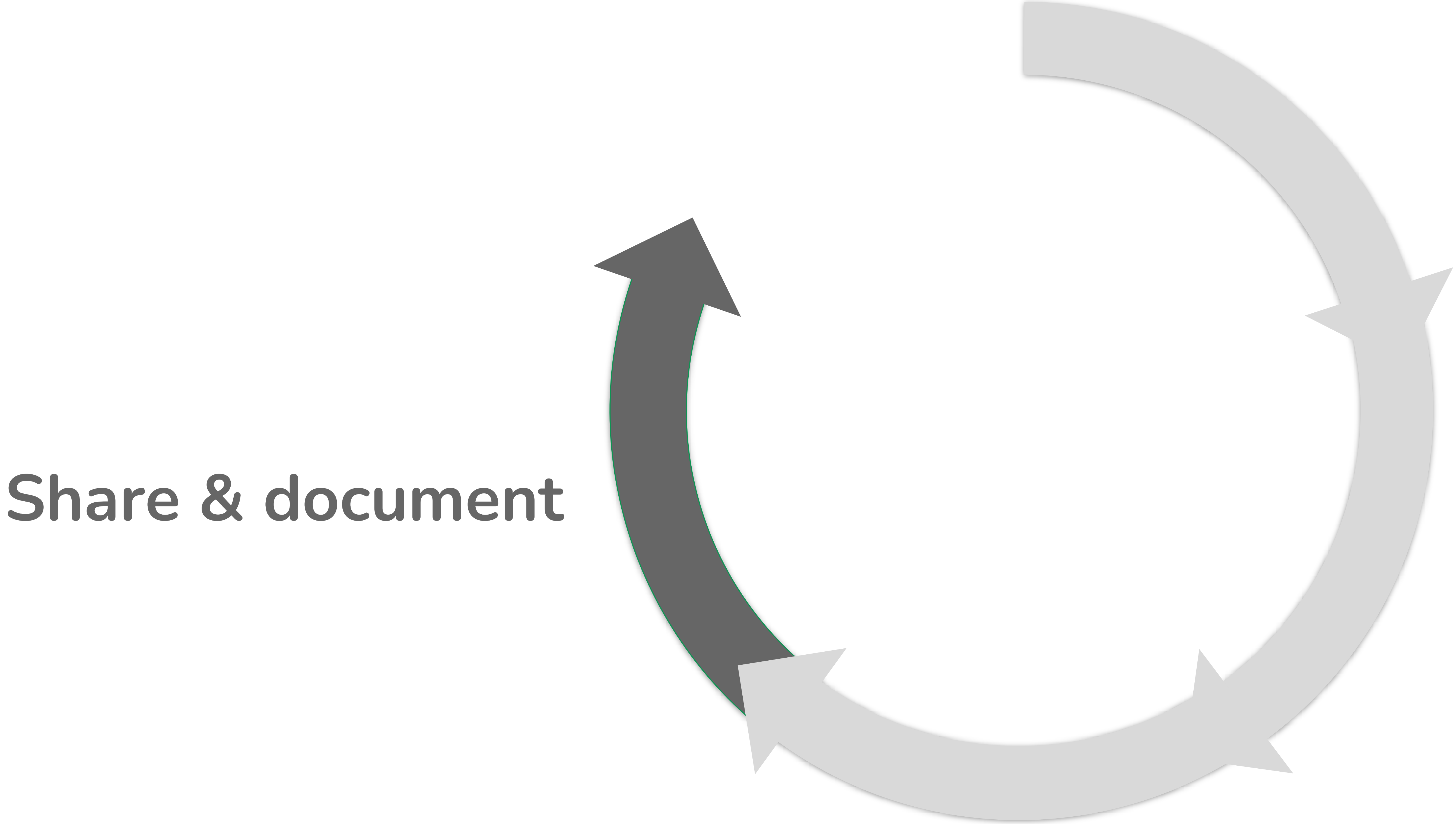
**Optimise**

Real world test data

Galaxy AUSTRALIA

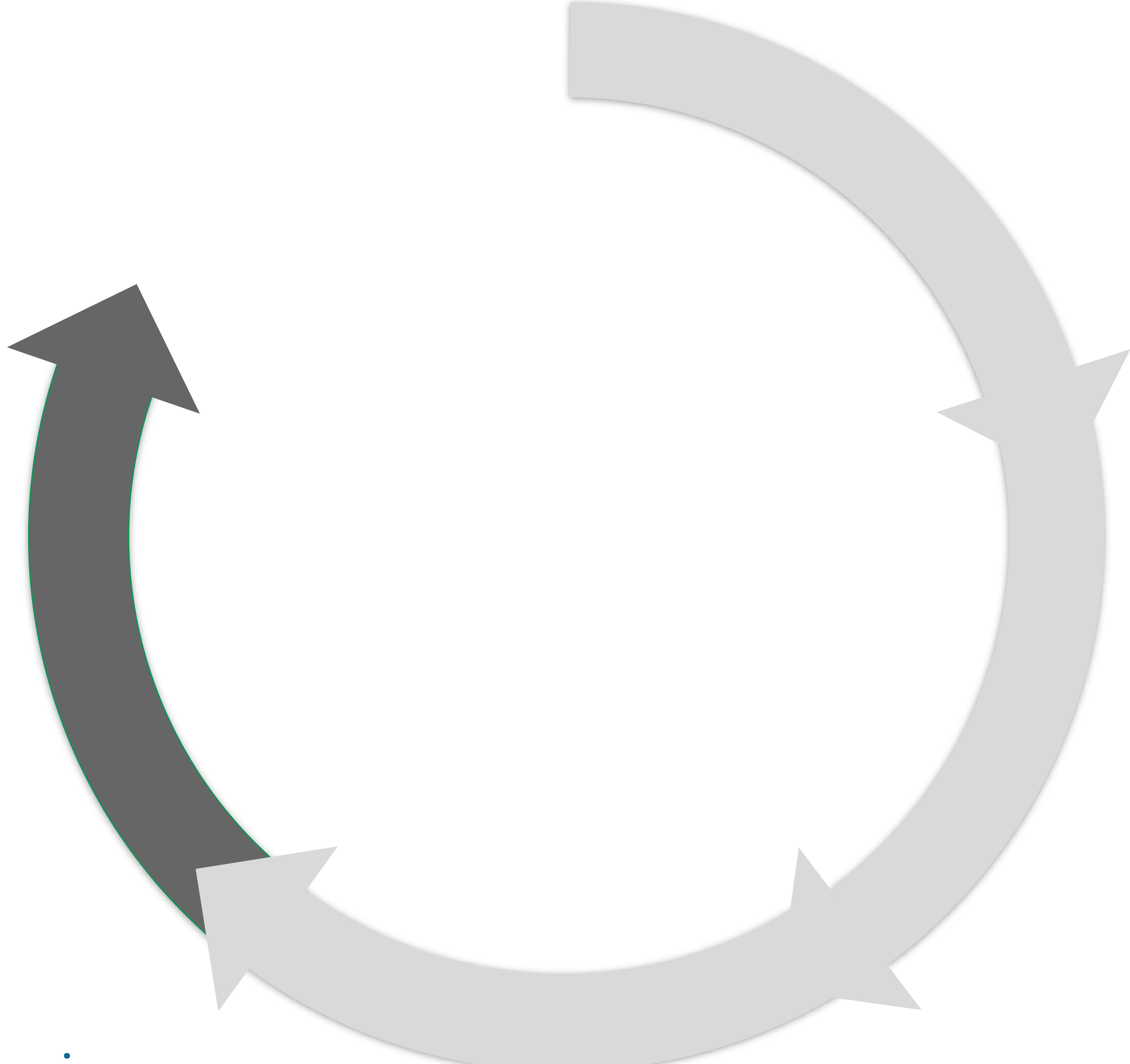
- ✓ Reusable
- ✓ Reproducibility
- ✓ ✓ Best practice
- ✓ Widely accepted
- ✓ Sustainable

