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

Daniel Kapitan   Femke Heddema   Andre Dekker

Melle Sieswerda   Bart-Jan Verhoeff   Matt Berg

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 contexts 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 availibility 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.

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 exhange 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, Sørensen, and Basole 2018) and the paradox of open (Keller and Tarkowski 2021). Subsequently, we outline the importance of the open source ecosystem by presenting 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.

Digital platforms require extensibility, availibility of complementary components and availibility 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. Open source implementations of components, such as software development kits 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 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, availibility of complements to the core technical platform (in our case the data standard itself) and availibility of executable pieces of software (de Reuver, Sørensen, and Basole 2018). Only when the majority of these aspects of digital platforms are met can we resonably expect that the platform will indeed be longlived.

In what they call the paradox of open, Keller and Tarkowsi argue that this conventional approach of open standards and open source components 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 Tarkoswki, 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) have 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 ("National Digital Health Mission" 2020; "HCX Protocol V0.9" 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. The upcoming legislation on the European Health Data Space (EHDS) mandates interoperability between electronic health record systems but has not specified which standard is to be used, although FHIR and openEHR have both been mentioned in the legislative discussion.

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. 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, although for the latter notable mature open

---

[1]See https://en.wikipedia.org/wiki/GSM#Initial_European_development.

[2]https://build.fhir.org/ig/FHIR/sql-on-fhir-v2/

source implementation are also emerging such as EHRbase.[3]

Hence, we stress that beyond evaluating the instrinic 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 and availibility of complementary components. 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.

Standardization 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 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 reference architecture of the infrastructure to be, comprising 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.

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, where data is transformed from the original source (component 6), stored in common data model (component 7) and properly annotated with metadata (component 8). Indeed, by now there are quite a few reports of real-world implementations of federated learning networks based on the OHDSI-OMOP stack,

---

[3]FHIR: https://confluence.hl7.org/display/FHIR/Open+Source+Implementations.   OMOP: https://www.ohdsi.org/software-tools/.   openEHR: https://openEHR.org/products_tools/platform/.   EHRbase: https://www.ehrbase.org/
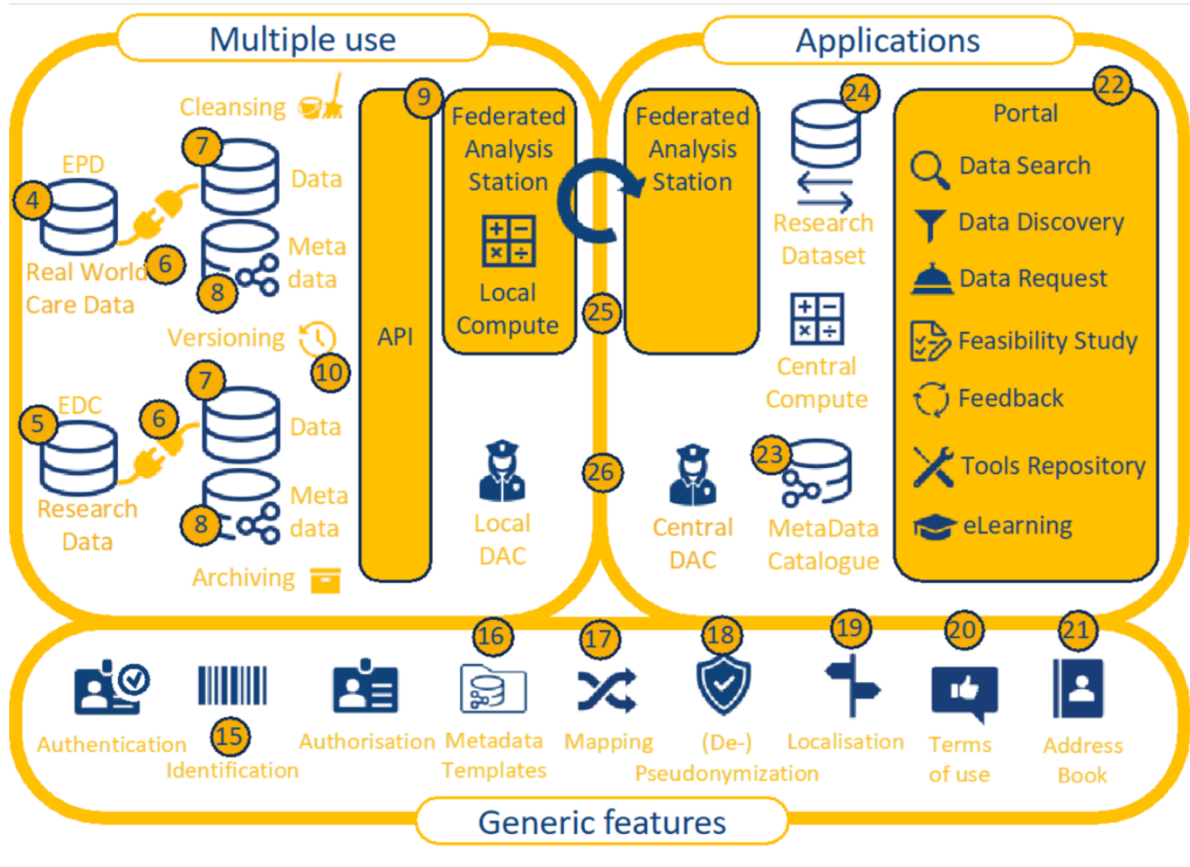
Figure 1: Reference architecture for the Dutch health data infrastructure for research and innovation (Health-RI 2024)

