# Comparative analysis of data stations in federated health data systems

Table 1 lists known examples of existing health data platform architectures along these two trade-offs.

Table 1: Broad categorization of health data platforms

| | primary | secondary |
|---|---|---|
| **centralized** | openHIE [1], Digizorg, Nordics | kapseli, Mayo, … |
| **decentralized** | RSO Zuid Limburg, Twiin portaal, … | many federated analytics research networks such as x-omics programme and EUCAIM |

## 1 EUCAIM

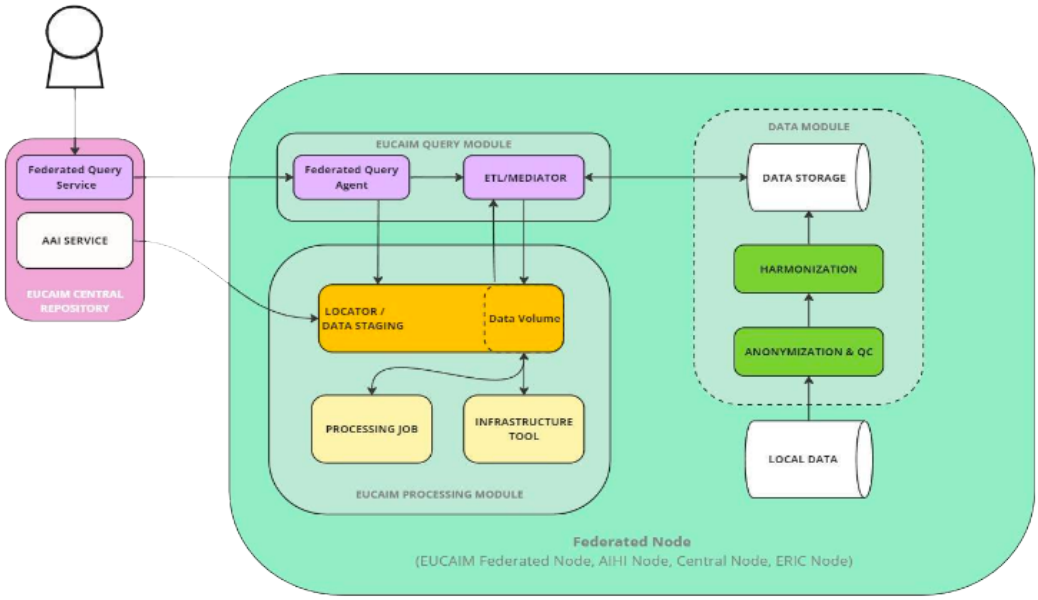



Figure 1: High-level architecture of EUCAIM node

## 2 PLUGIN

The PLUGIN federated learning network is an ongoing initiative initated in 2022 by DHD, IKNL and Expertisecentrum Zorgalgoritmen (EZA) [2]. Its main objective is to realize a federated learning network that includes all 70 hospitals in the Netherlands. The PLUGIN network is intended to support a wide variety of use-cases including:

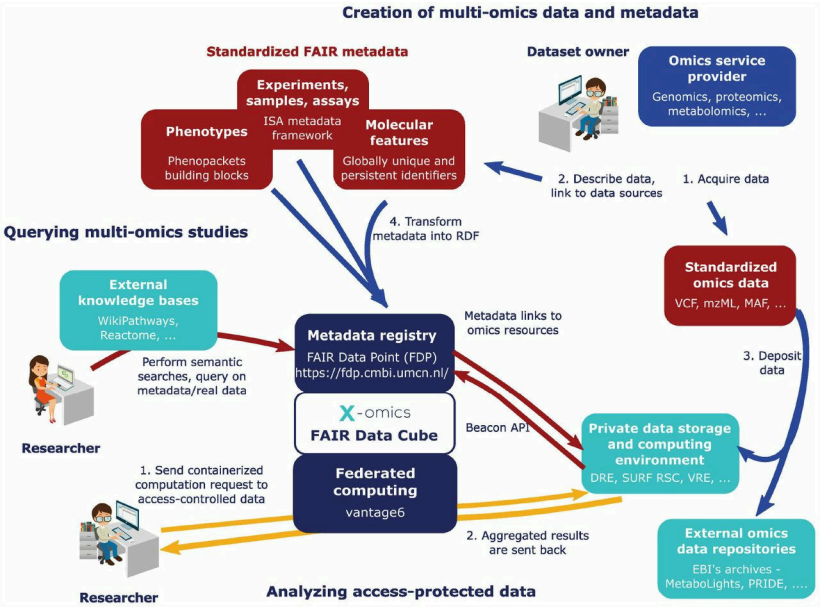
- AI-assisted coding (ICD10) based on supervised learning with language models

- Automated data submission for national registries such as the Dutch Cancer Registry managed by IKNL
- Descriptive analytics, for example, performance analysis across hospitals for benchmarking purpose

