

Comparative analysis of data stations in federated health data systems

1 EUCAIM

2 PLUGIN

The PLUGIN federated learning network is an ongoing initiative initiated in 2022 by DHD, IKNL and Expertisecentrum Zorgalgoritmen (EZA) [1]. 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 [2] 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 [3] and vantage6 [4]. Further elements of the FDCube are the Investigation-Study-Assay (ISA) metadata framework[5, 6] for capturing general study metadata, sample (including basic sample characteristics), and assay metadata, and the Phenopackets [7] standards for capturing phenotypic description of a patient/sample. The concept of the FDCube is illustrated Figure 1.

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 [8] 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 [9]

6 Cumuluz data station

[TO DO]

Bibliography

1. 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>
2. 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>
3. 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
4. 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
5. 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>
6. 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>
7. 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>
8. SPHN - Swiss Personalized Health Network (SPHN). <https://sphn.ch/>. Accessed 9 Jun 2025
9. (2024) KIK-V x GERDA