

Data inteoperability in context: the importance of open source implementations when choosing open standards

Daniel Kapitan^{a,b,c}, Femke Heddema^b, Andre Dekker^d, Melle Sieswerda^{e,d},
Bart-Jan Verhoeff^f

^a*Dutch Hospital Data, Utrecht, the Netherlands,*

^b*PharmAccess Foundation, Amsterdam, the Netherlands,*

^c*Eindhoven University of Technology, Eindhoven, the Netherlands,*

^d*Maastricht University, Maastricht, the Netherlands,*

^e*Integral Cancer Registry Netherlands, Eindhoven, the Netherlands,*

^f*Expertisecentrum Zorgalgoritmen, Utrecht, the Netherlands,*

Abstract

In response to the proposal of Tsafnat et al. to converge towards three open health data standards, this viewpoint provides a critical reflection on the proposed alignment of using OpenEHR, FHIR and OMOP as the default standards for clinical care and administration, data exchange and longitudinal analysis, respectively. We argue that open standards are a necessary but not sufficient condition to achieve health data interoperability. The ecosystem of open source implementations needs to be considered when choosing an appropriate standard for a given context. We discuss two specific contexts, namely standardization of i) health data for federated learning, and ii) health data sharing in low- and middle income countries (LMICs). Specific design principles, practical considerations and implementation choices for these two context are described, based on ongoing work in both areas. In the case of federated learning, we observe convergence towards OMOP and FHIR, where the two standards can effectively be used side-by-side given the availability of mediators between the two. In the case of health information exchanges in LMICs, we see a strong convergence towards FHIR as the primary standard, with as yet limited adoption of OMOP and OpenEHR. We propose practical guidelines for context-specific adaptation of open standards.

Keywords: OMOP, OpenEHR, FHIR, secondary use, data platform, digital platform

*Corresponding author

Email address: daniel@kapitan.net (Daniel Kapitan)

1. Open standards are a necessary but not sufficient condition for interoperability

“A paradox of health care interoperability is the existence of a large number of standards exists with significant overlap among them,” say Tsafnat et al., followed by a call to actions towards the health informatics community to put effort into establishing convergence and preventing collision (Tsafnat et al., 2024). To do so, they propose to converge on three open standards, namely i) OpenEHR for clinical care and administration; ii) Fast Health Interoperability Resources (FHIR) for data exchange and iii) Observational Medical Outcomes Partnership Common Data Model (OMOP) for longitudinal analysis. They argue that open data standards, backed by engaged communities, hold an advantage over proprietary ones and therefore should be chosen as the steppingstones towards achieving true interoperability.

While we support their high-level rationale and intention, we feel their proposed trichotomy does not do justice to details that are crucial in real-world implementations. This viewpoint provides a critical reflection on their proposed framework in three parts. First, we reflect on salient differences between the three open standards from the perspective of the notion of openness of digital platforms (de Reuver et al., 2018) and the paradox of open (Keller and Tarkowski, 2021). Subsequently, we present our findings in designing and implementing health data platforms in two specific contexts, namely i) platforms for federated learning on shared health data in high income countries; and ii) health data platforms for low and middle income countries (LMICs). We conclude with practical guidelines for context-specific adaptation of open standards.

2. Digital platforms require extensibility, availability of complementary components and availability of executable pieces of software

Besides the paradox of interoperability put forward by Tsafnat et al., we argue that open standards are a necessary, but not sufficient condition for convergence of health data standarization. We posit that open source implementations of components, libraries etc. constitute another necessary condition for establishing a flourishing health data sharing platform and associated ecosystem for any given context, be it regional, international or within a specific sub-domain like pandemic preparedness. Research on digital platforms underline the importance of the platform openness, not only in term of open standards, but also in term of extensibility of the code base, availability of complements to the core technical platform (in our case the data standard itself) and availability of executable pieces of software (de Reuver et al., 2018). Only when the majority of these aspects of digital platforms are fulfilled can we reasonably expect that the platform will indeed be longlived.

In what they call the paradox of open, Keller and Tarkowski argue that this conventional approach of open standards and open source flourish under two types of conditions (Keller and Tarkowski, 2021). First, projects where many people

contribute to the creation of a common resource have proven succesful. “This is the story of Wikipedia, OpenStreetMap, Blender.org, and the countless free software projects that provide much of the internet’s infrastructure.” (Keller and Tarkowski, 2021) Indeed, Tsafnat et al. have explicitly taken into account that “an engaged and vibrant community is a major advantage for the longevity of the data standards it uses,” which has informed their proposal to converge towards OMOP, FHIR en OpenEHR. However, the emphasis on open source implementations is somewhat overlooked. This point is only mentioned in passing and indirectly, when Tsafnat et al. reference work done by Reynolds and Wyatt who already argued in 2011 “... for the superiority of open source licensing to promote safer, more effective health care information systems. We claim that open source licensing in health care information systems is essential to rational procurement strategy” (Reynolds and Wyatt, 2011). We believe that a realistic assessment of the current position of an open standard within the wider context of availability of complementary components and open source implementations is equally important when choosing which standard to adopt.