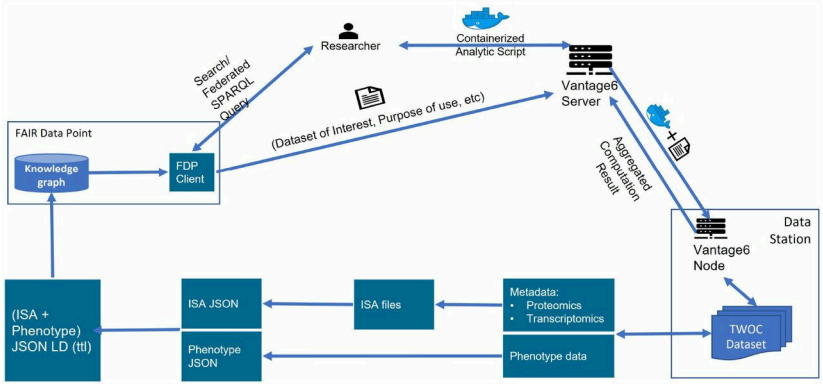
The architecture of

## 3 Fair Data Cube

The Fair Data Cube [3] is a framework for the storage, analysis and integration of multi-omics data. Fair Data Cube reuses and extends existing open software components/ modules and initiatives. This includes the FAIR Data Point [4] and vantage6 [5]. Further elements of the FDCube are the Investigation-Study-Assay (ISA) metadata framework[6, 7] for capturing general study metadata, sample (including basic sample characteristics), and assay metadata, and the Phenopackets [8] standards for capturing phenotypic description of a patient/sample. The concept of the FDCube is illustrated Figure 2.




(a) The concept of the FDCube. More details are available in https://github.com/Xomics/FAIRDataCube/wiki.




(b) Use of FDCube in the TWOC demonstrator. The FDCube workflows covers various functions including creating and publishing metadata, browsing and querying the metadata on FDP, and creating and running federated data analysis.

Figure 2:

Figure 2b shows an example of how Fair Data Cube is used on a public dataset on COVID-19 featuring multi-omics patient data. This dataset was prepared, harmonized and FAIRified as part of the TWOC project. The dataset consists of paired omics data layers describing transcriptomics, proteomics and metabolomics of blood samples, and includes comprehensive phenotype information. Both the ISA metadata schema and Phenopackets schema are adopted. The ISA metadata schema is used as a standard metadata schema to capture metadata about (-omics) experiments, and serializes in an ISA-json file using ISA tools. The ISA tools also provided additional functionalities to convert the ISA objects into linked data, for example a ttl (Terse RDF Triple Language) file. The FAIRified metadata of the TWOC dataset was published on a Fair Data Point portal allowing for querying using SPARQL. After finding an interesting dataset via browsing or

by SPARQL, the researcher could further run follow-up analyses on the target dataset by raising a computation request to the Vantage6 server and retrieve the returning results from the data station via Vantage6.

## 4 Swiss Personal Health Network

The Swiss SPHN network [9] as an example of a data station that uses graph databases both for the data and metadata

## 5 Datastation-as-a-Service in KIK-V

The Datastation-as-a-Service as defined by the Zorginstituut for federated analytics using privacy-enhancing technologies [10]

## 6 Cumuluz data station

[TO DO]

## Bibliography


1. (2024) OpenHIE Framework v5.2-En. https://ohie.org/. Accessed 27 Aug 2024
2. Kapitan D, Heddema F, Dekker A, Sieswerda M, Verhoeff B-J, Berg M (2025) Data Interoperability in Context: The Importance of Open-Source Implementations When Choosing Open Standards. Journal of Medical Internet Research 27(1):e66616. https://doi.org/10.2196/66616
3. Liao X, Ederveen T, Niehues A, et al (2024) FAIR Data Cube, a FAIR Data Infrastructure for Integrated Multi-Omics Data Analysis. Journal of Biomedical Semantics 15(1):20. https://doi.org/10.1186/s13326-024-00321-2
4. Silva Santos LOB da, Burger K, Kaliyaperumal R, Wilkinson MD (2023) FAIR Data Point: A FAIR-Oriented Approach for Metadata Publication. Data Intelligence 5(1):163–183. https://doi.org/10.1162/dint_a_00160
5. Moncada-Torres A, Martin F, Sieswerda M, Van Soest J, Geleijnse G (2021) VANTAGE6: An Open Source priVAcy preserviNg federaTed leArninG infrastructurE for Secure Insight eXchange. AMIA Annual Symposium Proceedings 2020:870–877
6. Sansone S-A, Rocca-Serra P, Field D, et al (2012) Toward Interoperable Bioscience Data. Nature Genetics 44(2):121–126. https://doi.org/10.1038/ng.1054
7. Johnson D, Batista D, Cochrane K, et al (2021) ISA API: An Open Platform for Interoperable Life Science Experimental Metadata. GigaScience 10(9):giab60. https://doi.org/10.1093/gigascience/giab060
8. Ladewig MS, Jacobsen JOB, Wagner AH, et al (2023) GA4GH Phenopackets: A Practical Introduction. Advanced Genetics 4(1):2200016. https://doi.org/10.1002/ggn2.202200016

9. SPHN - Swiss Personalized Health Network (SPHN). https://sphn.ch/. Accessed 9 Jun 2025
10. (2024) KIK-V x GERDA