# Real world example: Galaxy / Nextflow

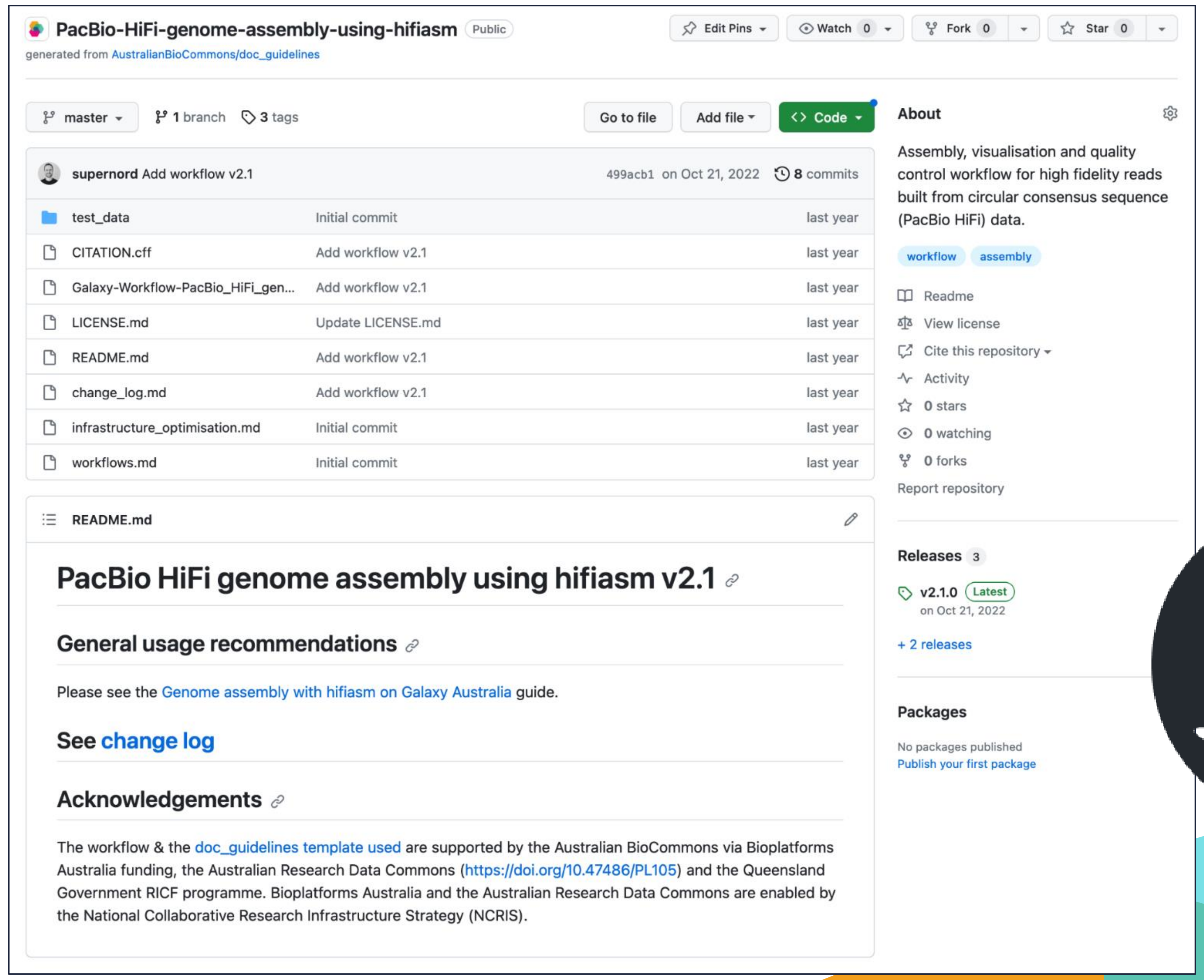


- ✓ Reusable
- ✓ Reproducibility
- ✓ ✓ Best practice
- ✓ Widely accepted
- ✓ Sustainable

# Real world example: Galaxy / Nextflow



Share & document



**PacBio-HiFi-genome-assembly-using-hifiasm** Public

generated from AustralianBioCommons/doc_guidelines

**supernord** Add workflow v2.1 499acb1 on Oct 21, 2022 8 commits

| | | |
|---|---|---|
| test_data | Initial commit | last year |
| CITATION.cff | Add workflow v2.1 | last year |
| Galaxy-Workflow-PacBio_HiFi_gen... | Add workflow v2.1 | last year |
| LICENSE.md | Update LICENSE.md | last year |
| README.md | Add workflow v2.1 | last year |
| change_log.md | Add workflow v2.1 | last year |
| infrastructure_optimisation.md | Initial commit | last year |
| workflows.md | Initial commit | last year |

README.md

## PacBio HiFi genome assembly using hifiasm v2.1

### General usage recommendations

Please see the Genome assembly with hifiasm on Galaxy Australia guide.

### See change log

### Acknowledgements

The workflow & the doc_guidelines template used are supported by the Australian BioCommons via Bioplatforms Australia funding, the Australian Research Data Commons (https://doi.org/10.47486/PL105) and the Queensland Government RICF programme. Bioplatforms Australia and the Australian Research Data Commons are enabled by the National Collaborative Research Infrastructure Strategy (NCRIS).

**About**

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

**Releases** 3

v2.1.0 Latest
on Oct 21, 2022

+ 2 releases

**Packages**

No packages published
Publish your first package

Version
License

✓ ✓ Reusable

✓ Reproducibility

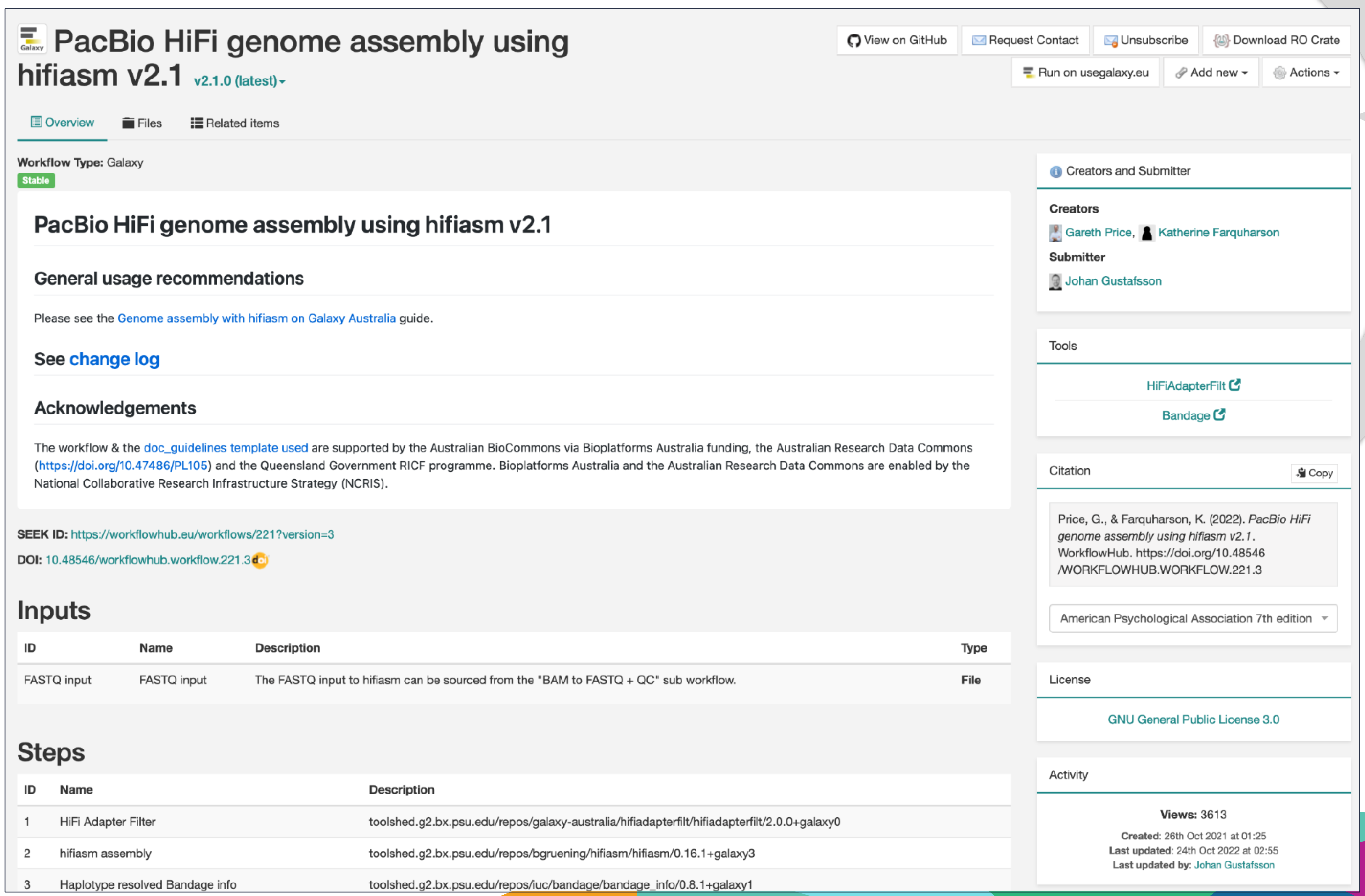
✓ ✓ ✓ Best practice

✓ Widely accepted

✓ Sustainable

# Real world example: Galaxy / Nextflow

WorkflowHub



## PacBio HiFi genome assembly using hifiasm v2.1 v2.1.0 (latest)

**Workflow Type:** Galaxy

Stable

### PacBio HiFi genome assembly using hifiasm v2.1

#### General usage recommendations

Please see the Genome assembly with hifiasm on Galaxy Australia guide.

#### See change log

#### Acknowledgements

The workflow & the doc_guidelines template used are supported by the Australian BioCommons via Bioplatforms Australia funding, the Australian Research Data Commons (https://doi.org/10.47486/PL105) and the Queensland Government RICF programme. Bioplatforms Australia and the Australian Research Data Commons are enabled by the National Collaborative Research Infrastructure Strategy (NCRIS).