This point is related to the second condition put forward by Keller and Tarkowski, namely that the conventional open approach has proven fruitful when “opening up” is the result of external incentives or requirements, rather than voluntary actions. “This is the story of publicly-funded knowledge production like Open Access academic publications, cultural heritage collections in the Public Domain, Open Educational Resources (OER), and Open Government data.” (Keller and Tarkowski, 2021) A canonical example is the birth of the GSM standard, which was mandated by European legislation.¹ Reflecting on this perspective on openness, we observe a salient difference between FHIR vis-a-vis OpenEHR and OMOP, namely that the former is the only one that has been mandated (or at least strongly recommended) in some jurisdictions. In the US, the Office of the National Coordinator for Health Information Technology (ONC) and the Centers for Medicare and Medicaid Services (CMS) has introduced a steady stream of new regulations, criteria, and deadlines in Health IT that has resulted in significant adoption of FHIR (Firely, 2023). In India, the open Health Claims Exchange protocol specification - which is based on FHIR - has been mandated by the Indian government as the standard for e-claims handling (ind, 2020; hcx, 2023). The African Union recommends all new implementations and digital health system improvements use FHIR as the primary mechanism for data exchange (Tilahun et al., 2023), but doesn’t say anything about the use of, for example, OpenEHR for administrative point-of-service systems.

These external incentives have resulted in a large boost in both commercial and open source development activities in the FHIR ecosystem. Illustrative of this is the speed with which the Bulk FHIR API has been defined and implemented in almost all major implementations (Mandl et al., 2020; Jones et al.,

¹See https://en.wikipedia.org/wiki/GSM#Initial_European_development for details.

2021), and the the SQL-on-FHIR specification to make large-scale analysis of FHIR data accessible to a larger audience and portable between systems.² It has also led to more people voluntarily contributing to FHIR-related open source projects, which has resulted in a wide offering of FHIR components across major technology stacks (Java, Python, .NET), thereby strengthening the first condition. By comparison, OMOP and OpenEHR have not yet profited from external incentives to spur the adoption and thereby growing the ecosystem beyond a certain critical mass. To illustrate this, a search on GitHub on “FHIR” yields 8.2 thousand results, “OMOP or OHDSI” one thousand results, and “OpenEHR” returns 400 results. A quick-scan of the available open source components listed on the website of the three governing bodies HL7, OHDSI and OpenEHR, indicates that the ecosystem of FHIR and OMOP have a significantly larger offering of extensible and complementary open source components than OpenEHR.³

Hence, we stress that beyond the evaluating the intrinsic structure of an open standard and the community that supports the standard, we need to take into account the wider ecosystem of open source implementations, complementary components etc. From this wider perspective of the whole ecosystem surrounding the three standards, FHIR stands out as having the most diverse and rich ecosystem because it has been mandated in certain jurisdictions. This is relevant when comparing these standards in real-world implementations. We now turn to two specific use cases c.q. contexts where these considerations are at play.

3. Standarization of health data for federated learning

The current fragmentation in health data is one of the major barriers towards leveraging the potential medical data for machine learning (ML). Without access to sufficient data, ML will be limited in its application to health improvement efforts and, ultimately, from making the transition from research to clinical practice. High quality health data, obtained from a research setting or a real-world clinical practice setting, is hard to obtain, because health data is highly sensitive and its usage is tightly regulated.

Federated learning (FL) is a learning paradigm that aims to address these issues of data governance and privacy by training algorithms collaboratively without moving (copying) the data itself (Rieke et al., 2020; Teo et al., 2024). Based on ongoing work with the PLUGIN healthcare consortium (<https://plugin.healthcare>, in Dutch) we have detailed an architecture for FL for secondary use of health data for hospitals in the Netherlands. Starting point for this implementation are the National Health Data Infrastructure agreements

²<https://build.fhir.org/ig/FHIR/sql-on-fhir-v2/>

³FHIR: <https://confluence.hl7.org/display/FHIR/Open+Source+Implementations>
 OMOP: <https://www.ohdsi.org/software-tools/>
 OpenEHR: https://openehr.org/products_tools/platform/

for research, policy and innovation for the Dutch healthcare sector, which have been adopted at the beginning of 2024 (Health-RI, 2024). Figure 1 shows a high level overview of the platform, which comprises three areas (multiple use, applications and generic features) and a total of 26 functional components (for details please refer to (Health-RI, 2024)). One of the prerequisites of this architecture is that organizations that participate in a federation of ‘data stations’ use the same common data model to make the data Findable, Accessible, Interoperable and Resusable (FAIR). These FAIR data stations comprise components 7, 8 and 9 in Figure 1, i.e. the data, metadata and APIs, respectively, through which this the data station can be accessed and used.