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 Federated Health Innovation Network (FHIN) [4].

For the PLUGIN project, however, we choose to adopt FHIR as the common 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, and the support of many healthcare terminologies and flexibility through the profiling mechanims (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 FL infrastructure in Canada similar to the PLUGIN project, 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 used FHIR as part of a FAIRification workflow to simplify the process of data extraction and preparation for clinical study analyses (Sinaci et al. 2024).

Given that OMOP can be conceptually 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 feasiblity 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 distiction between FHIR amd OMOP becomes less relevant if data can be effectively stored in either standard. We are hopeful that initiatives like OMOP-on-FHIR indeed will foster convergence rather than collision between these two standards.[5]

In the case of PLUGIN, another important consideration for choosing FHIR over OMOP is, that from a data architecture perspective, the mechanism of FHIR Profiles can be tied to principle 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 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 ingestion 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-exists 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.
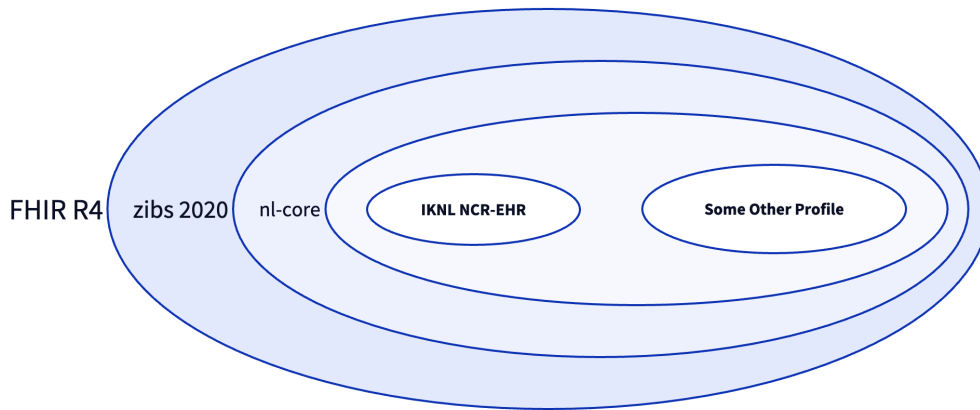

Figure 2: Principle of late binding with FHIR profiling mechanism, illustrated with FHIR Profiles that are currently in use in the Netherlands.

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 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. By explicitly aligining the mechanism of FHIR Profiles with this design pattern of a data lakehouse enables us to use complementary standards and open source components, most notably Apache Arrow as the standard columnar in-memory format with RPC-based data movement; Apache Parquet as the standard columnar on-disk format; and Apache Iceberg as

the open table format (Jain et al. 2023).[6]

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.