**SEEK ID:** https://workflowhub.eu/workflows/221?version=3

**DOI:** 10.48546/workflowhub.workflow.221.3

### 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 |

Creators and Submitter

**Creators**

Gareth Price, Katherine Farquharson

**Submitter**

Johan Gustafsson

Tools

HiFiAdapterFilt

Bandage

Citation

Price, G., & Farquharson, K. (2022). *PacBio HiFi genome assembly using hifiasm v2.1*. WorkflowHub. https://doi.org/10.48546/WORKFLOWHUB.WORKFLOW.221.3

American Psychological Association 7th edition

License

GNU General Public License 3.0

Activity

**Views:** 3613

**Created:** 26th Oct 2021 at 01:25
**Last updated:** 24th Oct 2022 at 02:55
**Last updated by:** Johan Gustafsson

✓ Visible

✓ ✓ Reusable

✓ Reproducibility

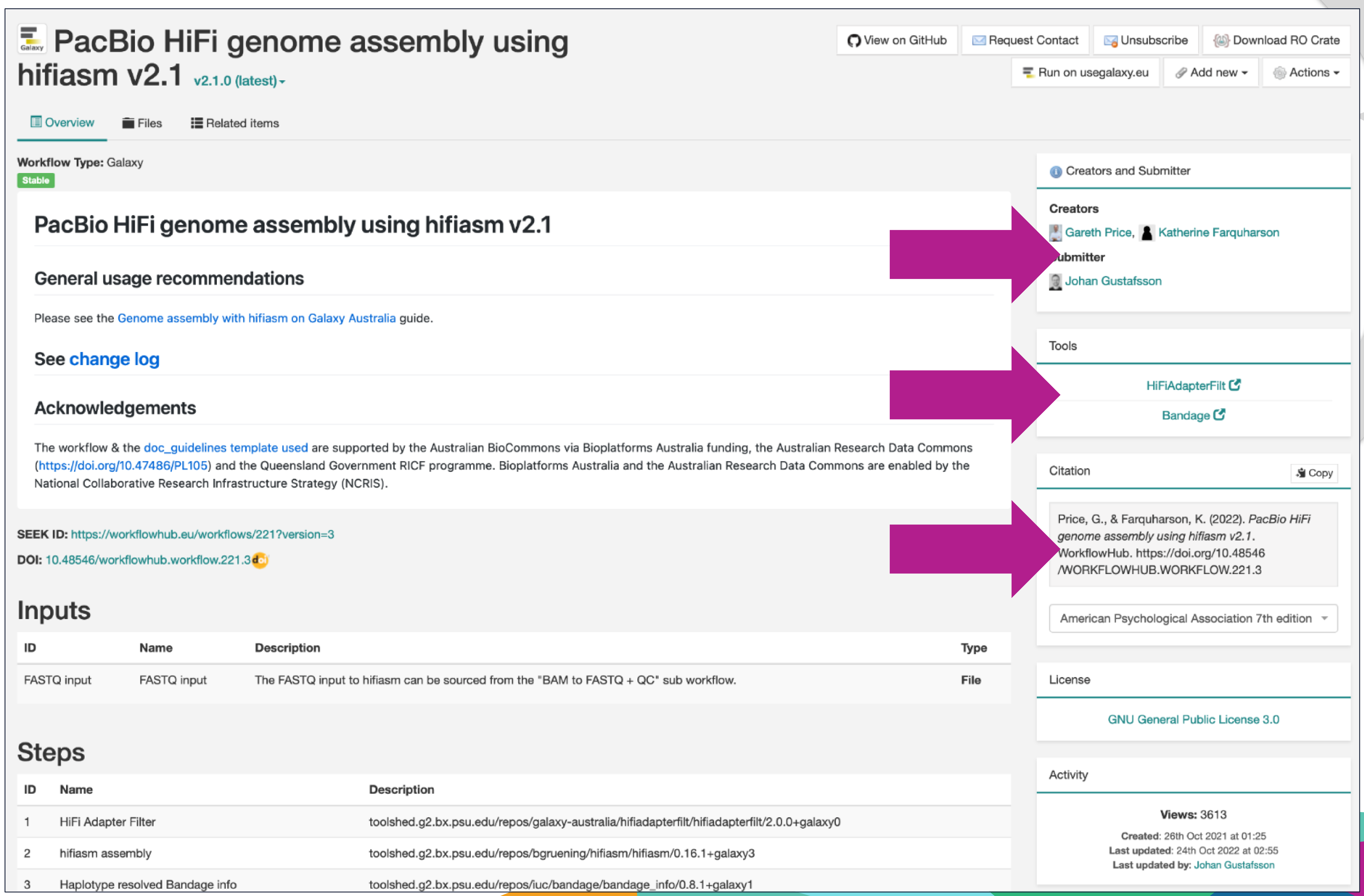
✓ ✓ ✓ ✓ Best practice

✓ Widely accepted

✓ Sustainable

# Real world example: Galaxy / Nextflow

WorkflowHub



## PacBio HiFi genome assembly using hifiasm v2.1 v2.1.0 (latest)

**Workflow Type:** Galaxy
Stable

### PacBio HiFi genome assembly using hifiasm v2.1

#### General usage recommendations

Please see the Genome assembly with hifiasm on Galaxy Australia guide.

#### See change log

#### Acknowledgements

The workflow & the doc_guidelines template used are supported by the Australian BioCommons via Bioplatforms Australia funding, the Australian Research Data Commons (https://doi.org/10.47486/PL105) and the Queensland Government RICF programme. Bioplatforms Australia and the Australian Research Data Commons are enabled by the National Collaborative Research Infrastructure Strategy (NCRIS).

**SEEK ID:** https://workflowhub.eu/workflows/221?version=3

**DOI:** 10.48546/workflowhub.workflow.221.3

### 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 |

Creators and Submitter

**Creators**
Gareth Price, Katherine Farquharson
**ubmitter**
Johan Gustafsson

Tools

HiFiAdapterFilt
Bandage

Citation

Price, G., & Farquharson, K. (2022). *PacBio HiFi genome assembly using hifiasm v2.1*. WorkflowHub. https://doi.org/10.48546/WORKFLOWHUB.WORKFLOW.221.3

American Psychological Association 7th edition

License

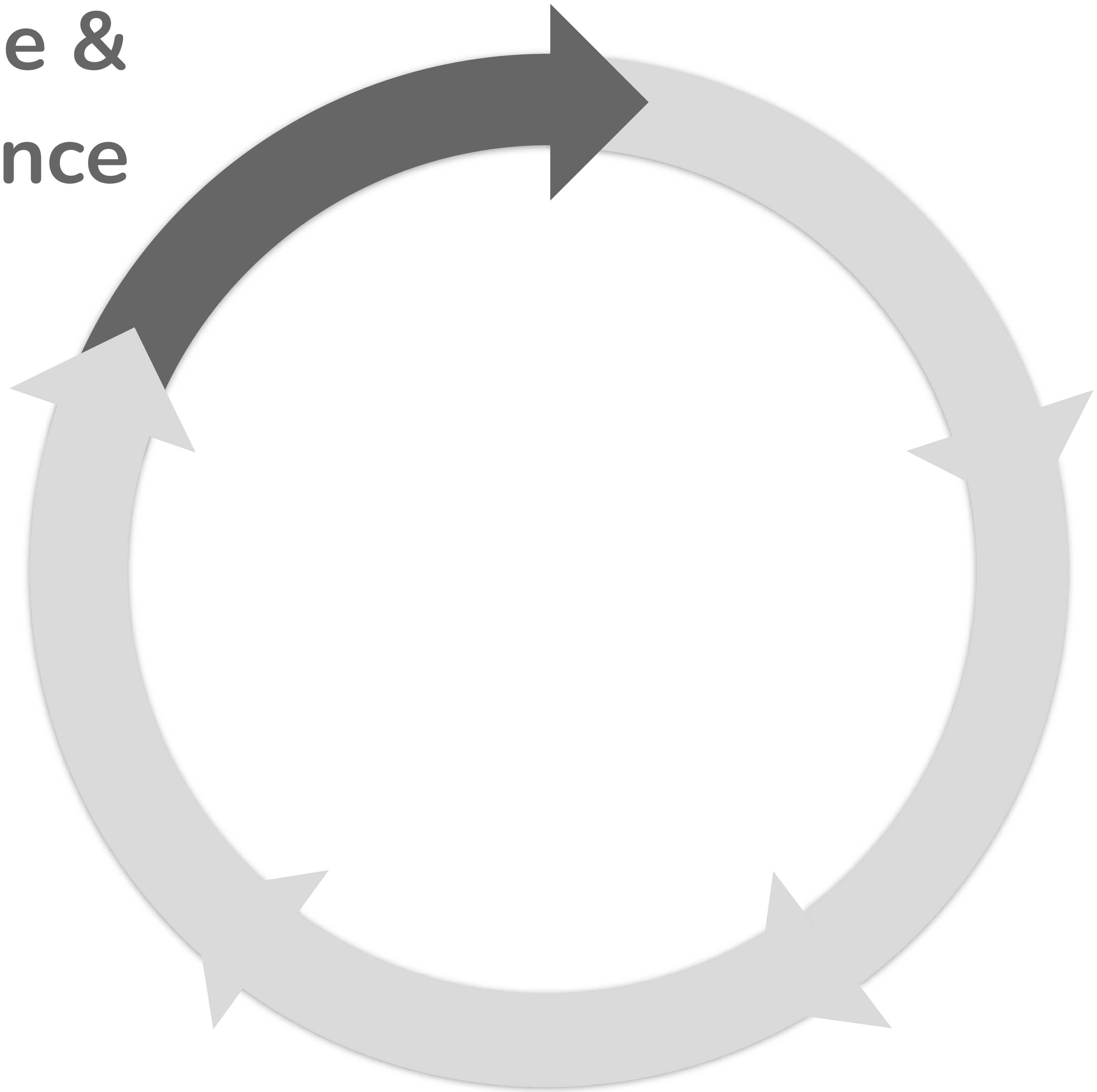
GNU General Public License 3.0

Activity

**Views:** 3613
**Created:** 26th Oct 2021 at 01:25
**Last updated:** 24th Oct 2022 at 02:55
**Last updated by:** Johan Gustafsson

✓ ✓ Visible
✓ ✓ Reusable
✓ Reproducibility
✓ ✓ ✓ ✓ Best practice
✓ Widely accepted
✓ Sustainable

