FAIR Workflows and WorkflowHub

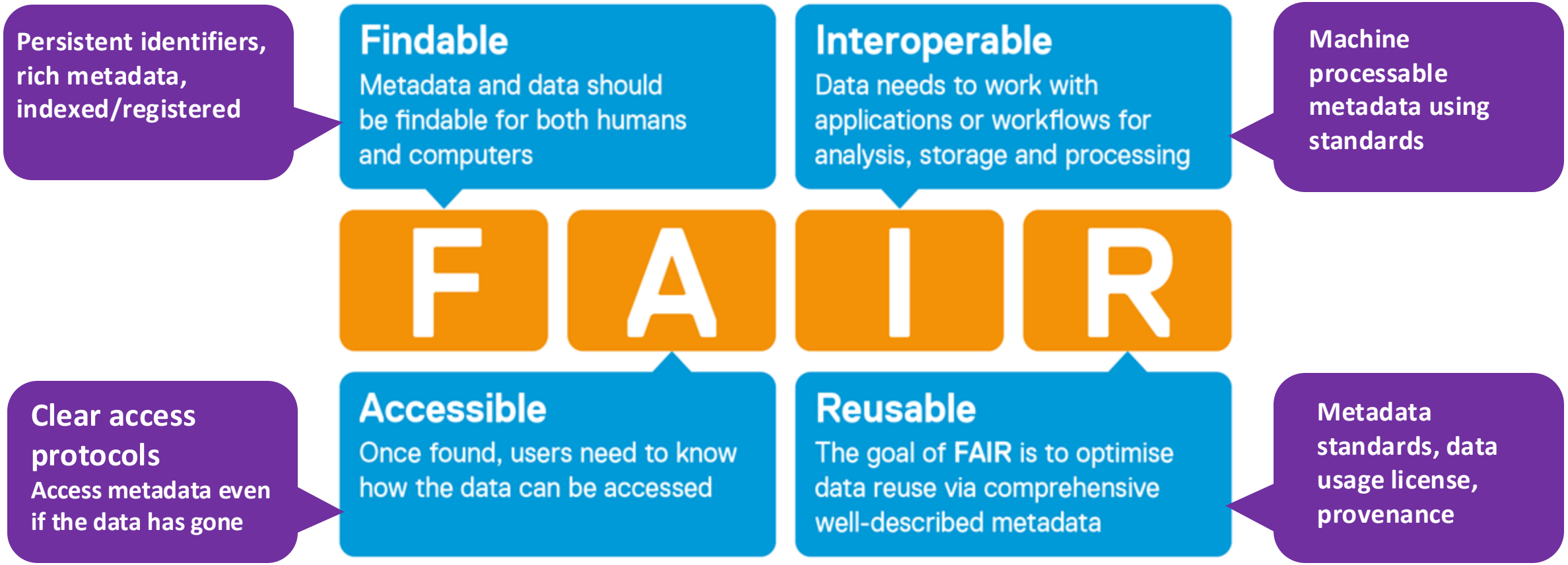
Policy for Findable, Accessible, Interoperable, Reusable Workflows

Professor Carole Goble
The University of Manchester, UK
ELIXIR UK

FAIR Principles Objects discoverable, reusable **using standards**

Guidelines to set expectations between creators, providers and users

What



How

Scientific Data **3**, 160018 (2016) doi:10.1038/sdata.2016.18

Why FAIR Computational Workflows?

How can I find already existing workflows?

Can I access them? Public or private? Git repository?

What language is it written in?

Can I rework it to use my tool?

Is it well enough described so I can understand it?

Can I use it? Adapt it?

Can I reuse it in our infrastructure?

Does it make FAIR data?

Can I establish a claim for it when I publish?

Will I get credit?

Can I track that credit?

Why FAIR Computational Workflows?

What workflow systems are used?

What is the impact of a workflow?

What workflows can be reused/exchanged?