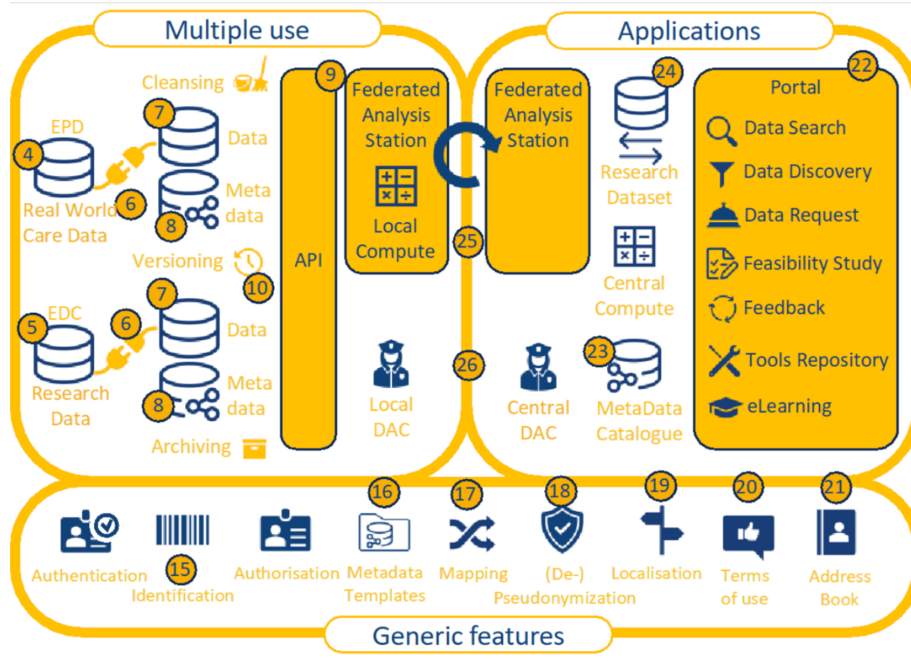


Figure 1: Reference architecture for the Dutch health data infrastructure for research and innovation (Health-RI, 2024)

Following the line of reasoning of Tsafnat et al., OMOP would be the go-to standard for storing the longitudinal data in each of the data stations. Indeed, by now there are quite a few reports of real-world implementations of federated learning networks based on the OHDSI-OMOP stack, including a global infrastructure with 22 centres for COVID19 prediction models (Khalid et al., 2021), FeederNet in South Korea with 57 participating hospitals (Lee et al., 2022), Dutch multi-cohort dementia research with 9 centres (Mateus et al., 2024), the European severe heterogeneous asthma research collaboration (Kroes et al., 2022) and the recently initiated Belgian FEIN network [TO DO: add reference].

For the PLUGIN project, however, we choose to adopt FHIR as the com-

mon data model because of its practicality and extensibility to be used in a Python-based data science stack, provenance of RESTful APIs out-of-the-box to facilitate easy integration with the container-based vantage6 FL framework, support of many healthcare terminologies and flexibility through the profiling mechanisms (Choudhury et al., 2020; Smits et al., 2022). Increasingly, other projects have reported the use of FHIR for persistent, longitudinal storage for FL. The CODA platform, which aims to implement a similar FL infrastructure in Canada, compared OMOP and FHIR and chose the latter as it has been found to support more granular mappings required for analytics (Mullie et al., 2023). The fair4health project has implemented also based on FHIR, using their own open source framework for the federated learning infrastructure itself (Sinaci et al., 2024).

Given that conceptually OMOP can be viewed as a strict subset of FHIR, hybrid solutions using OMOP and FHIR combined have also been reported, such as the German KETOS platform (Gruendner et al., 2019), and the preliminary findings from the European GenoMed4All project which aims to connect clinical and -omics data (Cremonesi et al., 2023). A collaboration of 10 university hospitals in Germany have shown that standardized ETL-processing from FHIR into OMOP can achieve 99% conformance (Peng et al., 2023), which confirms the feasibility of the solution pattern where FHIR acts as an intermediate sharing standard through which data from (legacy) systems are extracted and made available for reuse in a common data model. One could argue that the distinction between FHIR and OMOP becomes less relevant if data can be effectively stored in either standard. We are hopeful that initiatives like <https://omoponfhir.org> indeed will foster convergence rather than collision between these two standards.

In the case of PLUGIN, however, we had other considerations to choose FHIR as the persistent storage format of the data stations. One of the important considerations is that we found that the mechanism of FHIR Profiles can be tied to concept of late binding commonly applied in data lake/warehouse architectures: allow ingest of widely different sources, and gradually add more constraints and validations as you move closer to a specific use case. If machine learning is the primary objective for secondary use, we want to be able to cast a wider net of relevant data, rather than being too restrictive when ingesting the data at the start of processing pipeline. Late binding in data warehousing is a design philosophy where data transformation and schema enforcement are deferred until as late as possible in the data processing pipeline, sometimes even until query time. This approach contrasts with early binding, where data is transformed and structured as it is ingested into the data warehouse. This principle is visualized as concentric circles (Figure 2). The advantages of this design is that it allows for greater flexibility. During the initial ingest of the data, we only require the data to conform to the minimal syntactic standard defined by the base FHIR version (R4 in the diagram). As the data is processed, more strict checks and constraints are applied, whereby ultimately different profiles can co-exist next to one another (the two most inner circles), within a larger circle with fewer strictions. This approach does not support the extension mechanism

of FHIR, so we need to be cautious if we decide to use that.