The above considerations also show the conceptual difference of FHIR as a health data exchange standard versus openEHR as a persistent storage of healthcare data and OMOP as a persistent storage of health research data. For health data exchange and federated learning, the recipient of the data determines to a large extent what subset of data available in the source needs to be made available – i.e. the target data model is known late and this favors late binding. In a persistent storage setting, the holder of the source data determines what data needs to be stored – and typically everything – which favors early binding.
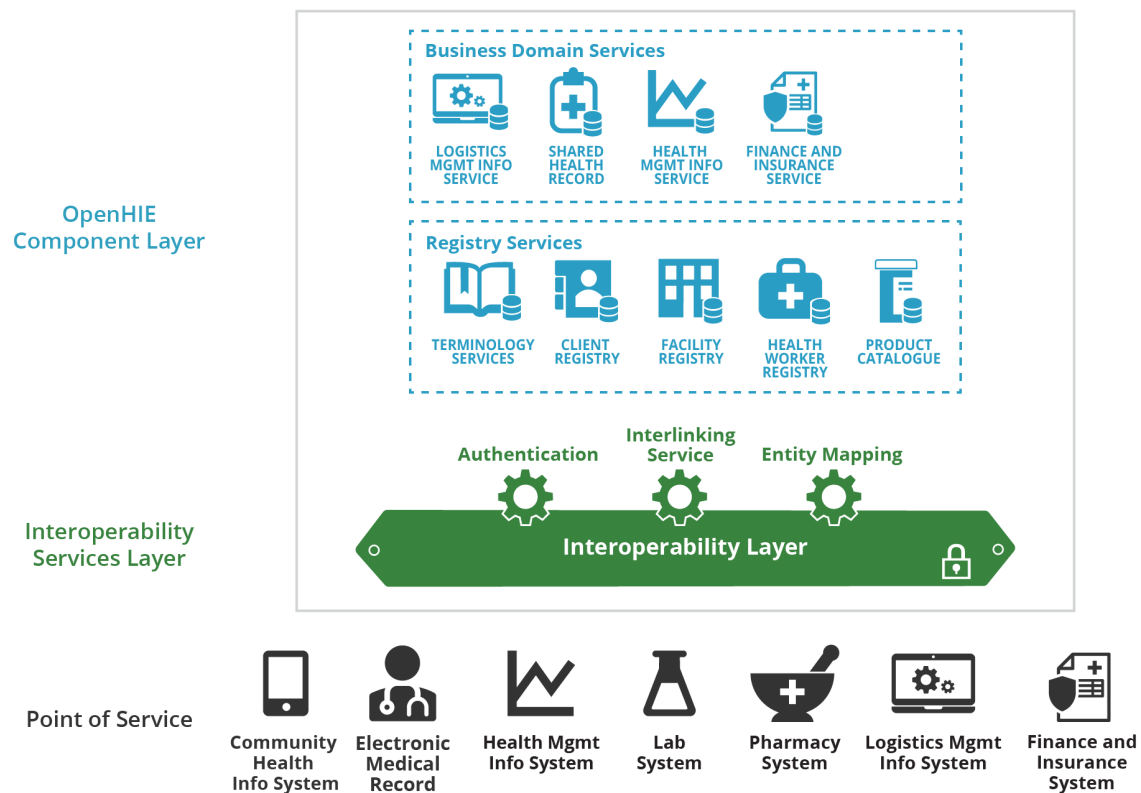
Health data standards in LMICs

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 fragmentation of health data stands in the way of scaling up digital health programmes beyond project-centric, vertical solutions into sustainable health information exchanges (Karamagi et al. 2022). Mehl et al. have called for convergence to open standards, similar to Tsafnat et al., but additionally stress the need for open source technologies (our main argument of this paper), open content (representations of public health, health system or clinical knowledge to guide implementations) and open architectures (reusable enterprise architecture patterns for health systems) (G. L. Mehl et al. 2023). As for the open architecture, we see a convergence towards the OpenHIE framework ("OpenHIE Framework V5.2-En" 2024), which has been adopted by many sub-Saharan African countries as the architectural blueprint for implementing nation-wide health information exchanges (HIE) (Mamuye et al. 2022), including Nigeria (Dalhatu et al. 2023), Kenya (Thaiya et al. 2021) and Tanzania (Nsaghurwe et al. 2021). Figure 3 shows an overview of the OpenHIE architecture.

While the OpenHIE specification is agnostic to which data standard should be used, in practice the digital health community in LMICs have *de facto* converged towards FHIR as the primary standard for health information exchange, in line with the proposal by Tsafnat et al. . To illustrate this point, consider the OpenHIM Platform architecture, which is currently the largest open source implementation of the OpenHIE specification Figure 4. Clients (Point-of-Service systems) can initiate various workflows to submit or query patient data. The Shared Health Record (SHR) acts as the core transactional system for the health information exchange, which in this case is realized with the HAPI FHIR server, being one of the most widely used open source implementations.[7]

---

[6]See https://arrow.apache.org, https://parquet.apache.org and https://iceberg.apache.org
[7]https://hapifhir.io