Can we reproduce a publication result?

What is the workflow provenance for a result?

Can I build citation graphs for workflows?

What is the DOI for a workflow?

FORCE11 The Future of Research Communications and e-Scholarship

RDA RESEARCH DATA ALLIANCE

<ReSA> Research Software Alliance

scientific data

Explore content | About the journal | Publish with us

nature > scientific data > comment > article

Comment | Open access | Published: 15 March 2016

The FAIR Guiding Principles for scientific data management and stewardship

Workflows are Epistemic Artefacts

scientific data

Explore content | About the journal | Publish with us

nature > scientific data > articles > article

Article | Open access | Published: 14 October 2022

Introducing the FAIR Principles for research software

Workflows are Software

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

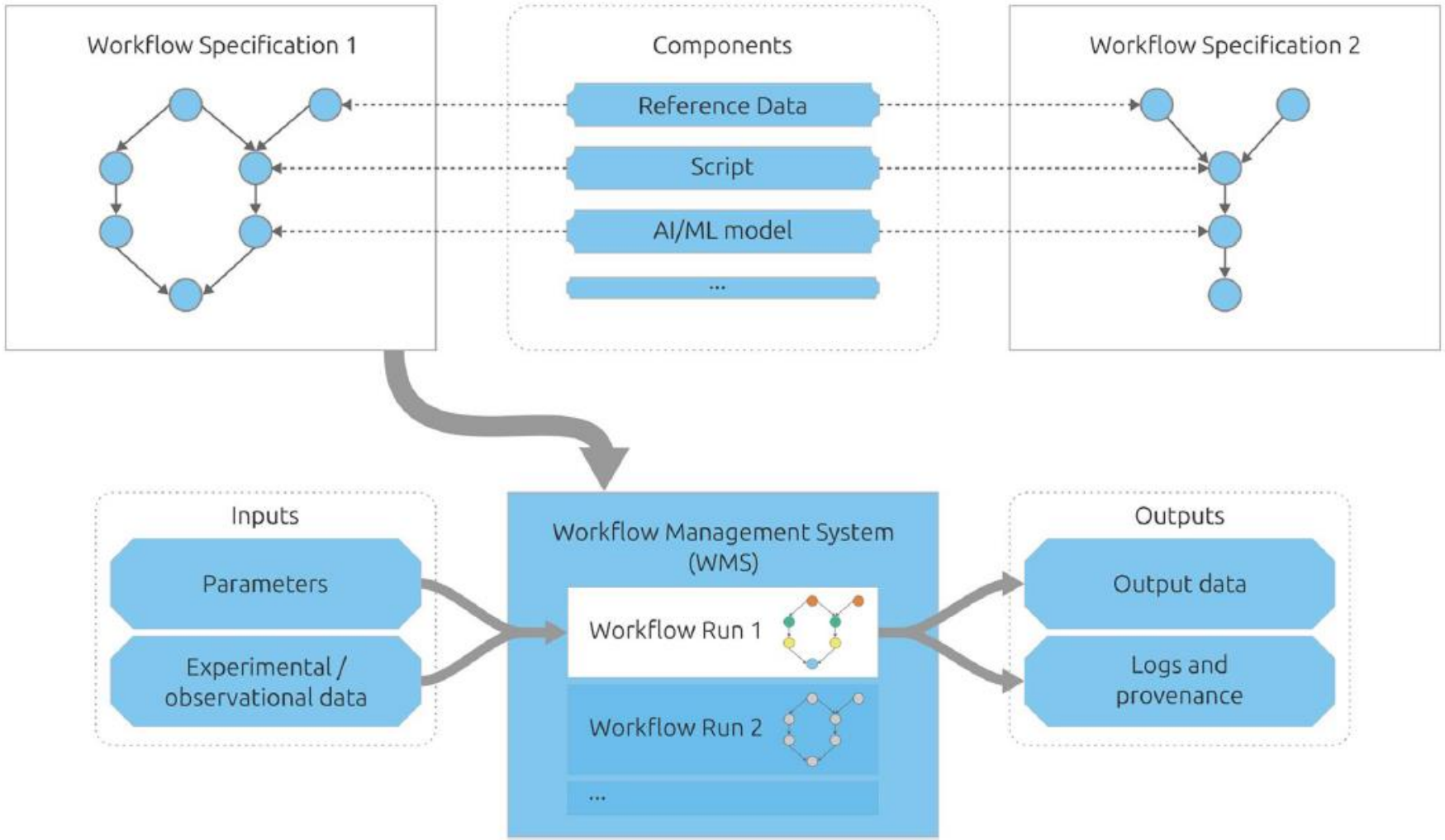
https://doi.org/10.1038/s41597-025-04451-9

