The Implementation and Demonstration of the X-omics FAIR Data Cube

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

[1]Center for Molecular and Biomolecular Informatics, Radboud UMC, Nijmegen, The Netherlands, [2]Translational Metabolic Laboratory, Radboud UMC, Nijmegen, The Netherlands, [3]SURF, Amsterdam, The Netherlands, [4]Genomics Coordination Center and Department of Genetics, University of Groningen and UMCG, The Netherlands, [5]Division of Analytical Biosciences, Leiden Academic Centre for Drug Research, Leiden University, The Netherlands.

Introduction

The FAIR (Findable, Accessible, Interoperable and Reusable) principles [1] were proposed to guide researchers to describe and share their data to increase data reuse and research reproducibility.
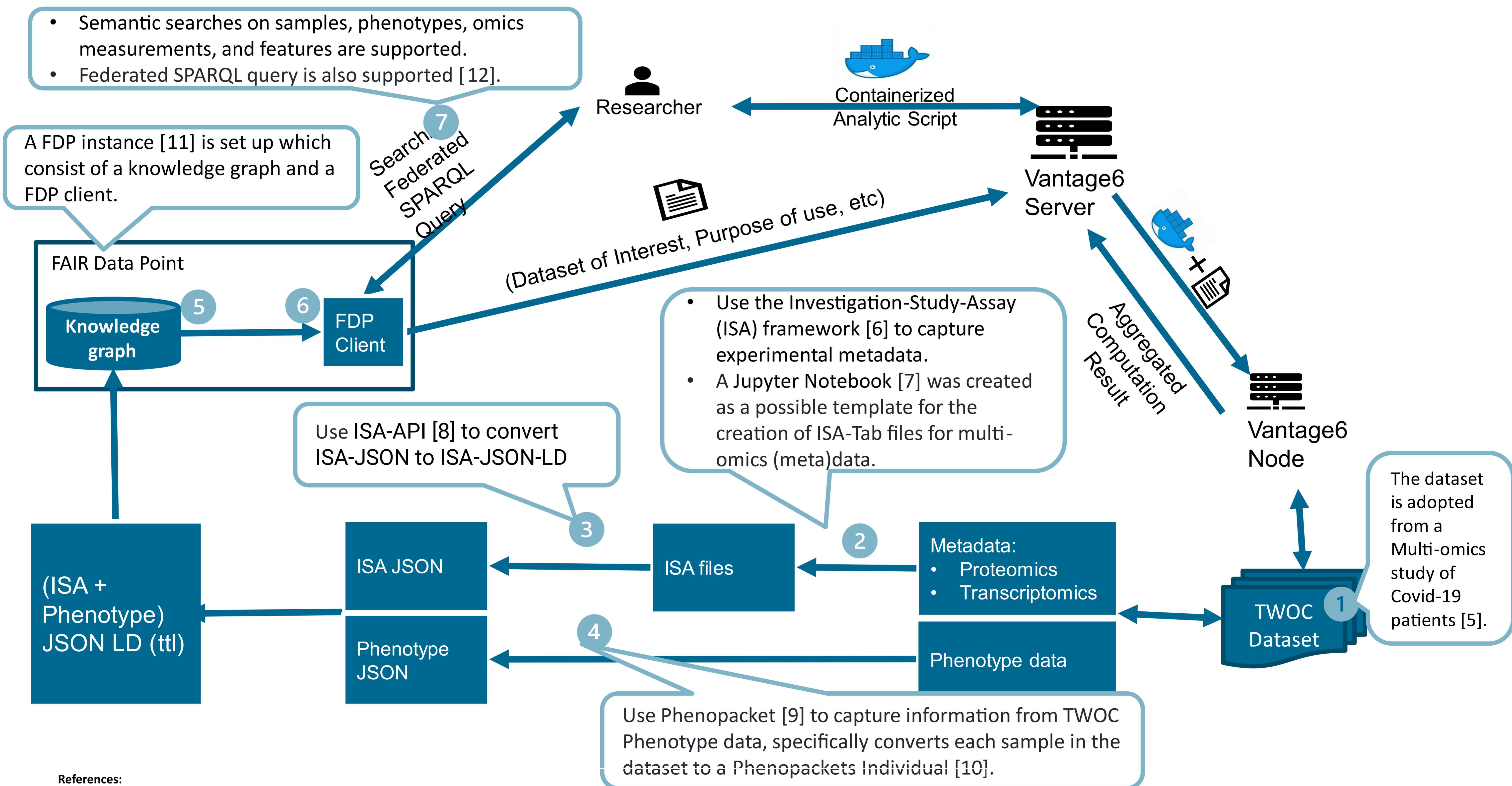
Creating FAIR data can be challenging for multi-omics researchers due to a lack of tooling and a diverse landscape of (meta)data standards differing across -omics types.