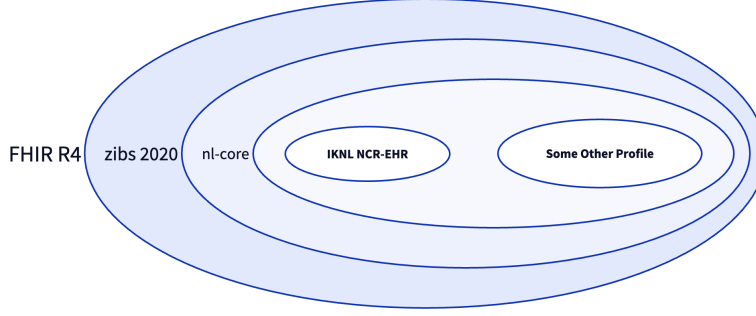


Figure 2: Principle of late binding with FHIR profiling mechanism

We found that this principle of late binding also allows flexible and efficient implementations of the data stations that make use of the current best practices of the a lakehouse architecture of (Hai et al., 2023; Harby and Zulkernine, 2022, 2024) and the composable data stack (Pedreira et al., 2023). Lakehouses typically have a zonal architecture that follow the Extract-Load-Transform pattern (ELT) where data is ingested from the source systems in bulk (E), delivered to storage with aligned schemas (L) and transformed into a format ready for analysis (T) (Hai et al., 2023). The discerning characteristic of the lakehouse architecture is its foundation on low-cost and directly-accessible storage that also provides traditional database management and performance features such as ACID transactions, data versioning, auditing, indexing, caching, and query optimization (Armbrust et al., 2021). Lakehouses thus combine the key benefits of data lakes and data warehouses: low-cost storage in an open format accessible by a variety of systems from the former, and powerful management and optimization features from the latter.

The main disadvantage in using FHIR in this way pertains to the need for upgrading the whole ELT pipeline when upgrading to a new primary FHIR version, for example R5. However, we expect that the development time required to upgrade FHIR versions is significantly less than the initial migration to FHIR.

[TO DO: add more arguments here, finish this section with conclusion/summary for this use-case]

4. Health data standards in low- and middle income countries

It is a widely held belief that digital technologies have an important role to play in strengthening health systems in LMICs. Yet, also here the current fragmen-

tation of health data stands in the way of scaling up digital health programme beyond project-centric, vertical solutions into sustainable health information exchanges. And also here, many have called to converge towards open standards (Mehl et al., 2023). Based on our direct involvement in implementing and designing health data platforms in sub-Saharan African countries, and in line with the approach put forward by Mehl et al., we emphasize the need for convergence not only on i) open standards, but also on ii) open technologies (similar to our arguments discussed in the above), iii) open architectures (documentation, using open standards, of reusable enterprise architecture patterns for health systems) and iv) open content (representations of public health, health system or clinical knowledge to guide implementations).

Many sub-Saharan African countries have adopted the OpenHIE framework (ope, 2022) as the architectural blueprint for implementing nation-wide health information exchanges (HIE) (Mamuye et al., 2022), including Nigeria (Dahatu et al., 2023), Kenya (Thaiya et al., 2021) and Tanzania (Nsaghurwe et al., 2021). While the OpenHIE specification is agnostic to which data standard should be used, in practice the digital health community in the global South have *de facto* converged towards FHIR as the common data model, not only for data exchange, but also for point-of-service systems to support clinical care and administration, and persistent, longitudinal storage of data in the so-called Shared Health Record (ohi, 2023). The adoption of FHIR as the default standard for all three domains is fueled by the widespread availability of open-access software infrastructure which enables an end-to-end integration. To illustrate this, consider for example, the Open Health Stack (OHS) consists of a new suite of digital public goods, including a software development kit for building FHIR-native apps on Android, and analytics tooling to generate insights from FHIR data. This solution design blurs the distinction between the three original domains for health data standardization.