workflows community initiative

Applying the FAIR Principles to computational workflows

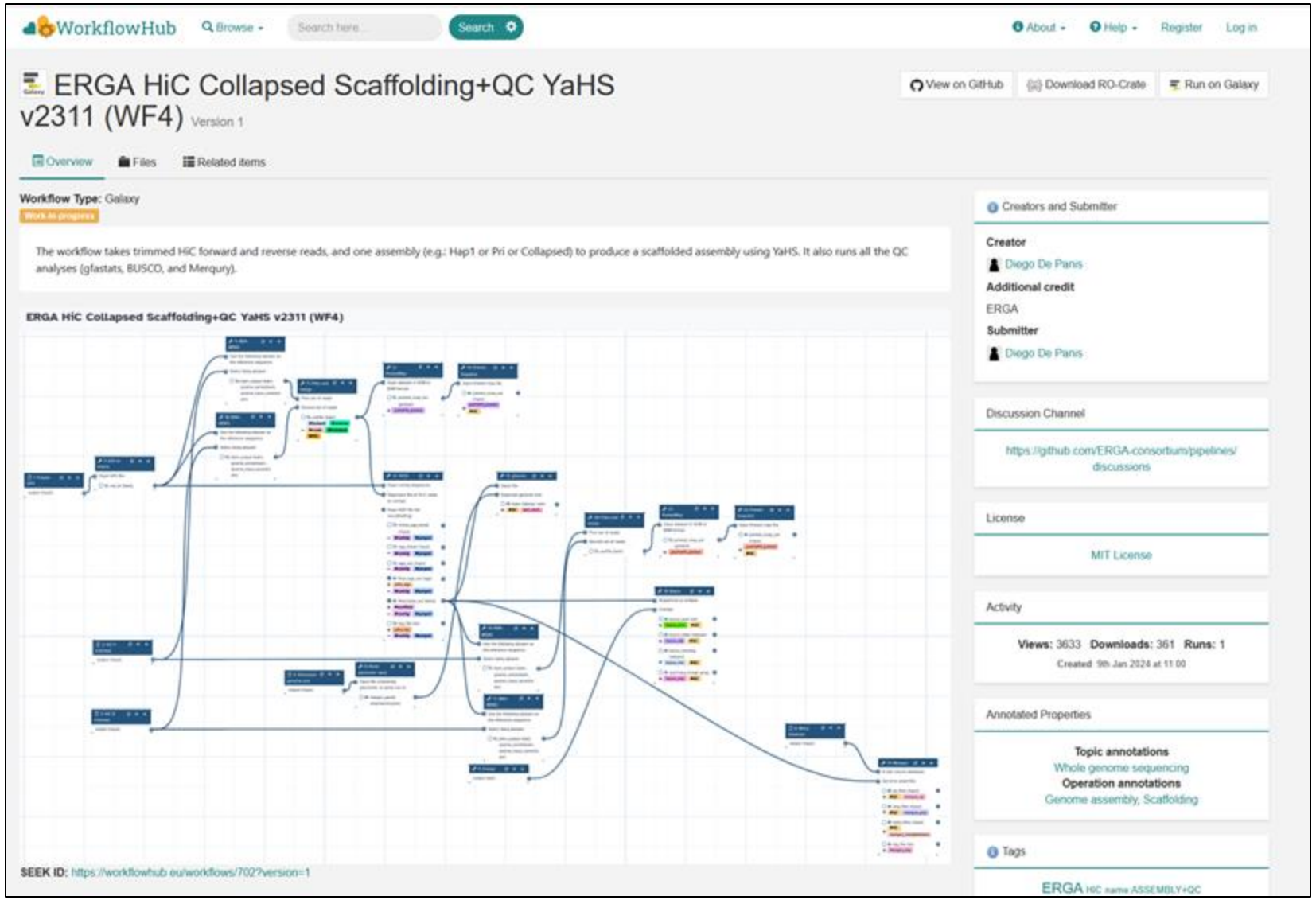
Sean R. Wilkinson[1]✉, Meznah Aloqalaa[2], Khalid Belhajjame[3], Michael R. Crusoe[4], Bruno de Paula Kinoshita[5], Luiz Gadelha[6], Daniel Garijo[7], Ove Johan Ragnar Gustafsson[8], Nick Juty[2], Sehrish Kanwal[9], Farah Zaib Khan[8], Johannes Köster[10], Karsten Peters-von Gehlen[11], Line Pouchard[12], Randy K. Rannow[13], Stian Soiland-Reyes[2,14], Nicola Soranzo[15], Shoaib Sufi[2], Ziheng Sun[16], Baiba Vilne[17], Merridee A. Wouters[18], Denis Yuen[19] & Carole Goble[2]

