

WorkflowHub

FAIR Workflow Registry

Justin Clark-Casey

EOSC Programme Manager, EMBL-EBI

ELIXIR Europe

On behalf of:

Professor Carole Goble

The University of Manchester, UK

ELIXIR-UK

Finn Bacall (ELIXIR UK)

Frederik Coppens (ELIXIR Belgium)

WorkflowHub Club

FAIR enabling practices, EOSC Symposium, Prague, CZ 2022-11-16



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824087

<https://doi.org/10.5281/zenodo.7323471>



Australian
BioCommons



eosc | EuroScienceGateway

eosc | cancer

eosc | FAIR-IMPACT
Expanding FAIR solutions across EOSC



Synthesis of Systematic Resources a DiSSCo project



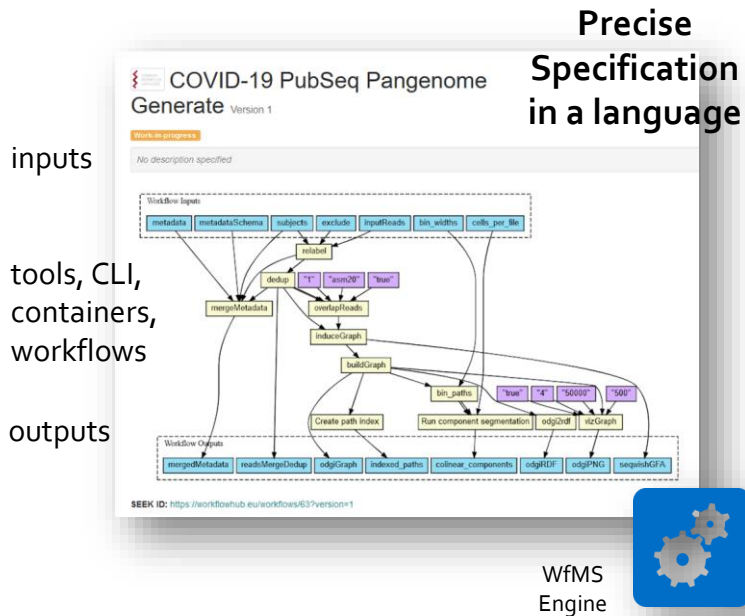


Computational Workflows make EOSC data analysis FAIR

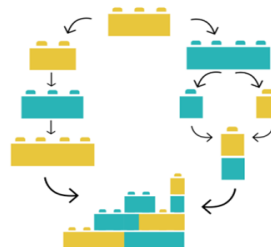




What is a workflow ...



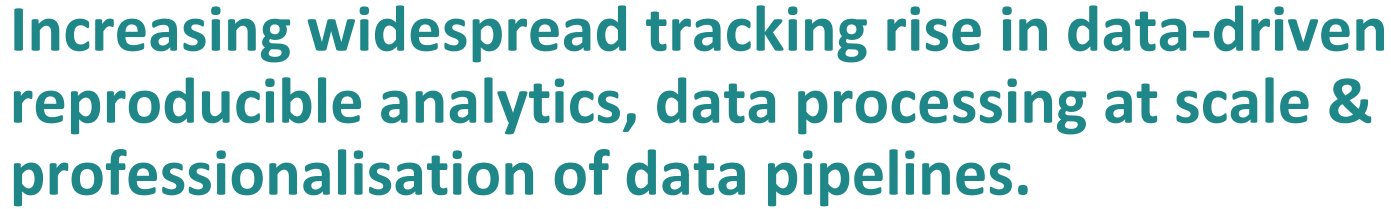
Computational data pipelines and analytics, chain codes, handle data flow, manage execution on computational platforms.



Access EOSC computation and data infrastructure, tool **interoperability**, processing **portability** and optimisation

Software Execution on EOSC computational platforms / clouds



[illegible]

The diagram illustrates a Galaxy workflow for protein structure prediction. It begins with a 'FASTA' file input, which is processed by 'Search databases to hit structures from a database of TMHMM models'. This step leads to 'Compound enumeration'. The workflow then proceeds through several steps: 'Generate 50 structure of each molecule', 'Split into size chunks', 'Remove solvents and ions', 'Produce protein in gpc 7A', 'Refine molecule (amc)', 'Docking', 'Score with YASARI and SuCoS', and finally 'Rank'. The output is a 'Rank' file. The workflow is part of the 'Galaxy PROJECT' and is associated with 'Compound enumeration' and 'Protein preparation'.