# Real world example: Galaxy / Nextflow



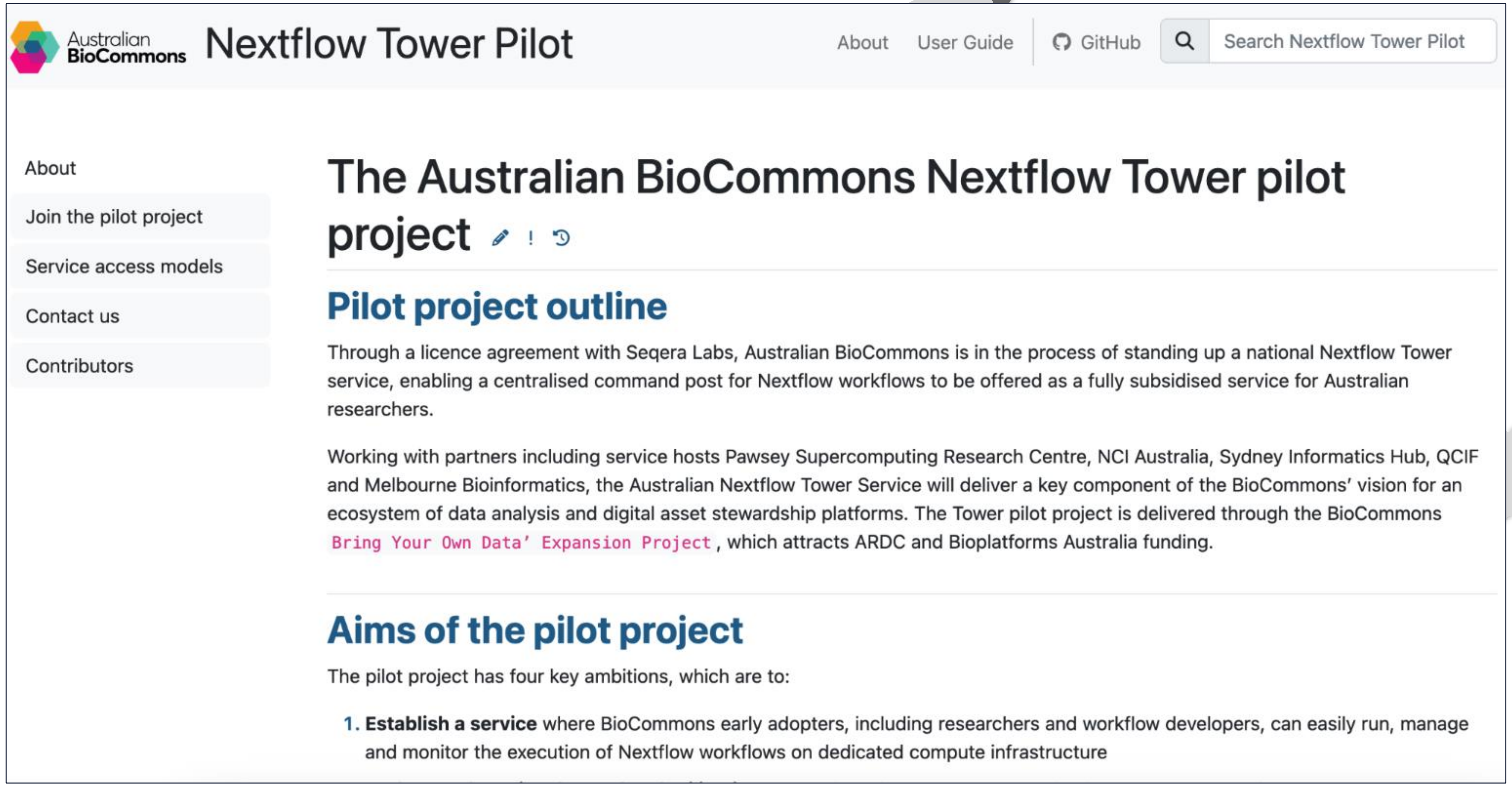
Support reuse & maintenance

| | | | | |
|---|---|---|---|---|
| | | ✓ | ✓ | Visible |
| | | ✓ | ✓ | Reusable |
| | | | ✓ | Reproducibility |
| ✓ | ✓ | ✓ | ✓ | Best practice |
| | | | ✓ | Widely accepted |
| | | | ✓ | Sustainable |

# Real world example: Galaxy / Nextflow

Support reuse &
maintenance

Nextflow Tower



## The Australian BioCommons Nextflow Tower pilot project

### Pilot project outline

Through a licence agreement with Seqera Labs, Australian BioCommons is in the process of standing up a national Nextflow Tower service, enabling a centralised command post for Nextflow workflows to be offered as a fully subsidised service for Australian researchers.

Working with partners including service hosts Pawsey Supercomputing Research Centre, NCI Australia, Sydney Informatics Hub, QCIF and Melbourne Bioinformatics, the Australian Nextflow Tower Service will deliver a key component of the BioCommons' vision for an ecosystem of data analysis and digital asset stewardship platforms. The Tower pilot project is delivered through the BioCommons `Bring Your Own Data' Expansion Project`, which attracts ARDC and Bioplatforms Australia funding.

### Aims of the pilot project

The pilot project has four key ambitions, which are to:

1. **Establish a service** where BioCommons early adopters, including researchers and workflow developers, can easily run, manage and monitor the execution of Nextflow workflows on dedicated compute infrastructure

✓ ✓ Visible

✓ ✓ Reusable

✓ Reproducibility

✓ ✓ ✓ ✓ Best practice

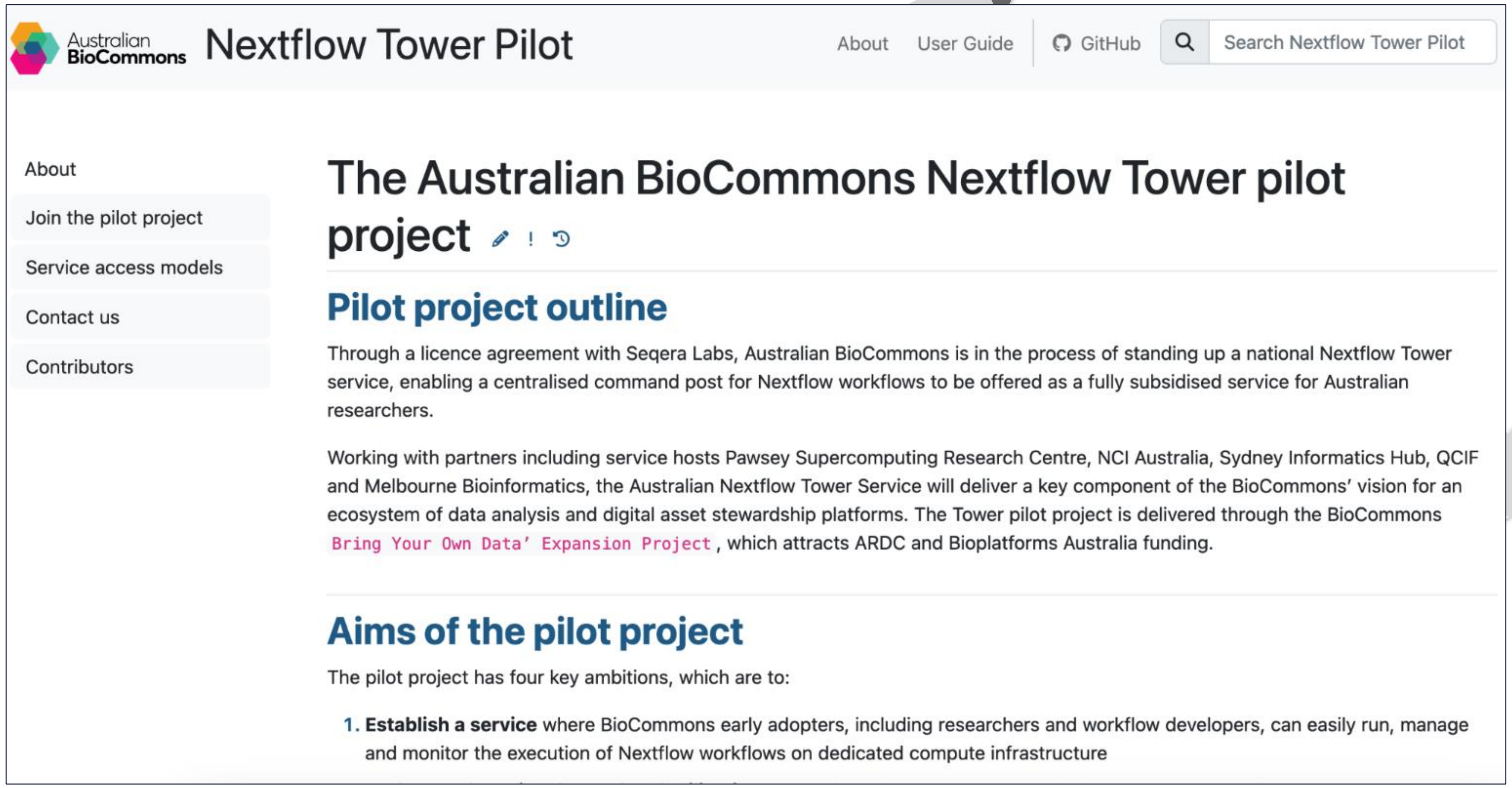
✓ Widely accepted

✓ Sustainable

# Real world example: Galaxy / Nextflow

Support reuse & maintenance

Nextflow Tower



## The Australian BioCommons Nextflow Tower pilot project

### Pilot project outline

Through a licence agreement with Seqera Labs, Australian BioCommons is in the process of standing up a national Nextflow Tower service, enabling a centralised command post for Nextflow workflows to be offered as a fully subsidised service for Australian researchers.