[1]Oak Ridge Leadership Computing Facility, Oak Ridge National Laboratory, Oak Ridge, Tennessee, USA. [2]Department of Computer Science, University of Manchester, Manchester, UK. [3]LAMSADE, PSL, Paris Dauphine University, Paris, France. [4]Mathematics of Complex Systems division, Visual and Data-Centric Computing department, Bioinformatics in Medicine group, Zuse Institute Berlin (ZIB), Berlin, Germany. [5]Earth Sciences, Barcelona Supercomputing Center, Barcelona, Spain. [6]German Human Genome-Phenome Archive (GHGA, W620), German Cancer Research Center (DKFZ), Heidelberg, Germany. [7]Ontology Engineering Group, Universidad Politécnica de Madrid, Madrid, Spain. [8]Australian BioCommons, University of Melbourne, Melbourne, Victoria, Australia. [9]Clinical Pathology, University of Melbourne Centre for Cancer Research (UMCCR), Parkville, Victoria, Australia. [10]Bioinformatics and computational oncology (Bioinformatische Algorithmen in der Onkologie), Institute for AI in Medicine (IKIM), University Medicine Essen, University of Duisburg-Essen, Essen, Germany. [11]Data Management Department, Deutsches Klimarechenzentrum GmbH, Hamburg, Germany. [12]Center for Computing Research, Sandia National Laboratories, Albuquerque, New Mexico, USA. [13]Silverdraft Supercomputing, Boise, Idaho, USA. [14]Informatics Institute, University of Amsterdam, Amsterdam, The Netherlands. [15]Earlham Institute, Norwich, UK. [16]Center for Spatial Information Science and Systems, Department of Geography and Geoinformation Science, George Mason University, Fairfax, Virginia, USA. [17]Bioinformatics Group, Riga Stradins University, Riga, Latvia. [18]School of Clinical Medicine, University of New South Wales, Kensington, New South Wales, Australia. [19]Ontario Institute for Cancer Research, Toronto, Ontario, Canada. ✉e-mail: wilkinsonsr@ornl.gov