[illegible][illegible][illegible]

```

graph TD
    A[Sequencing Data BCL] --> B[Demultiplexing + Adapter Removal Integration]
    B --> C[Intermediate Data Caching]
    B --> D[Annotation & Indexing]
    C --> E[Mapping]
    E --> F[Amplification Assignment & Feature Filtering]
    F --> G[Postprocessed BCL]
    G --> H[Variant Calling Workflow]
    H --> I[Annotation]
    I --> J[Variant Reports]
    J --> K[Variant Calling]
    K --> L[Variant Reports]
  
```

The flowchart illustrates the Nextflow architecture, showing the flow of data from sequencing to variant calling and reporting. The process starts with 'Sequencing Data (BCL)' which is processed through 'Demultiplexing + Adapter Removal (Integration)'. This leads to 'Intermediate Data Caching' (which also receives 'Probes & Name Resolution' and 'Integration (Integration Response)') and 'Annotation & Indexing' (which also receives 'Target (File & Name)' and 'Indexing (Indexing Response)'). 'Intermediate Data Caching' leads to 'Mapping' (which also receives 'Reference (FASTA)' and 'Probes (FASTA)'). 'Mapping' leads to 'Amplification Assignment & Feature Filtering' (which also receives 'Probes (FASTA)'). This leads to 'Postprocessed BCL' (which also receives 'Target (File & Name)' and 'Indexing (Indexing Response)'). 'Postprocessed BCL' leads to 'Variant Calling Workflow' (which also receives 'CIR definition (JSON)' and 'Target (File & Name)'). This leads to 'Annotation' (which also receives 'CIR Reports' and 'Target (File & Name)'). 'Annotation' leads to 'Variant Reports' (which also receives 'Table 1: Variant A/B' and 'Table 2: Variant A/B'). 'Variant Reports' leads to 'Variant Calling' (which also receives 'Table 1: Variant A/B' and 'Table 2: Variant A/B'). The final output is 'Variant Reports' (which also receives 'Table 1: Variant A/B' and 'Table 2: Variant A/B').

Explosion of Workflow Management Systems



1. Arvados - CWL-based distributed computing platform for data analysis on massive data sets. <https://arvados.org/>

2. Apache Taverna <https://www.taverna.org.uk/>

3. Galaxy <https://galaxyproject.org/>

4. SHIVA <https://www.shiva-workflow.eu/>

5. Apache Oozie <https://cosmos.hms.harvard.edu/2014/07/24/bioinformatic-specification-v1.0/>

6. DNANexus <https://www.dnanexus.com/>

7. BioDT <http://www.biodatomi.com/>

8. Agave <http://agave.dcc.ucsb.edu/>

9. Discoverly <https://discoverly.com/>

10. WINGS <https://wings.hawaii.edu/>

11. Knime <https://www.knime.com/>

12. make, rake, drake, compiling and parallel supporting pairs <https://code.google.com/p/make/>