In X-omics [2], we develop a FAIR Data Cube [3] – a set of tools and services that help researchers in different stages of the Research Data Life Cycle including Creation and publish of multi-omics and metadata, Querying multi-omics studies, Analyzing access-protected data.
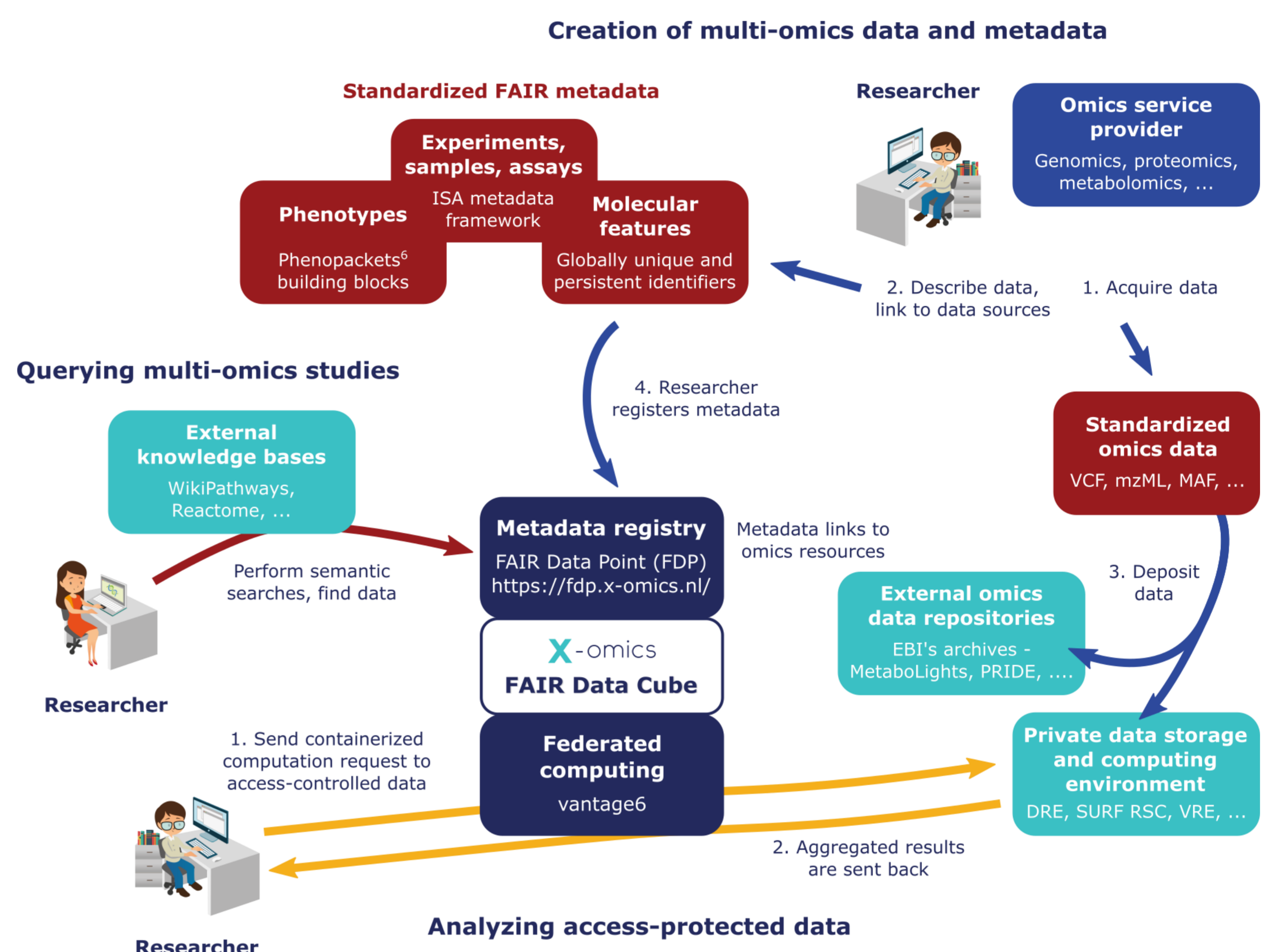
Implementation & Demonstration

The FAIR Data Cube is being developed in collaboration with the Trusted World of Corona (TWOC) [4] for demonstration.

- Semantic searches on samples, phenotypes, omics measurements, and features are supported.
- Federated SPARQL query is also supported [12].

A FDP instance [11] is set up which consist of a knowledge graph and a FDP client.

Use ISA-API [8] to convert ISA-JSON to ISA-JSON-LD

- Use the Investigation-Study-Assay (ISA) framework [6] to capture experimental metadata.
- A Jupyter Notebook [7] was created as a possible template for the creation of ISA-Tab files for multi-omics (meta)data.

The dataset is adopted from a Multi-omics study of Covid-19 patients [5].

Use Phenopacket [9] to capture information from TWOC Phenotype data, specifically converts each sample in the dataset to a Phenopackets Individual [10].

References:
[1]. Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 3, 160018 (2016).
[2]. www.x-omics.nl
[3]. https://github.com/Xomics/FAIRDataCube
[4]. https://www.health-holland.com/project/2020/trusted-world-of-corona
[5]. Su Y, Chen D, Yuan D, et al. Multi-Omics Resolves a Sharp Disease-State Shift between Mild and Moderate COVID-19. Cell. 2020;183(6):1479-1495.e20.
[6]. Sansone, SA., Rocca-Serra, P., Field, D. et al. Toward interoperable bioscience data. Nat Genet 44, 121–126 (2012).
[7] https://github.com/Xomics/ISA-ACTION-Template
[8]. https://github.com/ISA-tools/isa-api
[9]. https://phenopacket-schema.readthedocs.io/en/latest/index.html
[10]. https://github.com/Xomics/TWOCdemonstrator/tree/main/tools
[11]. https://fdp.combi.umcn.nl
[12]. https://fdp.cmbi.umcn.nl/blazegraph/

Contact
XiaoFeng.Liao@radboudumc.nl
Anna.Niehues@radboudumc.nl
Peter-Bram.tHoen@radboudumc.nl

Institute for Molecular Life Sciences
Radboudumc