Working with partners including service hosts Pawsey Supercomputing Research Centre, NCI Australia, Sydney Informatics Hub, QCIF and Melbourne Bioinformatics, the Australian Nextflow Tower Service will deliver a key component of the BioCommons' vision for an ecosystem of data analysis and digital asset stewardship platforms. The Tower pilot project is delivered through the BioCommons `Bring Your Own Data’ Expansion Project`, which attracts ARDC and Bioplatforms Australia funding.

### Aims of the pilot project

The pilot project has four key ambitions, which are to:

1. **Establish a service** where BioCommons early adopters, including researchers and workflow developers, can easily run, manage and monitor the execution of Nextflow workflows on dedicated compute infrastructure

✓ ✓ Visible

✓ ✓ ✓ Reusable

✓ Reproducibility

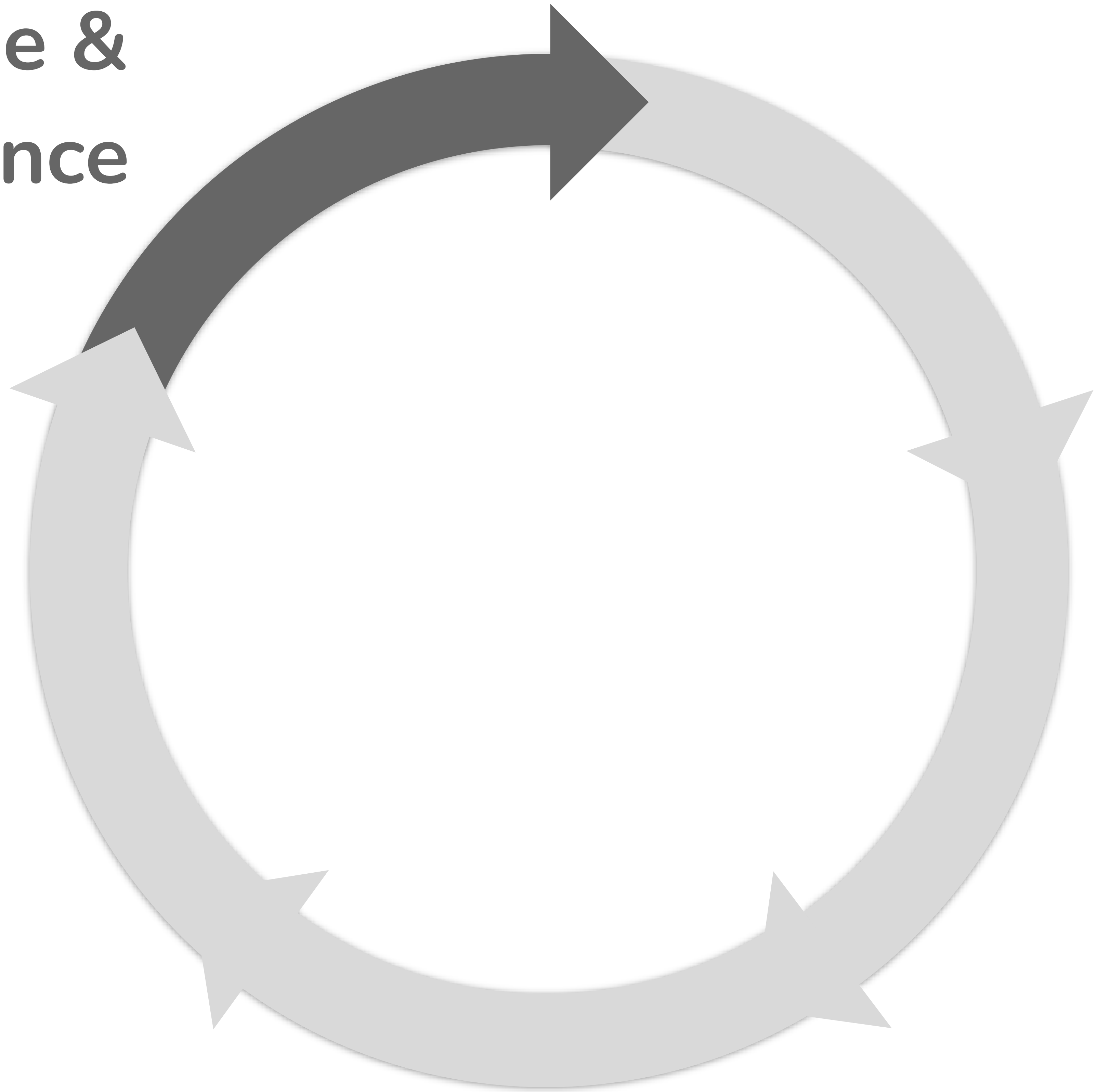
✓ ✓ ✓ ✓ Best practice

✓ Widely accepted

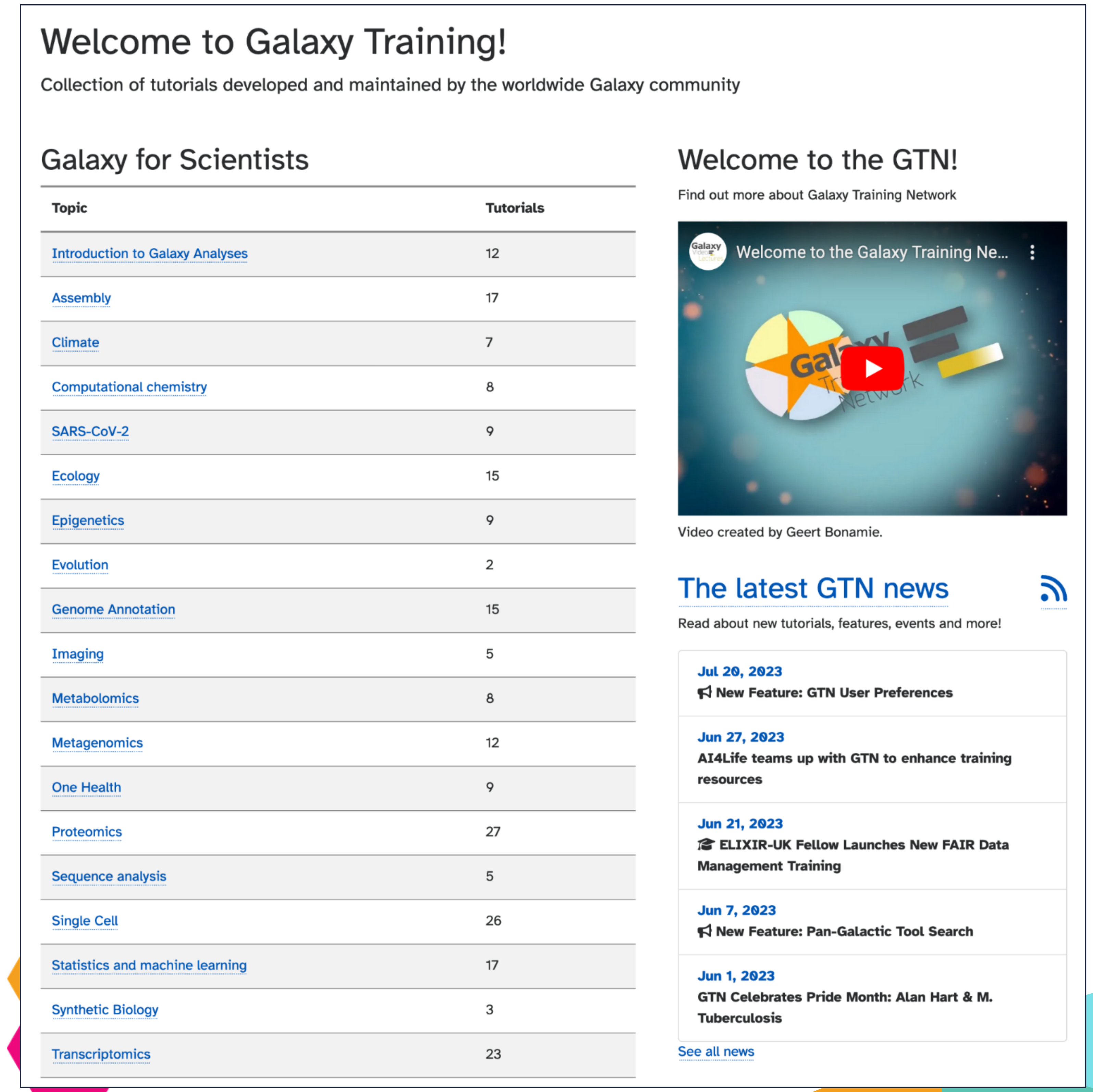
✓ ✓ Sustainable

Australian BioCommons

# Real world example: Galaxy / Nextflow


Support reuse & maintenance

Galaxy Training Network


## Welcome to Galaxy Training!

Collection of tutorials developed and maintained by the worldwide Galaxy community

### Galaxy for Scientists

| Topic | Tutorials |
|---|---|
| Introduction to Galaxy Analyses | 12 |
| Assembly | 17 |
| Climate | 7 |
| Computational chemistry | 8 |
| SARS-CoV-2 | 9 |
| Ecology | 15 |
| Epigenetics | 9 |
| Evolution | 2 |
| Genome Annotation | 15 |
| Imaging | 5 |
| Metabolomics | 8 |
| Metagenomics | 12 |
| One Health | 9 |
| Proteomics | 27 |
| Sequence analysis | 5 |
| Single Cell | 26 |
| Statistics and machine learning | 17 |
| Synthetic Biology | 3 |
| Transcriptomics | 23 |

### Welcome to the GTN!

Find out more about Galaxy Training Network

Video created by Geert Bonamie.

### The latest GTN news

Read about new tutorials, features, events and more!