13. Snakemake <https://snakemake.github.io/>

14. BPipe <http://bpipe.sourceforge.net/>

15. Ruffus <https://ruffus.sourceforge.net/>

16. NextFlow <https://www.nextflow.io/>

17. Luigi. Python pack <https://github.com/spotify/luigi>

18. SciLuigi. Helper lib <https://github.com/spotify/luigi>

19. Luigi Analysis Wor <https://github.com/spotify/luigi>

20. GATK Queue <https://github.com/spotify/luigi>

21. Abi <https://ccg.mcgill.ca/abi/>

22. Seqware Workflow <https://seqware.com/>

23. Pipelines <https://pipelines.psc.edu/>

24. Keweenaw <https://keweenaw.org/>

25. Apache Airavat <https://airavat.apache.org/>

26. Pyflow <https://pyflow.github.io/>

27. Cluster Flow <https://clusterflow.github.io/>

28. UniFlow <https://uniflow.github.io/>

29. CloudJung <https://cloudjung.github.io/>

30. Stacks <https://caterpillarstack.com/>

31. Lead <https://www.francisco.io/lead/>

32. OmicTools <https://omictools.com/>

33. Job Description Language <https://jdl.sourceforge.net/>

34. Autodes Cloud Comp <https://autodes.com/>

35. YAWL yet another workflow language <https://www.yawl-project.org/>

36. Triquetrum <https://projects.ece.cmu.edu/projects/techno/>

37. Kronos <https://github.com/jlaghiy/kronos>

38. GNU Guix Workflow Language <https://git.roel.nl/guix/gwl/guix/>

39. Porcine <https://github.com/porcine/porcine>

40. Parsl (a Parallel Scripting L) <https://parsl-project.org/>

41. ECFlow (Workflow primar) <https://ecflow.org/>

42. Ophidia <https://ophidia-project.github.io/>

43. Weblight <https://weblight-project.github.io/>

44. CloudEngine <https://cloudengine.eubac.eu/>

45. SCIPION <https://scipion-project.github.io/>

46. FASTR <https://fastr-project.github.io/>

47. BioMake <https://biomake.github.io/>

48. remake <https://github.com/ricmiller/remake>

49. SciFlow <https://www.sop.inria.fr/sciflow/>

50. OpenAlea <https://openalea.gforge.inria.fr/>

51. Pradai <https://pradai.github.io/>

52. LONI Pipeline <https://loni.github.io/>

53. Cipe <https://ciper-project.org/>

54. AWE <https://awe-project.github.io/>

55. PyCOMPS <https://pycomps.github.io/>

56. KLUKO <https://kluko.github.io/>

57. SoS Workflow <https://sos-workflow.github.io/>

58. Pachyderm <https://pachyderm.io/>

59. Digidag <https://digidag.io/>

60. XNAT Pipeline Engine <https://xnat.org/>

61. OCCAM (Open Curator) <https://occam-project.github.io/>

62. Copenius <https://copenius.github.io/>

63. iRODS Rule Language <https://irods.org/>

64. BioVista <https://biovista.github.io/>

65. Biode Watermill <https://biode-watermill.github.io/>

66. BIOVIA Pipeline Pilot <https://biobva.com/>

67. Dagman A meta-sched <https://dagman-project.github.io/>

68. UNICORE <https://unicore-project.org/>

69. Toil (A scalable, efficient) <https://toil.github.io/>

70. Nephelie <https://nephelie.nid.nhs.uk/>

71. TOPAS <https://topas-project.github.io/>

72. Sdpipes <https://sdpipes.github.io/>

73. Day-Header <https://day-header.github.io/>

74. GenomeVIP <https://genomevip.github.io/>

75. GDSIM <https://sourceforge.net/projects/gdsim/>

76. Roddy <https://roddy-project.github.io/>

77. SciFlow (historical, doesn't seem to be maintained anymore) <https://www.sci-flow.org/>

78. JTracker <https://jtracker.io/>

79. PipelineDog <https://pipeline-dog.github.io/>

80. DALiGE <https://dalige.github.io/>

81. Overseek <https://overseek.github.io/>

82. dynamic <https://dynamic-project.github.io/>

83. XIP <https://xip.hcg.jp/wiki/en/>

84. Eoulisan <https://eoulisan.github.io/>

85. TIGR <https://tigr-project.github.io/>

86. Archiver <https://archiver-project.github.io/>

87. NG6 <https://ng6.toulouse.inra.fr/>

88. BioMAJ <https://biomaj.github.io/>

89. Conveyor <https://conveyor-project.github.io/>

90. Biopipe <https://biopipe-project.github.io/>

91. Wildfire <https://wildfire-project.github.io/>

92. BioWB <https://biowb-project.github.io/>

93. BioWMS <https://biowms-project.github.io/>

94. BioMoby <https://biomoby-project.github.io/>

95. SIBIOS <https://sibios-project.github.io/>

96. NSANE <https://nsane-project.github.io/>

97. Pwake <https://pwake-project.github.io/>

98. Nephelie <https://nephelie-project.github.io/>

99. Skam <https://skam-project.github.io/>

100. TREMA <https://trema-project.github.io/>