Table 1: Number of healthcare facilities in Kenya. Source: Kenya Health Facility Census, Ministry of Health, September 2023.

| Level | Description | Number of facilities |
|-------|--|----------------------|
| 2 | Dispensaries and private clinics, typically located in a school, industrial plant or other organization that dispenses medication and sometimes basic medical and dental treatment | 8,806 |
| 3 | Health centres, medium-sized units which cater for a population of about 80,000 people | 2,559 |
| 4 | Sub-county hospital, similar to health centres with additional facilities for more complex procedures | 971 |

| Level | Description | Number of facilities |
|-------|---|----------------------|
| 5 | County referral hospital, regional centres which provide specialised care | 34 |
| 6 | National referral hospital | 5 |

Arguments to add:

- explain that we don't see OpenEHR playing any role of significance. Largest open source EHR implementations working with own data models. Unlikely this will change any time soon ([Syzdykova et al., 2017](#))
- instead, we see that FHIR is in fact being adopted as a point-of-service system. this is possible because data is of lower complexity and the majority of smaller health facilities are resource constrained anyway → moving towards a scenario where an app used by the health care worker is the system of record. Here we see a convergence of the three domains put forward by Tsafnat: the Shared Health Record serves as the back-end of the system-of-record, it provides a transactional, persistent storage engine for enabling information exchange and it is the longitudinal data that is used downstream for analytics, monitoring etc.
- explain example OHS/OpenSRP: maternal care
- – explain there is lot of complementary open technologies and open content, to get going
- – and SDK to build apps
- – Clinical workflows which can be configured in a FHIR-native front-and backend
- – InstantHIE to deploy a swarm of containers to provision all relevant components
- – leverage open source data & analytics components to build data stations/data warehouses at any scale
- also technical advantages of FHIR
 - explain that because FHIR is based on web technologies, it lends it self serverless implementations, separation of storage and compute(ref composable data stack); and downward scalability (running in the browser)
 - how does this relate to ontologies/graph-based? FHIR can be expressed in graphs ([Gebreslassie et al., 2023](#))]
- main disadvantages
 - FHIR profiling and versions. Still lot of work to be done prevent sprawl of profiles. Many LMICs still lack a standardized, national core profile. Initiatives such as the International Patient Summary can be used as a stepping stone to provide arrive at a consistent profile for high-volume, low-complexity care such as maternal and antenatal care (ANC) and non-communicable diseases (NCD). In our experience, deployment of these HIEs have a strong focus on enabling

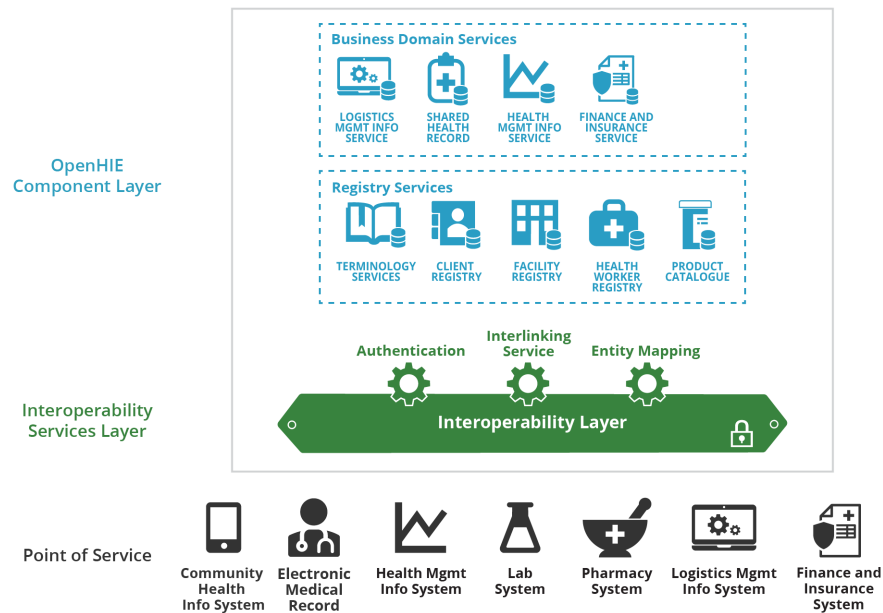
primary care networks, where it makes economic sense to combine the SHR as a shared system-of-record of such a network.

5. Conclusion and outlook

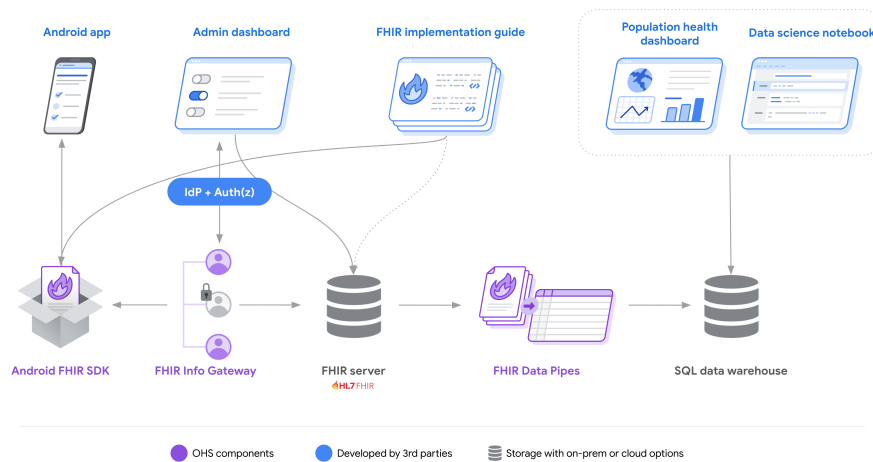
- We underline the need for open data standards as a necessary condition to achieve interoperability
- It is not a sufficient condition
- Suggestion: can we make OMOP and OpenEHR benefit from the open technologies on FHIR? We think that through making open source reference implementations, such as OMOP-on-FHIR (what's EHR equivalent of that?) we can
- MPC as next step up from FL
- More than a decade later, we observe that only a very small fraction of health IT systems are based on open source, the majority of which are used in LMICs which we will discuss later ([dig, 2024](#)).