**Jul 20, 2023**
New Feature: GTN User Preferences

**Jun 27, 2023**
AI4Life teams up with GTN to enhance training resources

**Jun 21, 2023**
ELIXIR-UK Fellow Launches New FAIR Data Management Training

**Jun 7, 2023**
New Feature: Pan-Galactic Tool Search

**Jun 1, 2023**
GTN Celebrates Pride Month: Alan Hart & M. Tuberculosis

See all news

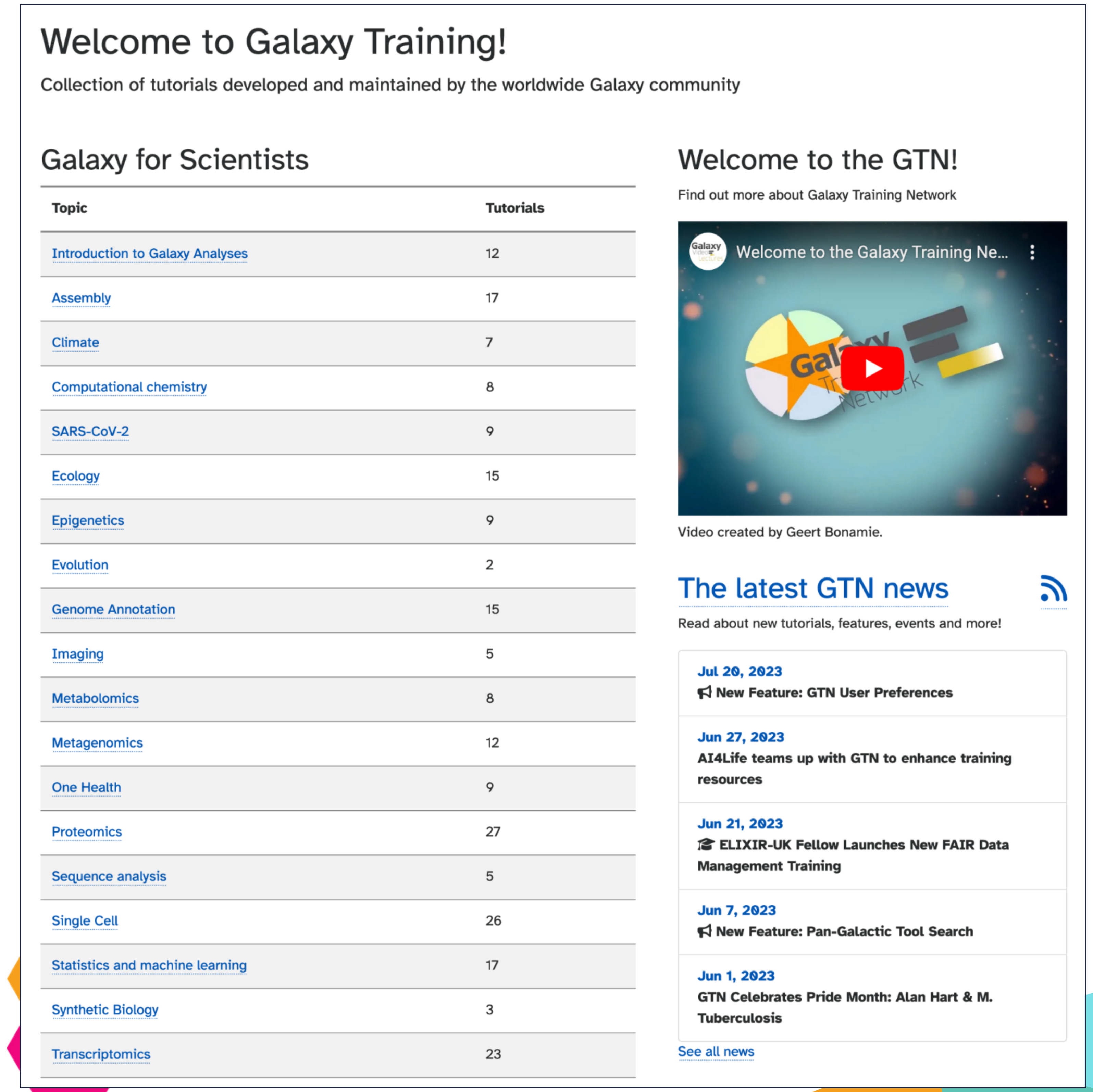
✓ ✓ Visible
✓ ✓ ✓ ✓ Reusable
✓ Reproducibility
✓ ✓ ✓ ✓ Best practice
✓ Widely accepted
✓ ✓ ✓ Sustainable

# Real world example: Galaxy / Nextflow

Support reuse & maintenance

How-to Guides

Galaxy Training Network



## Welcome to Galaxy Training!

Collection of tutorials developed and maintained by the worldwide Galaxy community

### Galaxy for Scientists

| Topic | Tutorials |
|---|---|
| Introduction to Galaxy Analyses | 12 |
| Assembly | 17 |
| Climate | 7 |
| Computational chemistry | 8 |
| SARS-CoV-2 | 9 |
| Ecology | 15 |
| Epigenetics | 9 |
| Evolution | 2 |
| Genome Annotation | 15 |
| Imaging | 5 |
| Metabolomics | 8 |
| Metagenomics | 12 |
| One Health | 9 |
| Proteomics | 27 |
| Sequence analysis | 5 |
| Single Cell | 26 |
| Statistics and machine learning | 17 |
| Synthetic Biology | 3 |
| Transcriptomics | 23 |

### Welcome to the GTN!

Find out more about Galaxy Training Network

Video created by Geert Bonamie.

### The latest GTN news

Read about new tutorials, features, events and more!

**Jul 20, 2023**
**New Feature: GTN User Preferences**

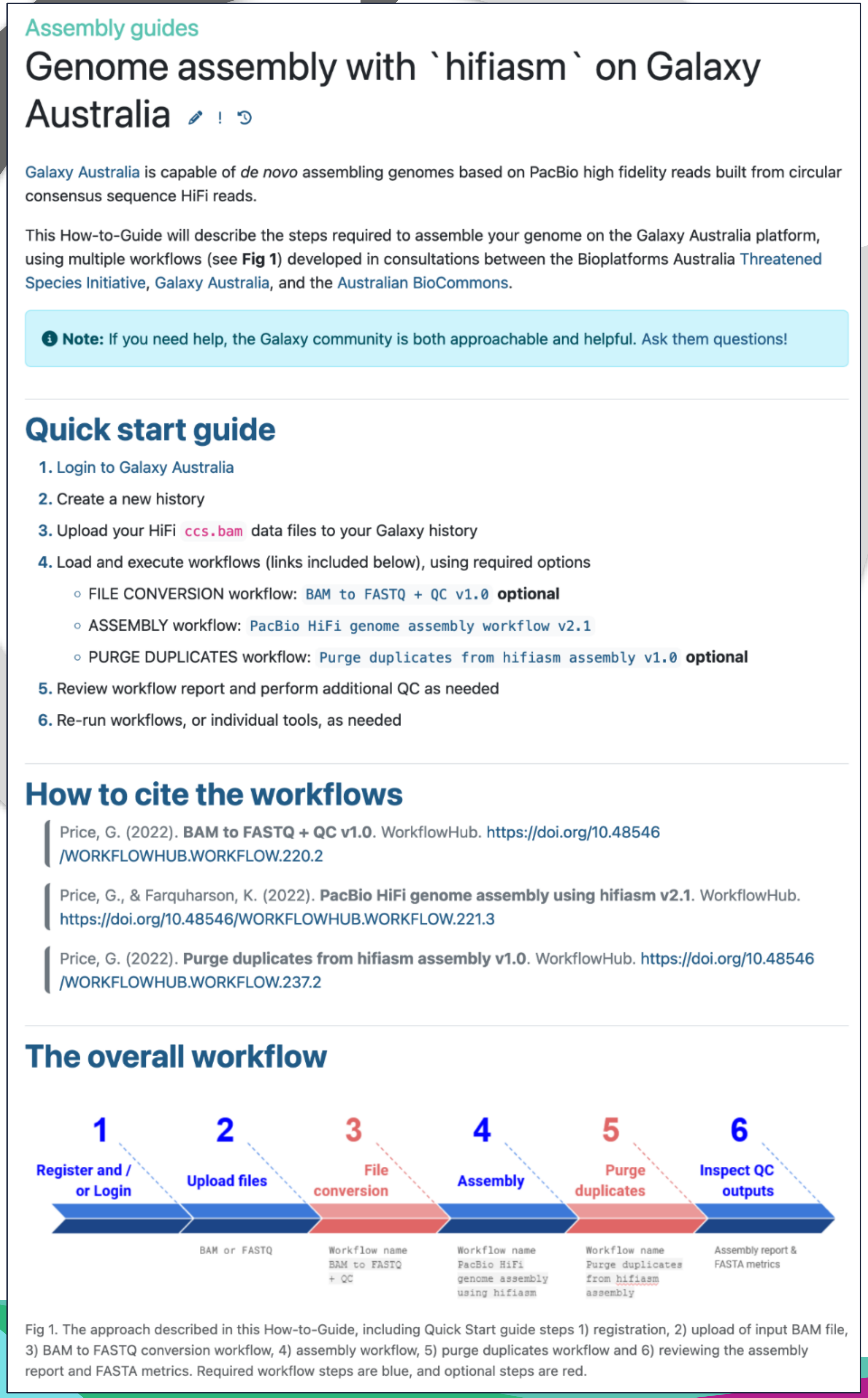
**Jun 27, 2023**
**AI4Life teams up with GTN to enhance training resources**

**Jun 21, 2023**
**ELIXIR-UK Fellow Launches New FAIR Data Management Training**

**Jun 7, 2023**
**New Feature: Pan-Galactic Tool Search**

**Jun 1, 2023**
**GTN Celebrates Pride Month: Alan Hart & M. Tuberculosis**

See all news



Assembly guides

## Genome assembly with `hifiasm` on Galaxy Australia

Galaxy Australia is capable of *de novo* assembling genomes based on PacBio high fidelity reads built from circular consensus sequence HiFi reads.