101. EGene <https://egene-project.github.io/>

102. Kashiwara <https://kashiwara-project.github.io/>

103. FlowMars <https://flowmars-project.github.io/>

104. Plumbing and Graph <https://plumbing-project.github.io/>

105. LabView <https://labview-project.github.io/>

106. MyOpenLab <https://myopenlab-project.github.io/>

107. Max/MSP <https://maxmsp-project.github.io/>

108. NoFlo <https://noflojs.org/>

109. Flowstone <https://flowstone-project.github.io/>

110. Hyperloom <https://hyperloom-project.github.io/>

111. Dask <https://dask-project.org/>

112. Stimela <https://stimela-project.github.io/>

113. JTracker <https://jtracker.io/>

114. JTracker <https://jtracker.io/>

115. PipelineDog <https://pipeline-dog.github.io/>

116. DALiGE <https://dalige.github.io/>

117. Overseek <https://overseek.github.io/>

118. dynamic <https://dynamic-project.github.io/>

119. XIP <https://xip.hcg.jp/wiki/en/>

120. Eoulisan <https://eoulisan.github.io/>

121. TIGR <https://tigr-project.github.io/>

122. Archiver <https://archiver-project.github.io/>

123. NG6 <https://ng6.toulouse.inra.fr/>

124. BioMAJ <https://biomaj.github.io/>

125. Conveyor <https://conveyor-project.github.io/>

126. Biopipe <https://biopipe-project.github.io/>

127. Wildfire <https://wildfire-project.github.io/>

128. BioWB <https://biowb-project.github.io/>

129. BioWMS <https://biowms-project.github.io/>

130. BioMoby <https://biomoby-project.github.io/>

131. SIBIOS <https://sibios-project.github.io/>

132. NSANE <https://nsane-project.github.io/>

133. Pwake <https://pwake-project.github.io/>

134. Nephelie <https://nephelie-project.github.io/>

135. Skam <https://skam-project.github.io/>

136. TREMA <https://trema-project.github.io/>

137. EGene <https://egene-project.github.io/>

138. Kashiwara <https://kashiwara-project.github.io/>

139. FlowMars <https://flowmars-project.github.io/>

140. Plumbing and Graph <https://plumbing-project.github.io/>

141. LabView <https://labview-project.github.io/>

142. MyOpenLab <https://myopenlab-project.github.io/>

143. Max/MSP <https://maxmsp-project.github.io/>

144. NoFlo <https://noflojs.org/>

145. Flowstone <https://flowstone-project.github.io/>

146. Hyperloom <https://hyperloom-project.github.io/>

147. Dask <https://dask-project.org/>

148. Stimela <https://stimela-project.github.io/>

149. JTracker <https://jtracker.io/>

150. JTracker <https://jtracker.io/>

151. PipelineDog <https://pipeline-dog.github.io/>

152. DALiGE <https://dalige.github.io/>

153. Overseek <https://overseek.github.io/>

154. dynamic <https://dynamic-project.github.io/>

155. XIP <https://xip.hcg.jp/wiki/en/>

156. Eoulisan <https://eoulisan.github.io/>

157. TIGR <https://tigr-project.github.io/>

158. Archiver <https://archiver-project.github.io/>

159. NG6 <https://ng6.toulouse.inra.fr/>

160. BioMAJ <https://biomaj.github.io/>

161. Conveyor <https://conveyor-project.github.io/>

162. Biopipe <https://biopipe-project.github.io/>

163. Wildfire <https://wildfire-project.github.io/>

164. BioWB <https://biowb-project.github.io/>

165. BioWMS <https://biowms-project.github.io/>

166. BioMoby <https://biomoby-project.github.io/>

167. SIBIOS <https://sibios-project.github.io/>

168. NSANE <https://nsane-project.github.io/>

169. Pwake <https://pwake-project.github.io/>

170. Nephelie <https://nephelie-project.github.io/>

171. Skam <https://skam-project.github.io/>

172. TREMA <https://trema-project.github.io/>

173. EGene <https://egene-project.github.io/>

174. Kashiwara <https://kashiwara-project.github.io/>

175. FlowMars <https://flowmars-project.github.io/>

176. Plumbing and Graph <https://plumbing-project.github.io/>

177. LabView <https://labview-project.github.io/>

178. MyOpenLab <https://myopenlab-project.github.io/>

179. Max/MSP <https://maxmsp-project.github.io/>

180. NoFlo <https://noflojs.org/>

181. Flowstone <https://flowstone-project.github.io/>

182. Hyperloom <https://hyperloom-project.github.io/>

183. Dask <https://dask-project.org/>

184. Stimela <https://stimela-project.github.io/>

185. JTracker <https://jtracker.io/>

Interactive Electronic Research Scripting environments Workflow management systems and execution platforms



Are workflows Findable, Accessible, Interoperable, Reusable?