https://doi.org/10.1038/s41597-025-04451-9

Guideline	Based on[1]
F1. A workflow is assigned a globally unique and persistent identifier.	
F1.1. Components of the workflow representing levels of granularity are assigned distinct identifiers.	
F1.2. Different versions of the workflow are assigned distinct identifiers.	
F2. A workflow and its components are described with rich metadata.	
F3. Metadata clearly and explicitly include the identifier of the workflow, and workflow versions, that they describe.	
F4. Metadata and workflow are registered or indexed in a searchable FAIR resource.	
A1. Workflow and its components are retrievable by their identifiers using a standardized communications protocol.	
A1.1. The protocol is open, free, and universally implementable.	
A1.2. The protocol allows for an authentication and authorization procedure, when necessary.	
A2. Metadata are accessible, even when the workflow is no longer available.	
I1. Workflow and its metadata (including workflow run provenance) use a formal, accessible, shared, transparent, and broadly applicable language for knowledge representation.	
I2. Metadata and workflow use vocabularies that follow FAIR principles.	
I3. Workflow is specified in a way that allows its components to read, write, and exchange data (including intermediate data), in a way that meets domain-relevant standards.	
I4. Workflow and its metadata (including workflow run provenance) include qualified references to other objects and the workflow's components.	
R1. Workflow is described with a plurality of accurate and relevant attributes.	
R1.1. Workflow is released with a clear and accessible license.	
R1.2. Components of the workflow representing levels of granularity are given clear and accessible licenses.	
R1.3. Workflow is associated with detailed provenance of the workflow and of the products of the workflow.	
R2. Workflow includes qualified references to other workflows.	
R3. Workflow meets domain-relevant community standards.	



WorkflowHub

https://workflowhub.eu

1200+ Workflows
332 Teams
26 Workflow languages/types

DOI
Versioning
Rich metadata (schema.org)
Git integration
WfMS integration

scientific data

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.

https://doi.org/10.1038/s41597-025-04786-3

FAIR Metadata packaging and exchange

https://www.researchobject.org/ro-crate/

https://www.researchobject.org/workflow-run-crate/

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


POLICY
evaluation
management
implementation
rules
strategy
principle
guide
benefit
agenda
governance
decisions

What?

Policy:

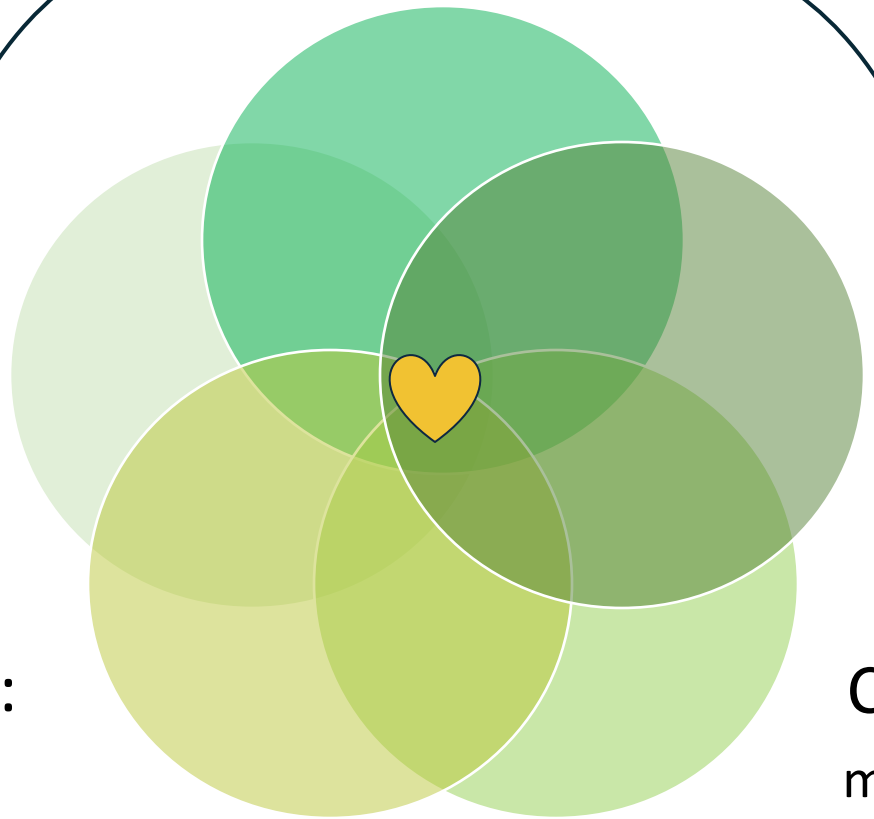
make it required

Infrastructure:

make it possible

Incentives:

make it rewarding



Skills/training:

make it easy

Communities:

make it normative

Adapted by Michelle Barker from Brian Nosek, Strategy for Culture Change (2019)

Who? Interlinked Policy Stakeholders

Policy Forums – OECD, ReSA, RDA, UNESCO, G7, CODATA…

Funders - National, European, Charities…

Publishers - Journals…

Institutions – Universities, National Labs…

Workflow Communities – nf-core, Galaxy, MGnify…

Research Communities/Societies – ELIXIR…

Enablers

Training Community Groups - Carpentries…

Infrastructure Providers – Galaxy, nextflow, WorkflowHub…

FAIR Policy

Policy

Share workflows

Workflow authorship recognised in career development

Reuse workflows

Workflow used in publication is reproducible

Policy

Register workflow

PIDs for workflows and workflow versions

PID, Metadata & licensing

PID, Metadata & provenance

FAIR Impl

Workflow registry
CodeMeta, schema.org

Workflow registry
Citation.cff

Infrastructure produce PIDs & metadata export

Infrastructure produce PIDs, metadata & history export

Community level guidance and policy …

Communities make workflow standards and registration:

- Normative
- Possible
- Required

CodeMeta

Citiation.cff

Galaxy PROJECT

WorkflowHub

nf-core

Requirements

All nf-core pipelines *must* follow the following guidelines:

- Nextflow: Workflows must be built using Nextflow.
- Community owned: Pipelines are owned by the community.
- Identity and branding: Primary development must on the nf-core organisation.
- Workflow specificity: There should only be a single pipeline per data / analysis type.
- Workflow size: Not too big, not too small.
- Workflow name: Names should be lower case and without punctuation.
- Use the template: All nf-core pipelines must be built using the nf-core template.
- Software license: Pipelines must open source, released with the MIT license.
- Bundled documentation: Pipeline documentation must be hosted on the nf-core website.
- Docker support: Software must be bundled using Docker and versioned.
- Continuous integration testing: Pipelines must run CI tests.
- Semantic versioning: Pipelines must use stable release tags.
- Standardised parameters: Strive to have standardised usage.
- Single command: Pipelines should run in a single command.
- Keywords: Excellent documentation and GitHub repository keywords.

https://nf-co.re/docs/guidelines/pipelines/overview

Publishers…

(GIGA)n SCIENCE

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.

Persistent Identifier proliferation
Feed KG and citation engines
Do not disrupt publication process

Career credit
Institutional support

Institutions…

- Normative
- Possible
- Required
- Rewarding

Infrastructure Providers…

Built in
PID minting, registration, metadata collection …

What about the rest of the ecosystem?

- Easy
- Normative
- Possible
- Required
- Rewarding

Policy Partnerships

Leverage/infiltrate current software and data policy

Research Software Policy Forum https://www.researchsoft.org/software-policies/

Committee for Scientific and Technological Policies, Policy Toolkit

Open Research Funders Group

Toolkit

or create our own WCI Policy?

Acknowledgements

Sean Wilkinson, Oak Ridge National Labs, USA
Johan Gustafsson, Australian Biocommons
Michael Crusoe, Zuse Institute, Berlin, ELIXIR Germany

Many others from the WCI FAIR Workflows WG and ELIXIR.