This How-to-Guide will describe the steps required to assemble your genome on the Galaxy Australia platform, using multiple workflows (see **Fig 1**) developed in consultations between the Bioplatforms Australia Threatened Species Initiative, Galaxy Australia, and the Australian BioCommons.

**Note:** If you need help, the Galaxy community is both approachable and helpful. Ask them questions!

### Quick start guide

1. Login to Galaxy Australia
2. Create a new history
3. Upload your HiFi `ccs.bam` data files to your Galaxy history
4. Load and execute workflows (links included below), using required options
    - FILE CONVERSION workflow: `BAM to FASTQ + QC v1.0` **optional**
    - ASSEMBLY workflow: `PacBio HiFi genome assembly workflow v2.1`
    - PURGE DUPLICATES workflow: `Purge duplicates from hifiasm assembly v1.0` **optional**
5. Review workflow report and perform additional QC as needed
6. Re-run workflows, or individual tools, as needed

### How to cite the workflows

> Price, G. (2022). **BAM to FASTQ + QC v1.0**. WorkflowHub. https://doi.org/10.48546/WORKFLOWHUB.WORKFLOW.220.2

> Price, G., & Farquharson, K. (2022). **PacBio HiFi genome assembly using hifiasm v2.1**. WorkflowHub. https://doi.org/10.48546/WORKFLOWHUB.WORKFLOW.221.3

> Price, G. (2022). **Purge duplicates from hifiasm assembly v1.0**. WorkflowHub. https://doi.org/10.48546/WORKFLOWHUB.WORKFLOW.237.2

### The overall workflow

Fig 1. The approach described in this How-to-Guide, including Quick Start guide steps 1) registration, 2) upload of input BAM file, 3) BAM to FASTQ conversion workflow, 4) assembly workflow, 5) purge duplicates workflow and 6) reviewing the assembly report and FASTA metrics. Required workflow steps are blue, and optional steps are red.

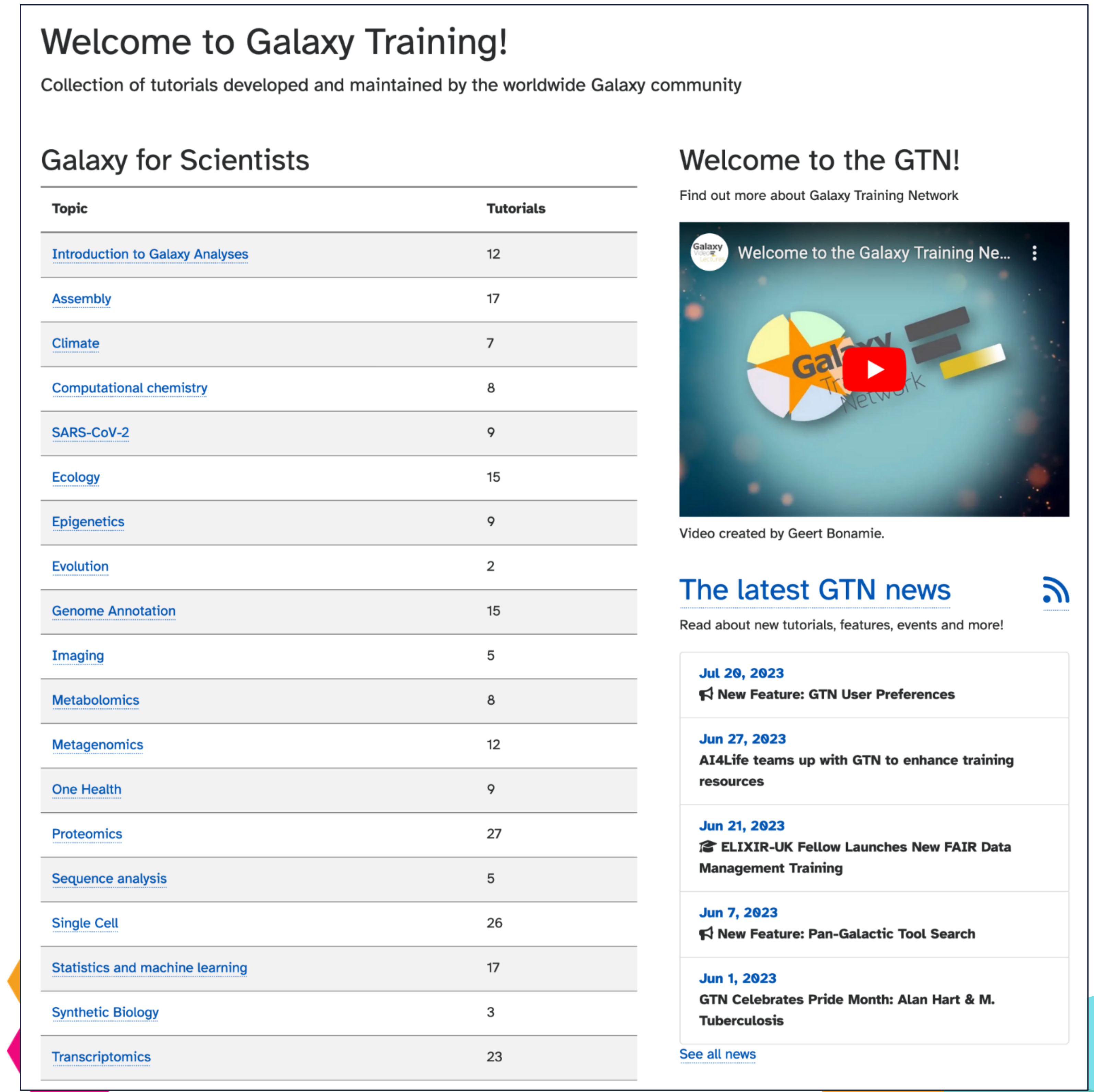
✓ ✓ Visible
✓ ✓ ✓ ✓ Reusable
✓ Reproducibility
✓ ✓ ✓ ✓ Best practice
✓ Widely accepted
✓ ✓ ✓ Sustainable

# Real world example: Galaxy / Nextflow

Training

## Support reuse & maintenance

## Galaxy Training Network



# Welcome to Galaxy Training!

Collection of tutorials developed and maintained by the worldwide Galaxy community

## Galaxy for Scientists

| Topic | Tutorials |
|---|---|
| Introduction to Galaxy Analyses | 12 |
| Assembly | 17 |
| Climate | 7 |
| Computational chemistry | 8 |
| SARS-CoV-2 | 9 |
| Ecology | 15 |
| Epigenetics | 9 |
| Evolution | 2 |
| Genome Annotation | 15 |
| Imaging | 5 |
| Metabolomics | 8 |
| Metagenomics | 12 |
| One Health | 9 |
| Proteomics | 27 |
| Sequence analysis | 5 |
| Single Cell | 26 |
| Statistics and machine learning | 17 |
| Synthetic Biology | 3 |
| Transcriptomics | 23 |

## Welcome to the GTN!

Find out more about Galaxy Training Network

Video created by Geert Bonamie.

## The latest GTN news

Read about new tutorials, features, events and more!

**Jul 20, 2023**
**New Feature: GTN User Preferences**

**Jun 27, 2023**
**AI4Life teams up with GTN to enhance training resources**

**Jun 21, 2023**
**ELIXIR-UK Fellow Launches New FAIR Data Management Training**

**Jun 7, 2023**
**New Feature: Pan-Galactic Tool Search**

**Jun 1, 2023**
**GTN Celebrates Pride Month: Alan Hart & M. Tuberculosis**

See all news

Assembly guides

# Genome assembly with `hifiasm` on Galaxy Australia

Galaxy Australia is capable of *de novo* assembling genomes based on PacBio high fidelity reads built from circular consensus sequence HiFi reads.

This How-to-Guide will describe the steps required to assemble your genome on the Galaxy Australia platform, using multiple workflows (see **Fig 1**) developed in consultations between the Bioplatforms Australia Threatened Species Initiative, Galaxy Australia, and the Australian BioCommons.