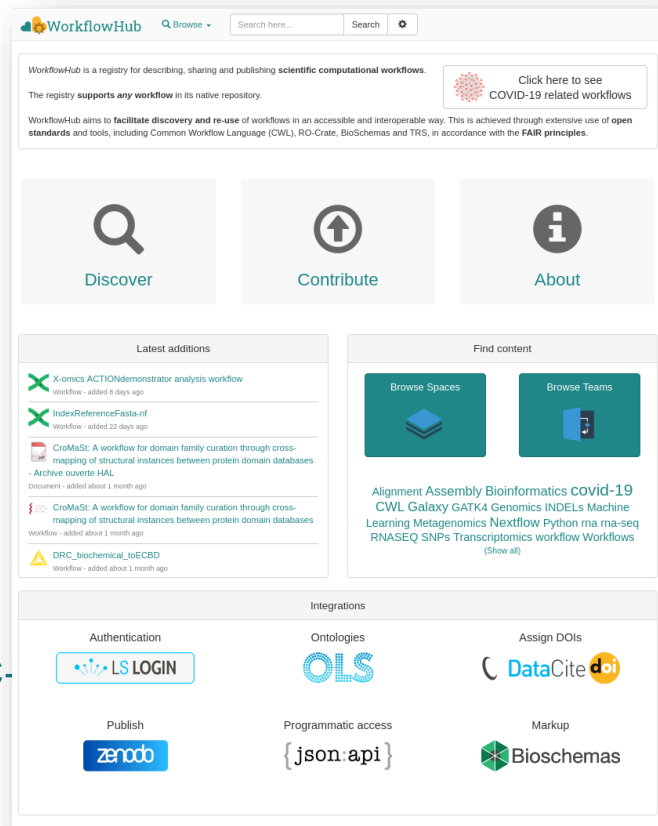
<https://workflowhub.eu>



WorkflowHub

FAIR Workflow Registry

Launched Sept 2020
EOSC service provided by ELIXIR, EOSC-
Life, The University of Manchester
Open Development



283 workflows

11 system types

110 teams

103 organisations

360 people



This project has received funding from the European Union's Horizon
2020 research and innovation programme under grant agreement No 824087

FAIR Workflow registry for EOSC



Workflow-system **agnostic**



Search for and **discover** workflows



Metadata standardization

(CWL, schema.org, custom tags, RO-Crate)



DOI publication, citation & credit



Collections, Teams, Organizations and **Communities**



Programmatic access: GA4GH TRS API, RO-Crate



Registry, not repository

Workflows can live elsewhere, e.g. GitHub



Integration with EOSC **execution** platforms

The screenshot displays the WorkflowHub web interface. At the top, there is a search bar and navigation links. The main content area shows a workflow titled "Jupyter Notebook Mutation Free Energy Calculations using BioExcel Building Blocks (biobb)" with version 4 (latest). The workflow is categorized as "Jupyter" and "Stable". It includes a description of the workflow, which is based on the official pmx tutorial. The workflow is created by Genis Bayram and Adam Hospital. The interface also shows a version history section, a citation section, and a discussion channels section. The workflow is available for download as a RO-Crate and can be viewed on GitHub. The interface is designed to be user-friendly and accessible, with clear navigation and detailed information about the workflow.



SEEK ID: <https://workflowhub.eu/workflows/55?version=4>

DOI: [10.48546/workflowhub.workflow.55.4](https://doi.org/10.48546/workflowhub.workflow.55.4)

Version History

Version 4 (latest) Created 16th Sep 2022 at 10:35 by Genis Bayram
Update to BioBB 3.8.*, taken from Git commit 5b20069

Versions &
status

Citation &
License



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824087

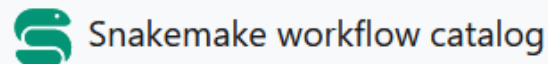


Low barrier to entry for publishing workflows

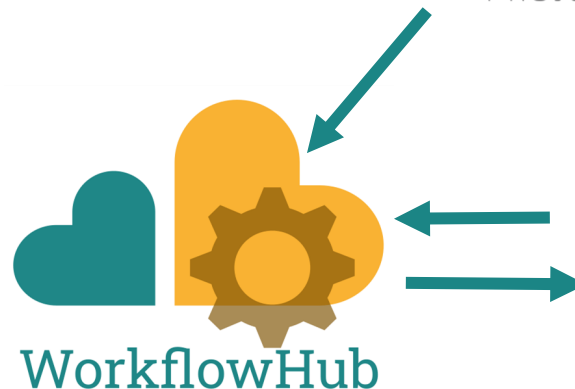


- ✓ Workflow-system agnostic
- ✓ Import from native repositories
- ✓ Git integration
- ✓ Versions, any stage of development
- ✓ Automated metadata extraction
- ✓ Include documents, test data