References

- , 2020. National Digital Health Mission. Technical Report. India National Health Authority.
- , 2022. OpenHIE Framework v5.0. <https://ohie.org/>.
- , 2023. HCX Protocol v0.9.
- , 2023. OHIE23 Unconference 101.
- , 2024. Digital Public Goods Alliance. <https://digitalpublicgoods.net/>.
- Armbrust, M., Ghodsi, A., Xin, R., Zaharia, M., 2021. Lakehouse: A New Generation of Open Platforms that Unify Data Warehousing and Advanced Analytics, in: 11th Annual Conference on Innovative Data Systems Research (CIDR '21), p. 8.
- Choudhury, A., van Soest, J., Nayak, S., Dekker, A., 2020. Personal Health Train on FHIR: A Privacy Preserving Federated Approach for Analyzing FAIR Data in Healthcare, in: Bhattacharjee, A., Borgohain, S.K., Soni, B., Verma, G., Gao, X.Z. (Eds.), Machine Learning, Image Processing, Network Security and Data Sciences, Springer, Singapore. pp. 85–95. doi:[10.1007/978-981-15-6315-7_7](https://doi.org/10.1007/978-981-15-6315-7_7).
- Cremonesi, F., Planat, V., Kalokyri, V., Kondylakis, H., Sanavia, T., Miguel Mateos Resinas, V., Singh, B., Uribe, S., 2023. The need for multimodal health data modeling: A practical approach for a federated-learning healthcare platform. Journal of Biomedical Informatics 141, 104338. doi:[10.1016/j.jbi.2023.104338](https://doi.org/10.1016/j.jbi.2023.104338).



(a) a)



(a) b)

Figure 4: The OpenHIE reference architecture (a) and themain components in the Open Health Stack (b).

- Dalhatu, I., Aniekwe, C., Bashorun, A., Abdulkadir, A., Dirlikov, E., Ohakanu, S., Adedokun, O., Oladipo, A., Jahun, I., Murie, L., Yoon, S., Abdu-Aguye, M.G., Sylvanus, A., Indyer, S., Abbas, I., Bello, M., Nalda, N., Alagi, M., Odafe, S., Adebajo, S., Ogorry, O., Akpu, M., Okoye, I., Kakanfo, K., Onovo, A.A., Ashefor, G., Nzelu, C., Ikpeazu, A., Aliyu, G., Ellerbrock, T., Boyd, M., Stafford, K.A., Swaminathan, M., 2023. From Paper Files to Web-Based Application for Data-Driven Monitoring of HIV Programs: Nigeria’s Journey to a National Data Repository for Decision-Making and Patient Care. *Methods of Information in Medicine* 62, 130–139. doi:[10.1055/s-0043-1768711](https://doi.org/10.1055/s-0043-1768711).
- de Reuver, M., Sørensen, C., Basole, R.C., 2018. The Digital Platform: A Research Agenda. *Journal of Information Technology* 33, 124–135. doi:[10.1057/s41265-016-0033-3](https://doi.org/10.1057/s41265-016-0033-3).
- Firely, 2023. FHIR in US Healthcare Regulations. Technical Report.
- Gebreslassie, T.G., van Reisen, M., Amare, S.Y., Taye, G.T., Plug, R., 2023. FHIR4FAIR: Leveraging FHIR in health data FAIRfication process: In the case of VODAN-A. *FAIR Connect* 1, 49–54. doi:[10.3233/FC-230504](https://doi.org/10.3233/FC-230504).
- Gruendner, J., Schwachhofer, T., Sippl, P., Wolf, N., Erpenbeck, M., Gulden, C., Kapsner, L.A., Zierk, J., Mate, S., Stürzl, M., Croner, R., Prokosch, H.U., Toddenroth, D., 2019. KETOS: Clinical decision support and machine learning as a service – A training and deployment platform based on Docker, OMOP-CDM, and FHIR Web Services. *PLOS ONE* 14, e0223010. doi:[10.1371/journal.pone.0223010](https://doi.org/10.1371/journal.pone.0223010).
- Hai, R., Koutras, C., Quix, C., Jarke, M., 2023. Data Lakes: A Survey of Functions and Systems. *IEEE Transactions on Knowledge and Data Engineering* 35, 12571–12590. doi:[10.1109/TKDE.2023.3270101](https://doi.org/10.1109/TKDE.2023.3270101).
- Harby, A.A., Zulkernine, F., 2022. From Data Warehouse to Lakehouse: A Comparative Review, in: 2022 IEEE International Conference on Big Data (Big Data), IEEE, Osaka, Japan. pp. 389–395. doi:[10.1109/BigData55660.2022.10020719](https://doi.org/10.1109/BigData55660.2022.10020719).
- Harby, A.A., Zulkernine, F., 2024. Data Lakehouse: A Survey and Experimental Study. doi:[10.2139/ssrn.4765588](https://doi.org/10.2139/ssrn.4765588).
- Health-RI, 2024. Agreements on the National Health Data Infrastructure for Research, Policy and Innovation - Health-RI Nationale Gezondheidsdata-infrastructuur - Confluence. <https://health-ri.atlassian.net/wiki/spaces/HNG/pages/249073646/Agreements+on+the+National+Health+Data+Infrastructuur>
- Jones, J., Gottlieb, D., Mandel, J.C., Ignatov, V., Ellis, A., Kubick, W., Mandl, K.D., 2021. A landscape survey of planned SMART/HL7 bulk FHIR data access API implementations and tools. *Journal of the American Medical Informatics Association* 28, 1284–1287. doi:[10.1093/jamia/ocab028](https://doi.org/10.1093/jamia/ocab028).