**Note:** If you need help, the Galaxy community is both approachable and helpful. Ask them questions!

## Quick start guide

1. Login to Galaxy Australia
2. Create a new history
3. Upload your HiFi `ccs.bam` data files to your Galaxy history
4. Load and execute workflows (links included below), using required options
   - FILE CONVERSION workflow: `BAM to FASTQ + QC v1.0` **optional**
   - ASSEMBLY workflow: `PacBio HiFi genome assembly workflow v2.1`
   - PURGE DUPLICATES workflow: `Purge duplicates from hifiasm assembly v1.0` **optional**
5. Review workflow report and perform additional QC as needed
6. Re-run workflows, or individual tools, as needed

## How to cite the workflows

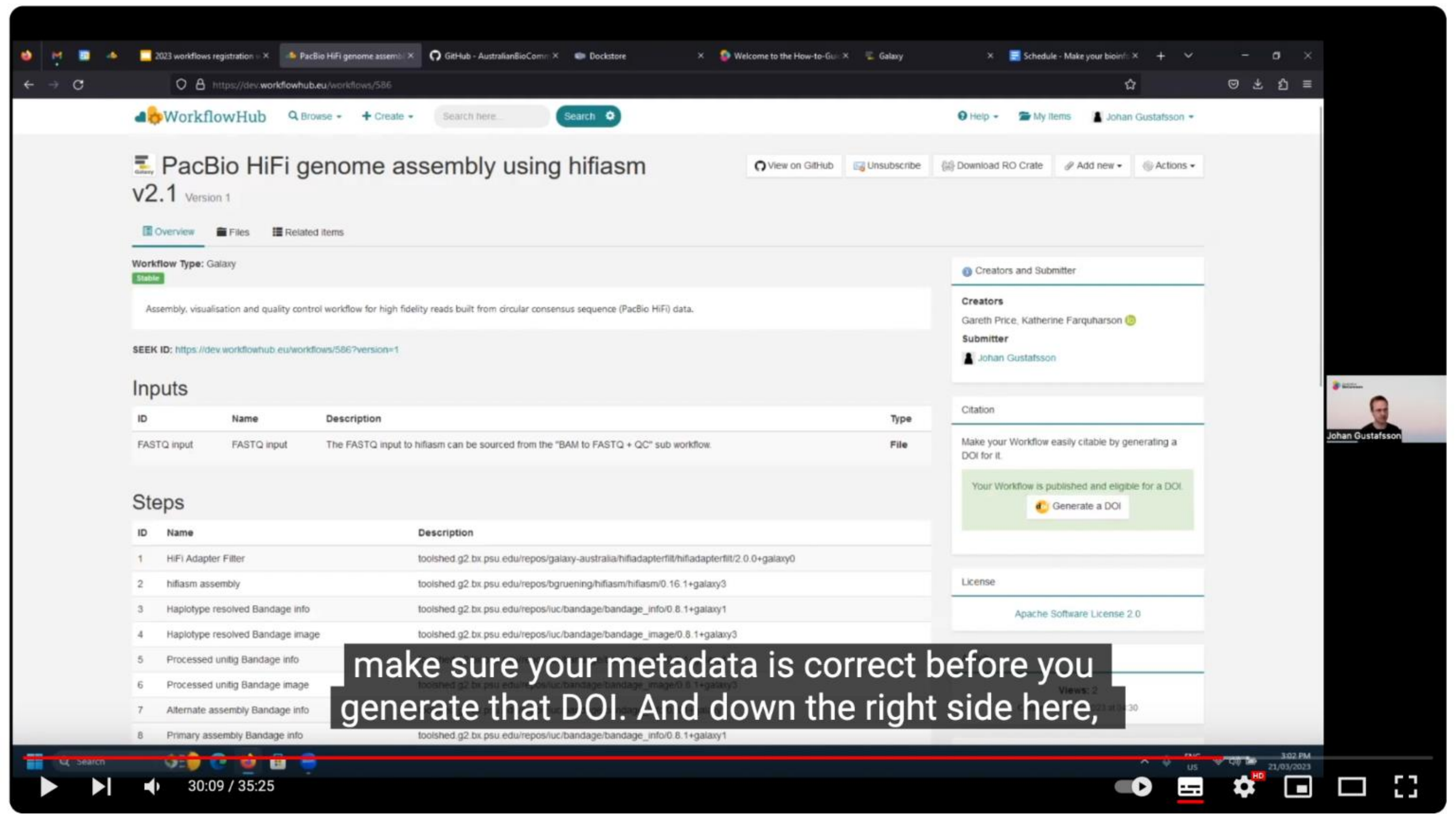
> Price, G. (2022). **BAM to FASTQ + QC v1.0**. WorkflowHub. https://doi.org/10.48546/WORKFLOWHUB.WORKFLOW.220.2

> Price, G., & Farquharson, K. (2022). **PacBio HiFi genome assembly using hifiasm v2.1**. WorkflowHub. https://doi.org/10.48546/WORKFLOWHUB.WORKFLOW.221.3

> Price, G. (2022). **Purge duplicates from hifiasm assembly v1.0**. WorkflowHub. https://doi.org/10.48546/WORKFLOWHUB.WORKFLOW.237.2

## The overall workflow

Fig 1. The approach described in this How-to-Guide, including Quick Start guide steps 1) registration, 2) upload of input BAM file, 3) BAM to FASTQ conversion workflow, 4) assembly workflow, 5) purge duplicates workflow and 6) reviewing the assembly report and FASTA metrics. Required workflow steps are blue, and optional steps are red.



PacBio HiFi genome assembly using hifiasm v2.1 Version 1

**Workflow Type:** Galaxy
Stable

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/586?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 |
| 5 | Processed unitig Bandage info | |
| 6 | Processed unitig Bandage image | |
| 7 | Alternate assembly Bandage info | |
| 8 | Primary assembly Bandage info | toolshed.g2.bx.psu.edu/repos/iuc/bandage/bandage_info/0.8.1+galaxy1 |

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.
Generate a DOI

License

Apache Software License 2.0

make sure your metadata is correct before you generate that DOI. And down the right side here,

Gustafsson, Johan, & Samaha, Georgina. (2023). WORKSHOP: Make your bioinformatics workflows findable and citable. Zenodo. https://doi.org/10.5281/zenodo.7787488

✓ ✓ Visible

✓ ✓ ✓ ✓ **Reusable**

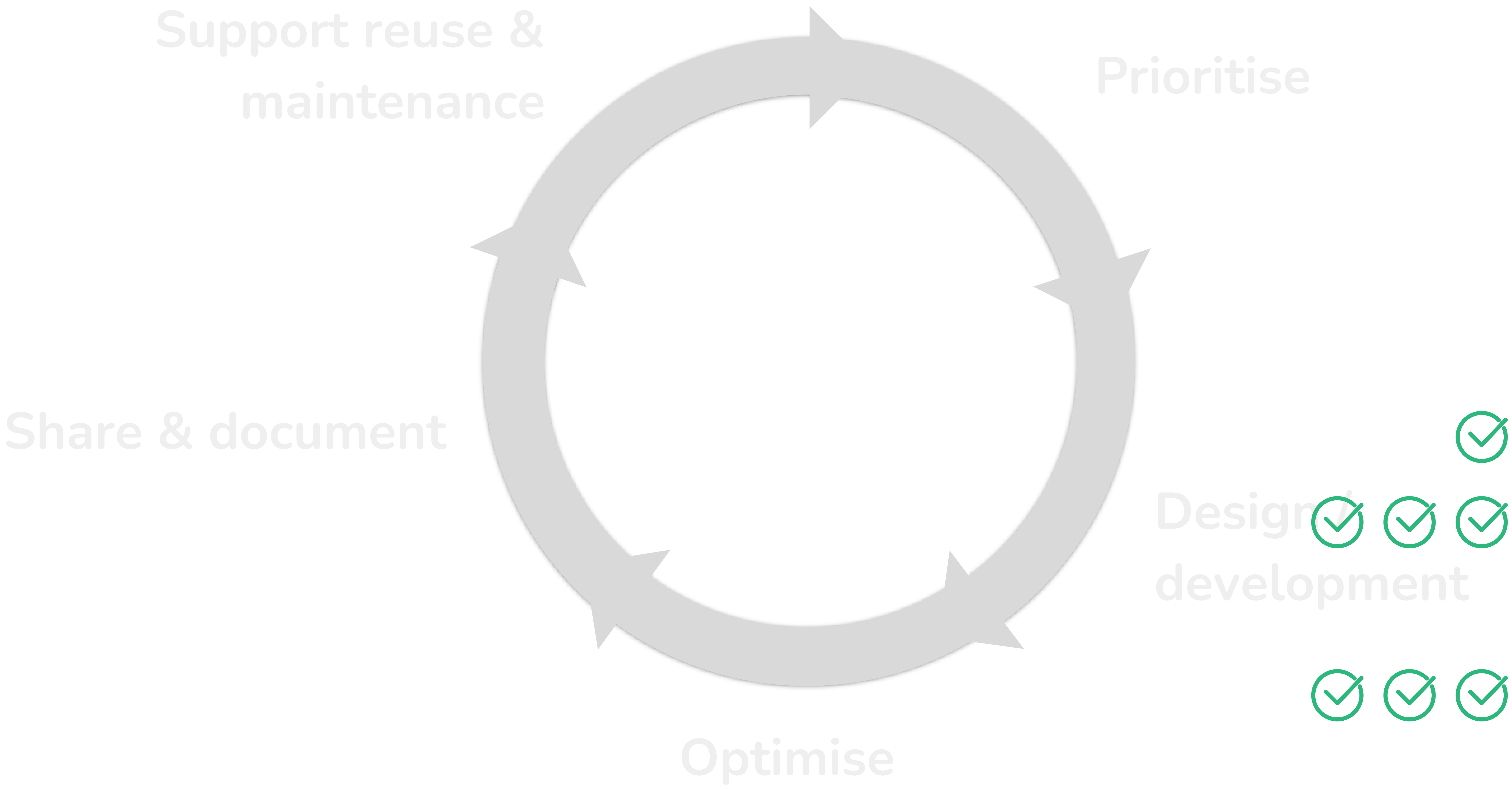
✓ Reproducibility

✓ ✓ ✓ ✓ Best practice

✓ Widely accepted

✓ ✓ ✓ **Sustainable**

# Real world example: Galaxy / Nextflow



✓ ✓ Visible

✓ ✓ ✓ ✓ Reusable

✓ Reproducibility

✓ ✓ ✓ ✓ Best practice

✓ Widely accepted

✓ ✓ ✓ Sustainable

# Thanks!

Any questions?

You can email me at: johan@biocommons.org.au

Australian
BioCommons