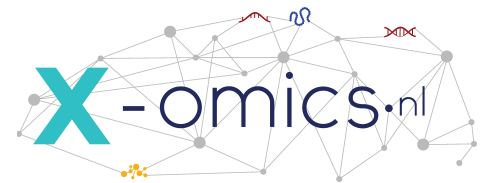


Ten simple rules for building FAIR workflows

Casper de Visser¹, Lennart F. Johansson^{2,3}, Purva Kulkarni¹, Hailiang Mei⁴, Pieter Neerincx^{2,3}, K. Joeri van der Velde^{2,3}, Péter Horvatovich³, Alain J. van Gool¹, Morris A Swertz^{2,3}, Peter A.C. 't Hoen¹, Anna Niehues¹
¹Radboudumc, ²GUMC, ³GU, ⁴LUMC



Introduction

Research data is accumulating rapidly, and with it the challenge of irreproducible science [1,2]. As a consequence, implementation of high quality management of scientific data has become a global priority. The FAIR (Findable, Accessible, Interoperable and Reusable) principles [3] provide practical guidelines for maximizing the value of research data, however, processing data using workflows - systematic executions of a series computational tools - is equally important for good data management [4]. The FAIR principles have recently been adapted to Research Software (FAIR4RS Principles) [5] to promote the reproducibility and reusability of any type of research software. Here we propose a set of 10 simple rules, drafted by experienced workflow developers that will help researchers to apply FAIR4RS principles to workflows.

The FAIR4RS principles

F	Software, and its associated metadata, is easy for both humans and machines to find.
A	Software, and its metadata, is retrievable via standardised protocols.
I	Software interoperates with other software by exchanging data and or metadata, and or through interaction via application programming interfaces (APIs), described through standards.
R	Software is both usable (can be executed) and reusable (can be understood, modified, built upon, or incorporated into other software).

The rules

Findability

1. Register the workflow
2. Describe the workflow with rich metadata

Accessibility

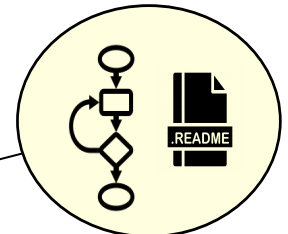
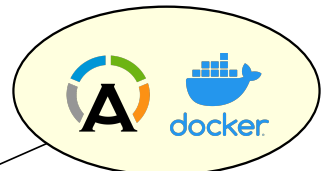
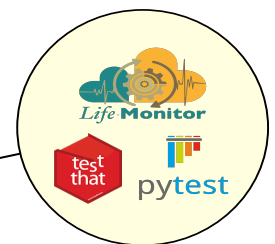
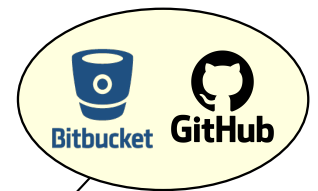
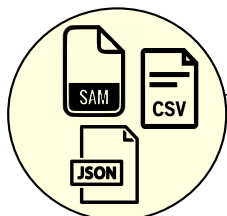
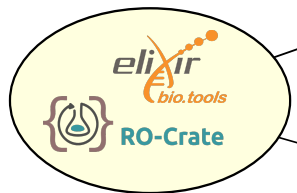
3. Make source code available in a public code repository
4. Provide example input data and results along with the workflow

Interoperability

5. The tools integrated in a workflow should adhere to file format standards
6. Make the workflow portable

Reusability

7. Provide a reproducible computational environment to run the workflow
8. Add a configuration file with defaults
9. Modularize the workflow
10. Provide clear and concise workflow documentation



Conclusion

Altogether, these rules can be seen as practical guidelines for workflow developers who aim to contribute to more reproducible and sustainable computational science, in order to positively impact the open science and FAIR community. For each rule, we propose multiple technical implementations that can be used, because we are aware that not every technical is suitable for every workflow. We believe in the added value of these rules to build a stronger and sustainable workflow community, where reusable, trustworthy and validated workflows are the standard in any data-driven research field.

References

1. Horton N, Alexander R, Parker M, Piekut A, Rundel C. The Growing Importance of Reproducibility and Responsible Workflow In the Data Science and Statistics Curriculum. *Journal of Statistics and Data Science Education* (2022).
2. Madduri R, Chard K, D'Arcy M, Jung SC, Rodriguez A, Sulakhe D, et al. Reproducible big data science: A case study in continuous FAIRness. *PLOS ONE* (2019).
3. Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, et al. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data* (2016).
4. Atkinson M, Gesing S, Montagnat J, Taylor I. Scientific workflows: Past, present and future. (2017).
5. Barker M, Chue Hong NP, Katz DS, Lamprecht AL, Martinez-Ortiz C, Psomopoulos F, et al. Introducing the FAIR Principles for research software. *Sci Data* (2022).