FAIR Data Cube, a FAIR data infrastructure for integrated multi-omics data analysis

Xiaofeng Liao[1,*], Yuliia Orlova[5], Cenna Doornbos[1], Anna Niehues[1,2], Casper de Visser[1], Junda Huang[1], Thomas H.A. Ederveen[1], Purva Kulkarni[1,2,3], K. Joeri van der Velde[4], Morris A. Swertz[4], Martin Brandt[5], Alain J. van Gool[2,3] and Peter A.C. 't Hoen[1,*]

[1]*Medical BioSciences department, Radboud university medical center, Nijmegen, The Netherlands*

[2]*Translational Metabolic Laboratory, Department of Laboratory Medicine, Radboud university medical center, Nijmegen, The Netherlands*

[3]*Department of Human Genetics, Radboud university medical center, Nijmegen, The Netherlands*

[4]*Genomics Coordination Center, University of Groningen and University Medical Center Groningen, Groningen, The Netherlands*

[5]*SURF, Science Park 140, Amsterdam,The Netherlands*

Abstract

We are witnessing an enormous growth in the amount of molecular profiling (omics) data enabeling the integration of multi-omics data. Nonetheless, this is challenging due to the lack of FAIR -omics data and metadata. The storage of human -omics data in secure silos, for privacy reasons, further complicates their reuse. Federated analysis of FAIR data is a privacy-preserving solution to make optimal use of these multi-omics data and transform them into actionable knowledge.

The Netherlands X-omics Initiative is a National Roadmap Large-Scale Research Infrastructure aiming for efficient integration of data generated within X-omics and external datasets. To facilitate this, we developed the FDCube, which adopts and applies the FAIR principles and helps researchers to create FAIR data and metadata, facilitate reuse of their data, perform federated analysis, and make their data analysis workflows transparent.

Keywords

FAIR, Multi-omics, FAIR Data Cube, Metadata, Federated Analysis

1. Introduction

It is now widely acknowledged that in order to truly advance our understanding of health, it is required to combine -omics data from different sources. Nonetheless, this remains challenging as data and their associated metadata are not always findable, accessible, interoperable, and

SWAT4HCLS 2024: The 15th International Conference on Semantic Web Applications and Tools for Health Care and Life Sciences, February 26–29, 2024, Leiden, The Netherlands

*Corresponding author.
XiaoFeng.Liao@radboudumc.nl (X. Liao); yuliia.orlova@surf.nl (Y. Orlova); Cenna.Doornbos@radboudumc.nl (C. Doornbos); annaniehues@eatris.eu (A. Niehues); Casper.deVisser@radboudumc.nl (C. de Visser); Junda.Huang@radboudumc.nl (J. Huang); Tom.Ederveen@radboudumc.nl (T. H.A. Ederveen); Purva.Kulkarni@radboudumc.nl (P. Kulkarni); joeriv@gmail.com (K. J. van der Velde); m.a.swertz@gmail.com (M. A. Swertz); martin.brandt@surf.nl (M. Brandt); Alain.vanGool@radboudumc.nl (A. J. v. Gool); Peter-Bram.tHoen@radboudumc.nl (Peter A.C. 't Hoen)
0000-0002-4706-1084 (X. Liao)

© 2024 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)

CEUR-WS.org/Vol-3890/paper-44.pdf

CEUR Workshop Proceedings ceur-ws.org ISSN 1613-0073

reusable (FAIR). Furthermore, as most -omics data are derived from a human source, these data are mainly stored in secure and protected data silos. It remains a challenge to re-use these highly secured data sets without the risk of infringing the privacy of the individuals from which the data are derived. Hence, there is a need for tools that enable federated data analysis.

Here we present the X-omics FAIR Data Cube (FDCube). The FDCube helps to make -omics data comply with the FAIR principles and provides a federated data analysis mechanism to bring algorithms to data stations, in order to facilitate data reusing and analysis, while ensuring data privacy.

2. Result

The architecture of FDCube is presented in Figure 1A. The FDCube infrastructure allows dataset owners to register data on the FAIR Data Point (FDP) and it incorporates the FAIR Data Station, a metadata capture platform that facilitates making data FAIR at the source. Using the Investigation-Study-Assay (ISA) metadata schema, metadata is transformed into a FAIR machine-actionable resource stored in an RDF triplestore.

Researchers can exploit the FDCube to find datasets and initiate computation requests to dataset owners. These federated analysis requests are executed on the respective datasets through Vantage6, and the results are communicated back.

We adopted the Trusted World of Corona (TWOC)[1] project as a demonstration to show how to utilize the FDCube for integrated multi-omics federated analysis. TWOC is developing an information platform containing scientific data & information and real world clinical observations on Corona. Figure 1B illustrates the use of FDCube on the TWOC dataset by showing the pipelines covering multiple functionalities.

The FDCube is now listed as a catalog item in the SURF Research Cloud, in which an in-a-box solution is provided to deploy the collection of software applications used by FDCube, including the FAIR Data Point, GraphDB, and FAIR Data Station.

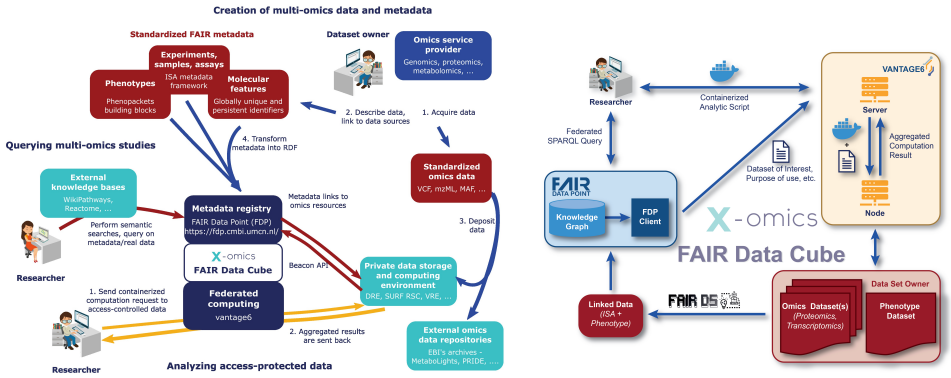

Figure 1: High level architecture of FDCube and its demonstration on TWOC project

[1]https://www.health-holland.com/project/2020/trusted-world-of-corona