

# FAIR data and metadata – The X-omics FAIR Data Cube and its added value for multi-omics researchers

Anna Niehues<sup>1,2</sup>, XiaoFeng Liao<sup>1</sup>, Martin Brandt<sup>3</sup>, Cenna Doornbos<sup>1</sup>, Tom Ederveen<sup>1</sup>, Fiona Hagenbeek<sup>4</sup>, Junda Huang<sup>1</sup>, Purva Kulkarni<sup>1,2</sup>, Joeri van der Velde<sup>5</sup>, Casper de Visser<sup>1</sup>, Michael van Vliet<sup>6</sup> and Peter-Bram 't Hoen<sup>1</sup>

<sup>1</sup>Center for Molecular and Biomolecular Informatics, Radboud university medical center, Nijmegen, The Netherlands

<sup>2</sup>Translational Metabolic Laboratory, Radboud university medical center, Nijmegen, The Netherlands

<sup>3</sup>SURF, Amsterdam, The Netherlands

<sup>4</sup>Department of Biological Psychology/Netherlands Twin Register, Faculty of Behavioural and Movement Sciences, Vrije Universiteit Amsterdam, The Netherlands

<sup>5</sup>Genomics Coordination Center and Department of Genetics, University of Groningen and UMCG, The Netherlands

<sup>6</sup>Division of Analytical Biosciences, Leiden Academic Centre for Drug Research, Leiden University, The Netherlands

E-mail: Anna.Niehues@radboudumc.nl, XiaoFeng.Liao@radboudumc.nl, Peter-Bram.tHoen@radboudumc.nl

## 1. Introduction

The FAIR (Findable, Accessible, Interoperable and Reusable) (FAIR) principles were proposed [1] to guide researchers to describe and share their data to increase data reuse and research reproducibility. Creating FAIR data can be challenging for multi-omics researchers due to a lack of tooling and a diverse landscape of (meta)data standards differing across -omics types. Linked data structures and graph representations allow semantic queries and open up new possibilities of data analysis. However, large multi-omics data sets cannot easily be converted to such structures. In X-omics, we develop a FAIR Data Cube (FDCube) [2] – a set of tools and services that help researchers in different stages of the Research Data Life Cycle including creating and describing new data, and finding, understanding and reusing existing FAIR multi-omics data.

## 2. Approach

To facilitate creation of FAIR multi-omics data and metadata, we collaborate with different initiatives such as the FAIR Genomes project [3]. We adopt and develop metadata schemas for different omics data types, and make use of the Investigation-Study-Assay (ISA) metadata framework [4] to capture experimental metadata. Example workflows to create such metadata are publicly shared [5]. Researchers can find and query multi-omics studies via a FAIR Data Point (FDP) instance [6], which links to public or access-protected data repositories. A set of accompanying tools allows the import of general study metadata to the FDP as well as performing semantic queries on additional metadata on samples, phenotypes, or molecular features represented in an RDF-based knowledge graph. In order to allow analysis of access-protected data, we further implement a vantage6-based architecture that allows bioinformaticians to send containerised computing requests to access-controlled omics data storage and receive aggregated results.

## 3. Results

A prototype FDCube implementation is being developed in collaboration with the Trusted World of Corona (TWOC) [7], in which we use public COVID-19 multi-omics data sources to demonstrate the strength and added value of the FDCube and its FAIR-based methodologies. We invite researchers to discuss with us about their own experiences, how the FDCube can facilitate their research, and how X-omics tools can further support them.

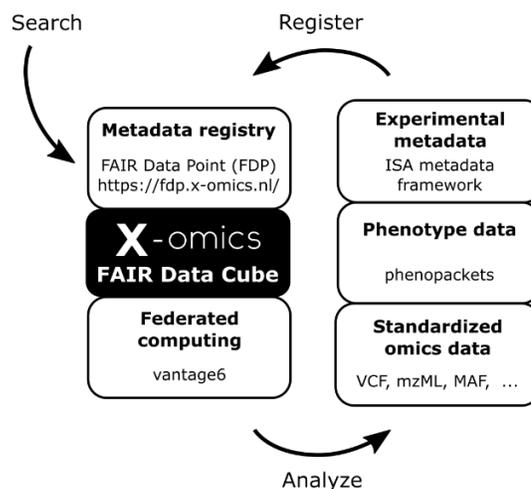


Figure 1. Schematic overview of the FAIR Data Cube architecture.

## References

1. M. Wilkinson, M. Dumontier, I. Aalbersberg et al. “The FAIR Guiding Principles for scientific data management and stewardship”. *Sci Data* 3:160018, 2016.
2. <https://github.com/Xomics/FAIRDataCube>
3. K.J. van der Velde, G. Singh, R. Kaliyaperumal. et al. “FAIR Genomes metadata schema promoting Next Generation Sequencing data reuse in Dutch healthcare and research”. *Sci Data* 9:169, 2022.
4. S.A., Sansone, P. Rocca-Serra, D. Field et al. “Toward interoperable bioscience data”. *Nat Genet* 44:121–126, 2012.
5. <https://github.com/Xomics/ISA-ACTION-Template>
6. <https://fdp.x-omics.nl>, hosted by SURF
7. <https://health-holland.com/project/2020/trusted-world-of-corona>