- Keller, P., Tarkowski, A., 2021. The Paradox of Open. Open Future .
- Khalid, S., Yang, C., Blacketer, C., Duarte-Salles, T., Fernández-Bertolín, S., Kim, C., Park, R.W., Park, J., Schuemie, M.J., Sena, A.G., Suchard, M.A., You, S.C., Rijnbeek, P.R., Reips, J.M., 2021. A standardized analytics pipeline for reliable and rapid development and validation of prediction models using observational health data. *Computer Methods and Programs in Biomedicine* 211, 106394. doi:[10.1016/j.cmpb.2021.106394](https://doi.org/10.1016/j.cmpb.2021.106394).
- Kroes, J.A., Bansal, A.T., Berret, E., Christian, N., Kremer, A., Alloni, A., Gabetta, M., Marshall, C., Wagers, S., Djukanovic, R., Porsbjerg, C., Hamerlijnck, D., Fulton, O., ten Brinke, A., Bel, E.H., Sont, J.K., 2022. Blueprint for harmonising unstandardised disease registries to allow federated data analysis: Prepare for the future. *ERJ Open Research* 8. doi:[10.1183/23120541.00168-2022](https://doi.org/10.1183/23120541.00168-2022).
- Lee, S., Kim, C., Chang, J., Park, R.W., 2022. FeederNet (Federated E-Health Big Data for Evidence Renovation Network) platform in Korea – OHDSI.
- Mamuye, A.L., Yilma, T.M., Abdulwahab, A., Broomhead, S., Zondo, P., Kyeng, M., Maeda, J., Abdulaziz, M., Wuhib, T., Tilahun, B.C., 2022. Health information exchange policy and standards for digital health systems in africa: A systematic review. *PLOS Digital Health* 1, e0000118. doi:[10.1371/journal.pdig.0000118](https://doi.org/10.1371/journal.pdig.0000118).
- Mandl, K.D., Gottlieb, D., Mandel, J.C., Ignatov, V., Sayeed, R., Grieve, G., Jones, J., Ellis, A., Culbertson, A., 2020. Push Button Population Health: The SMART/HL7 FHIR Bulk Data Access Application Programming Interface. *npj Digital Medicine* 3, 1–9. doi:[10.1038/s41746-020-00358-4](https://doi.org/10.1038/s41746-020-00358-4).
- Mateus, P., Moonen, J., Beran, M., Jaarsma, E., van der Landen, S.M., Heuvelink, J., Birhanu, M., Harms, A.G.J., Bron, E., Wolters, F.J., Cats, D., Mei, H., Oomens, J., Jansen, W., Schram, M.T., Dekker, A., Bermejo, I., 2024. Data harmonization and federated learning for multi-cohort dementia research using the OMOP common data model: A Netherlands consortium of dementia cohorts case study. *Journal of Biomedical Informatics* 155, 104661. doi:[10.1016/j.jbi.2024.104661](https://doi.org/10.1016/j.jbi.2024.104661).
- Mehl, G.L., Seneviratne, M.G., Berg, M.L., Bidani, S., Distler, R.L., Gorgens, M., Kallander, K.E., Labrique, A.B., Landry, M.S., Leitner, C., Lubell-Doughtie, P.B., Marcelo, A.D., Matias, Y., Nelson, J., Nguyen, V., Nsengimana, J.P., Orton, M., Oztzy Garcia, D.R., Oyaole, D.R., Ratanaprayul, N., Roth, S., Schaefer, M.P., Settle, D., Tang, J., Tien-Wahser, B., Wanyee, S., Hersch, F., 2023. A full-STAC remedy for global digital health transformation: Open standards, technologies, architectures and content. *Oxford Open Digital Health* 1, oqad018. doi:[10.1093/oodh/oqad018](https://doi.org/10.1093/oodh/oqad018).