tl;dr: Workflows can remain in existing repositories



IWC - Intergalactic Workflow Commission





Helping EOSC communities make FAIR workflows



Supporting workflow communities in EOSC



WorkflowHub Teams

95 Teams found

Query: Search here... Go

Created At: Any time

Space: Independent Teams

COVID-19 Biohackathon 63

Australian BioCommons 7

BioExcel 2

EOSC-Life 2

ELIXIR 2

Organism: SARS-CoV-2 10, Homo sapiens 9

Default Condensed Table

AGRF BIO

AGRF is a non-profit organisation providing sequencing services. The AGRF

Space Independent Teams

Public web page: <https://www.agrf.org.au/bioinformatics>

Applied Computational Biology at IEGHMGU

The members of the Applied Computational Biology group at the Institute of disciplines, all possessing deep IT knowledge and experience and a strong sciences in our institute as well as in external collaborations.

Space Independent Teams

Public web page: <https://www.helmholtz-muenchen.de/ieg/research/research/applied-computational-biology>

Australian BioCommons

The Australian BioCommons enhances digital life science research through remains globally competitive, through sustained strategic leadership, research

Space Australian BioCommons

Public web page: <https://www.biocommons.org.au/>

Ay Lab

Dr. Ferhat Ay and his lab are currently located at the La Jolla Institute of Immunology, where they are currently working on the 3D genome and other

Space Independent Teams

Public web page: <https://www.ijl.org/laics/>

BioBB Building Blocks

The BioBB Building Blocks (bioBB) software library is a collection of Python between the wrapped tools, which make them compatible and prepared to different workflow systems, including Galaxy, CWL, Jupyter Notebook and

WorkflowHub Teams serve numerous EOSC organizations, projects and ad-hoc collaborations

Galaxy PROJECT

Freely accessible ready to use global infrastructure and workflows for SARS-CoV-2 monitoring

Request Contact Unsubscribe Actions

Overview Related Items

The COVID-19 pandemic is the first global health crisis to occur in the age of big genomic data. Although data generation capacity is well established and sufficiently standardized, analytical capacity is not. To establish analytical capacity it is necessary to pull together global computational resources and deliver the best open source tools and analysis workflows within a ready to use, universally accessible resource. Such a resource should not be controlled by a single research group, institution, or country. Instead it should be maintained by a community of users and developers who ensure that the system remains operational and populated with current tools. A community is also essential for facilitating the types of discourse needed to establish best analytical practices. Bringing together public computational research infrastructure from the USA, Europe, and Australia, we developed a distributed data analysis platform that accomplishes these goals. It is immediately accessible to anyone in the world and is designed for the analysis of rapidly growing collections of deep sequencing datasets. We demonstrate its utility by detecting allelic variants in high-quality existing SARS-CoV-2 sequencing datasets and by continuous reanalysis of COG-UK data.

The scientific publication associated to the workflows in this collection can be currently accessed as [preprint](#). All data and documentation is available at the project website <https://covid19.galaxyproject.org>.

SEEK ID: <https://workflowhub.eu/collections/2>

Items

- sars-cov-2-pe-illumina-artic-variant-calling/COVID-19-PE-ARTIC-ILLUMINA - Added 11 months ago
- sars-cov-2-pe-illumina-wgs-variant-calling/COVID-19-PE-WGS-ILLUMINA - Added 11 months ago
- sars-cov-2-onl-artic-variant-calling/COVID-19-ARTIC-ONT - Added 11 months ago
- sars-cov-2-se-illumina-wgs-variant-calling/COVID-19-SE-WGS-ILLUMINA - Added 11 months ago
- sars-cov-2-variation-reporting/COVID-19-VARIATION-REPORTING - Added 11 months ago

Galaxy PROJECT

Change picture

Maintainers

Creator: Björn Grüning

Submitter: Björn Grüning

License: Other (Public Domain)

