

Critical steps towards large-scale implementation of the FAIR data principles

Bruna dos Santos Vieira^{1,2,9}, César Henrique Bernabé³, Ines Henriques¹, Shuxin Zhang^{4,5}, Alberto Ballasteros Cámara⁶, Jose Antonio Ramírez García⁷, Joeri van der Velde⁸, Philip van Damme^{2,3}, Pablo Alarcón Moreno⁶, Nirupama Benis^{4,5}, Jolanda Strubel⁹, Fieke Schoots⁹, Pauline L'Henaff⁹, P.A.C. 't Hoen¹, Marco Roos³, Annika Jacobsen³, Ronald Cornet^{4,5}, Mark D. Wilkinson⁶, Franz Schaefer⁷, Morris Swertz⁸, Mijke Jetten⁹

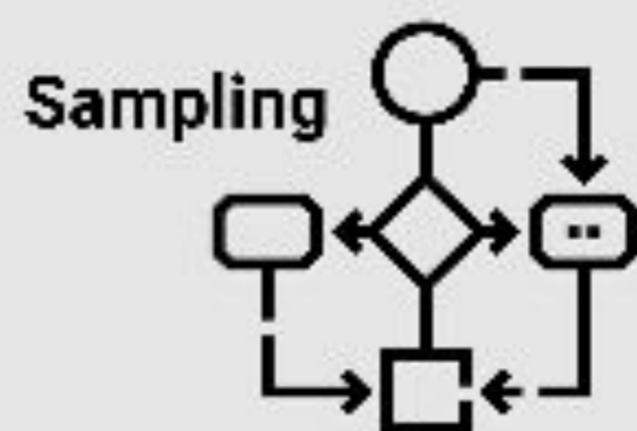
¹Centre for Molecular and Biomolecular Informatics, Radboud University Medical Center, Nijmegen, The Netherlands, ²Department of Medical Imaging, Radboud University Medical Center, Nijmegen, The Netherlands, ³Department of Human Genetics, Leiden University Medical Center, Leiden, The Netherlands, ⁴Department of Medical Informatics, Amsterdam UMC location University of Amsterdam, Amsterdam, The Netherlands, ⁵Amsterdam Public Health, Digital Health & Methodology, Amsterdam, The Netherlands, ⁶Departamento de Biotecnología-Biología Vegetal, Escuela Técnica Superior de Ingeniería Agronómica, Alimentaria y de Biosistemas, Centro de Biotecnología y Genómica de Plantas. Universidad Politécnica de Madrid (UPM) - Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria-CSIC (INIA-CSIC), Madrid, Spain, ⁷Division of Paediatric Nephrology, Center for Paediatrics and Adolescent Medicine, University of Heidelberg, Heidelberg, Germany, ⁸University of Groningen and University Medical Center Groningen, Genomics Coordination Center, Groningen, The Netherlands, ⁹Health-RI, Utrecht, The Netherlands.

Introduction

The process in which data is made Findable, Accessible, Interoperable and Reusable (FAIR), known as FAIRification, varies across projects, domains, and objectives.

To identify the most efficient way to turn resources FAIR on a larger-scale, we aim to compare the most common FAIRification workflows used in the Netherlands and identify shared crucial FAIRification steps.

Methods



Seven workflows were selected based on the criteria:

- Workflows described by **FAIR demonstrator projects** within (at least) the Netherlands;
- Workflows containing **FAIRification steps** (not only Research data management steps);
- Workflows **published** in scientific communication or **endorsed** by a university medical center.



A group of EJP RD and Health-RI experts was formed to analyse the workflows by identifying if:

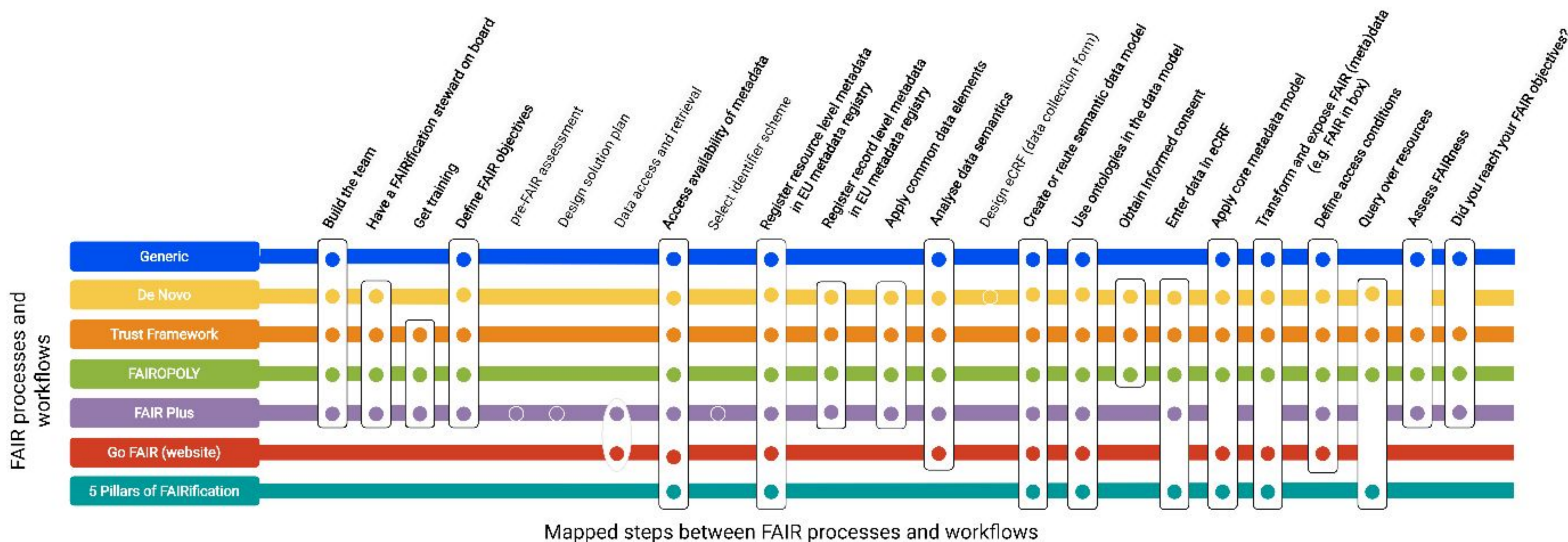
- Several steps meant the same thing;
- The naming convention was the same;
- Subtasks were also a match;
- A step appears in at least 50% of workflows (≥ 4 workflows).



Each mapped step was detailed into a recipe for completing the steps, including:

- Step checklist;
- Needed expertise;
- Related resources.

Results



Discussion and Conclusion

The majority of the processes focused on making existing data FAIR (retrospective oriented). Most observed overlaps related to identifying FAIR objectives, team's expertise, (meta)data assessment and semantic modeling, and defining data licensing and access conditions.

Future Work

- To identify domain-specific steps;
- To complete detailing all critical steps into recommended resources and checklists and to list the needed expertise;
- To endorse and support (e.g. training and capacity building) the common steps in the national research infrastructure.