- Mullie, L., Afilalo, J., Archambault, P., Bouchakri, R., Brown, K., Buckeridge, D.L., Cavayas, Y.A., Turgeon, A.F., Martineau, D., Lamontagne, F., Lebrasseur, M., Lemieux, R., Li, J., Sauthier, M., St-Onge, P., Tang, A., Witteman, W., Chassé, M., 2023. CODA: An open-source platform for federated analysis and machine learning on distributed healthcare data. *Journal of the American Medical Informatics Association*, ocad235doi:[10.1093/jamia/ocad235](https://doi.org/10.1093/jamia/ocad235).
- Nsaghurwe, A., Dwivedi, V., Ndesanjo, W., Bamsi, H., Busiga, M., Nyella, E., Massawe, J.V., Smith, D., Onyejekwe, K., Metzger, J., Taylor, P., 2021. One country’s journey to interoperability: Tanzania’s experience developing and implementing a national health information exchange. *BMC Medical Informatics and Decision Making* 21, 139. doi:[10.1186/s12911-021-01499-6](https://doi.org/10.1186/s12911-021-01499-6).
- Pedreira, P., Erling, O., Karanasos, K., Schneider, S., McKinney, W., Valluri, S.R., Zait, M., Nadeau, J., 2023. The Composable Data Management System Manifesto. *Proceedings of the VLDB Endowment* 16, 2679–2685. doi:[10.14778/3603581.3603604](https://doi.org/10.14778/3603581.3603604).
- Peng, Y., Henke, E., Reinecke, I., Zoch, M., Sedlmayr, M., Bathelt, F., 2023. An ETL-process design for data harmonization to participate in international research with German real-world data based on FHIR and OMOP CDM. *International Journal of Medical Informatics* 169, 104925. doi:[10.1016/j.ijmedinf.2022.104925](https://doi.org/10.1016/j.ijmedinf.2022.104925).
- Reynolds, C.J., Wyatt, J.C., 2011. Open Source, Open Standards, and Health Care Information Systems. *Journal of Medical Internet Research* 13, e1521. doi:[10.2196/jmir.1521](https://doi.org/10.2196/jmir.1521).
- Rieke, N., Hancox, J., Li, W., Milletari, F., Roth, H.R., Albarqouni, S., Bakas, S., Galtier, M.N., Landman, B.A., Maier-Hein, K., Ourselin, S., Sheller, M., Summers, R.M., Trask, A., Xu, D., Baust, M., Cardoso, M.J., 2020. The future of digital health with federated learning. *npj Digital Medicine* 3, 1–7. doi:[10.1038/s41746-020-00323-1](https://doi.org/10.1038/s41746-020-00323-1).
- Sinaci, A.A., Gencturk, M., Alvarez-Romero, C., Laleci Erturkmen, G.B., Martinez-Garcia, A., Escalona-Cuaresma, M.J., Parra-Calderon, C.L., 2024. Privacy-preserving federated machine learning on FAIR health data: A real-world application. *Computational and Structural Biotechnology Journal* 24, 136–145. doi:[10.1016/j.csbj.2024.02.014](https://doi.org/10.1016/j.csbj.2024.02.014).
- Smits, D., Van Beusekom, B., Martin, F., Veen, L., Geleijnse, G., Moncada-Torres, A., 2022. An Improved Infrastructure for Privacy-Preserving Analysis of Patient Data, in: Mantas, J., Gallos, P., Zoulias, E., Hasman, A., Househ, M.S., Diomidous, M., Liaskos, J., Charalampidou, M. (Eds.), *Studies in Health Technology and Informatics*. IOS Press. doi:[10.3233/SHTI220682](https://doi.org/10.3233/SHTI220682).

- Syzdykova, A., Malta, A., Zolfo, M., Diro, E., Oliveira, J.L., 2017. Open-Source Electronic Health Record Systems for Low-Resource Settings: Systematic Review. *JMIR Medical Informatics* 5, e44. doi:[10.2196/medinform.8131](https://doi.org/10.2196/medinform.8131).
- Teo, Z.L., Jin, L., Liu, N., Li, S., Miao, D., Zhang, X., Ng, W.Y., Tan, T.F., Lee, D.M., Chua, K.J., Heng, J., Liu, Y., Goh, R.S.M., Ting, D.S.W., 2024. Federated machine learning in healthcare: A systematic review on clinical applications and technical architecture. *Cell Reports Medicine* 5, 101419. doi:[10.1016/j.xcrm.2024.101419](https://doi.org/10.1016/j.xcrm.2024.101419).
- Thaiya, M.S., Julia, K., Joram, M., Benard, M., Nambiro, D.A., 2021. Adoption of ICT to Enhance Access to Healthcare in Kenya .
- Tilahun, B., Mamuye, A., Yilma, T., Shehata, Y., 2023. African Union Health Information Exchange Guidelines and Standards. Technical Report.
- Tsafnat, G., Dunscombe, R., Gabriel, D., Grieve, G., Reich, C., 2024. Converge or collide? Making sense of a plethora of open data standards in healthcare: An editorial.