Activity

Views: 367

Created: 23rd Jul 2021 at 08:03

Last updated: 23rd Jul 2021 at 08:11

Tags

covid-19 Galaxy

Add your tags

Catalogue the team's workflow collections, related resources, people and institutions

Request Contact

Subscribe



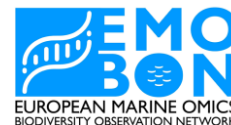
This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824087

Communities at WorkflowHub

nf-core



Australian
BioCommons



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824087

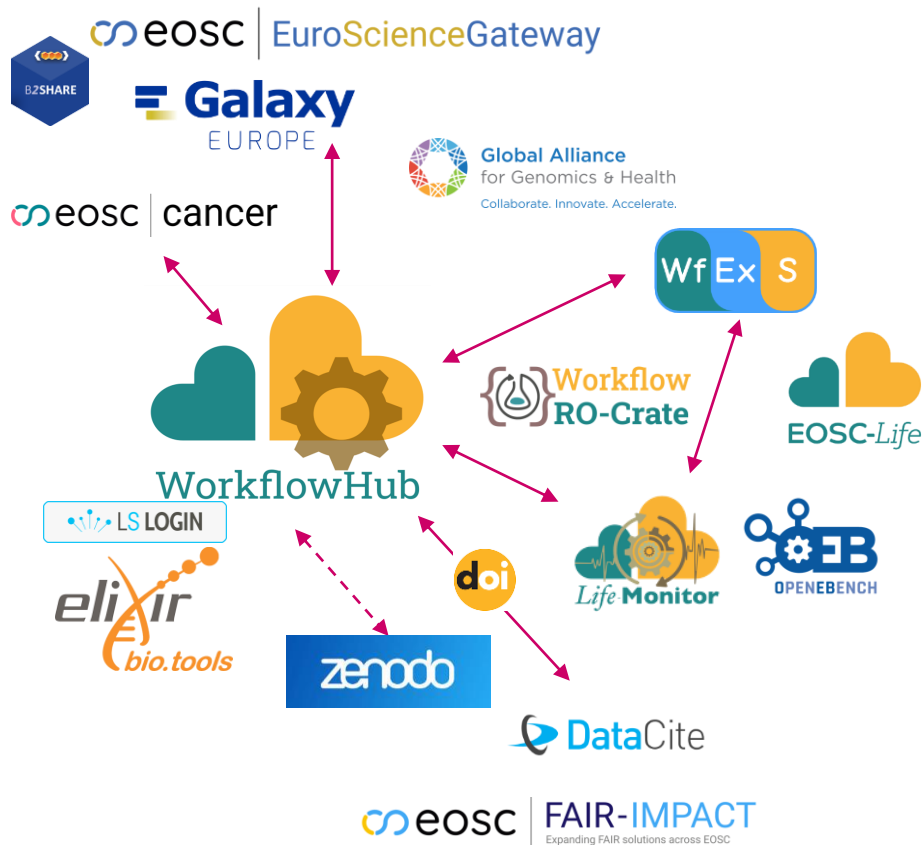
<https://workflowhub.eu/projects>



WorkflowHub integrates with EOSC services



WorkflowHub in the EOSC ecosystem



Services in the

Workflow Collaboratory exchange workflows as
FAIR Digital Objects using **RO-Crate**

Packaging workflow files & companion objects

Exchange between services & systems

Reproducibility & Testing

Citation

Workflows in the EOSC PID graph

Standards-based exchange

<https://doi.org/10.5281/zenodo.4605654>



Acknowledgements

WorkflowHub Club

<https://about.workflowhub.eu/>

Open Development

Bi-weekly community calls

Workflow Community Initiative <https://workflows.community/about>
EOSC-Life <https://www.eosc-life.eu/>

ELIXIR <http://elixir-europe.org>

RO-Crate <https://www.researchobject.org/ro-crate/>

WorkflowHub <https://workflowhub.eu/>

Galaxy Europe <https://galaxyproject.eu/>

Bioschemas <https://bioschemas.org>

Common Workflow Language <https://www.commonwl.org/>

Life Monitor https://crs4.github.io/life_monitor/

Finn Bacall, The University of Manchester

Stuart Owen, The University of Manchester

Stian Soiland-Reyes, The University of Manchester



 eosc | EuroScienceGateway

 eosc | cancer

 eosc | FAIR-IMPACT
Expanding FAIR solutions across EOSC



FAIR Computational Workflows at WCI