8

OpenHIE Component Layer

Business Domain Services

LOGISTICS MGMT INFO SERVICE · SHARED HEALTH RECORD · HEALTH MGMT INFO SERVICE · FINANCE AND INSURANCE SERVICE

Registry Services

TERMINOLOGY SERVICES · CLIENT REGISTRY · FACILITY REGISTRY · HEALTH WORKER REGISTRY · PRODUCT CATALOGUE

Interoperability Services Layer

Authentication · Interlinking Service · Entity Mapping

Interoperability Layer

Point of Service

Community Health Info System · Electronic Medical Record · Health Mgmt Info System · Lab System · Pharmacy System · Logistics Mgmt Info System · Finance and Insurance System

OpenHIE 2021-09-29; CC BY 4.0

Figure 3: OpenHIE architecture showing the Point of Service systems (black), the Interoperability Layer (green) and the Component Layer (blue).
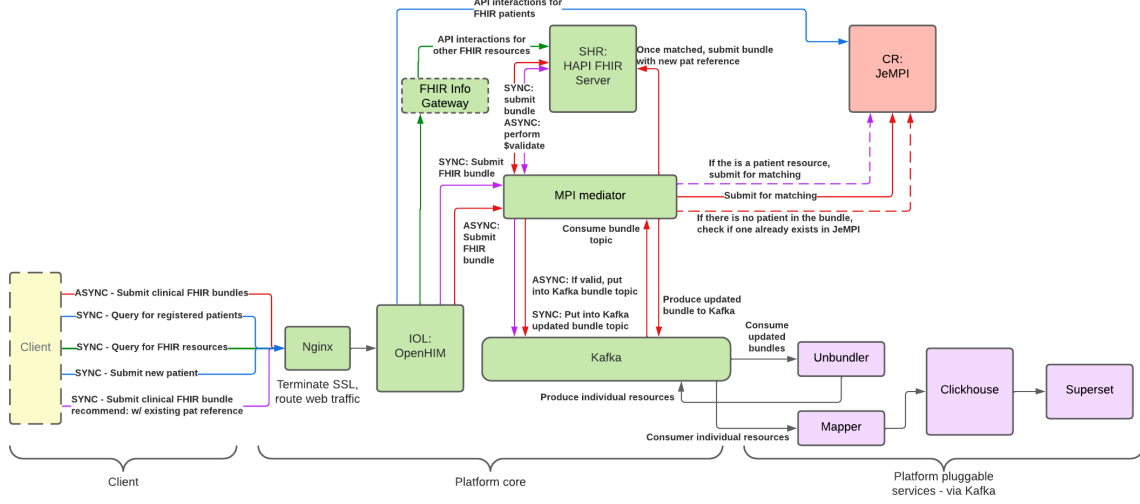
Figure 4: OpenHIM Platform Architecture, illustrating the use of FHIR-based workflows between the components as specified in OpenHIE. CR: Client Registry. IOL: Interoperability Layer. MPI: Master Patient Index. SHR: Shared Health Record. Image taken from https://jembi.gitbook.io/.

Looking at the Point-of-Service systems, we see that as of today openEHR does not play any role of significance as the standard for clinical administration. Within the context of LMICs, the largest open source EHR implementations are based on proprietary data models, and it is unlikely this will change any time soon (Syzdykova et al. 2017). Instead, we see that FHIR-native software development frameworks such as OpenSRP (G. Mehl 2020) and the Open Health Stack[8] are being used more and more, where Android apps are used for clinical administration by health professionals (Figure 5). OpenSRP has been deployed in 14 countries targeting various patient populations, amongst which a reference implementation of the WHO antenatal and neonatal care guidelines for midwives in Lombok, Indonesia (Summit Institute for Development 2023; Kurniawan et al. 2019). This solution design is particularly useful for mid-size and smaller healthcare facilities, which are often resource constrained, lacking basic IT infrastructure to deploy a full-blown electronic medical record system. Hence, by necessity, the FHIR-based SHR functions as the clinical administration system of record and as the hub for information exchange at the same time.

Finally, regarding longitudinal data analysis we also see a convergence towards FHIR as the primary standard. As is the case of federated learning, the choice for FHIR to implement datawarehouse and analytic platforms is the preferred method due to the widespread availibility of complementary open source technologies. FHIR-specific technologies such as Bulk FHIR data access and SQL-on-FHIR mentioned earlier, allow the FHIR ecosystem to be used,
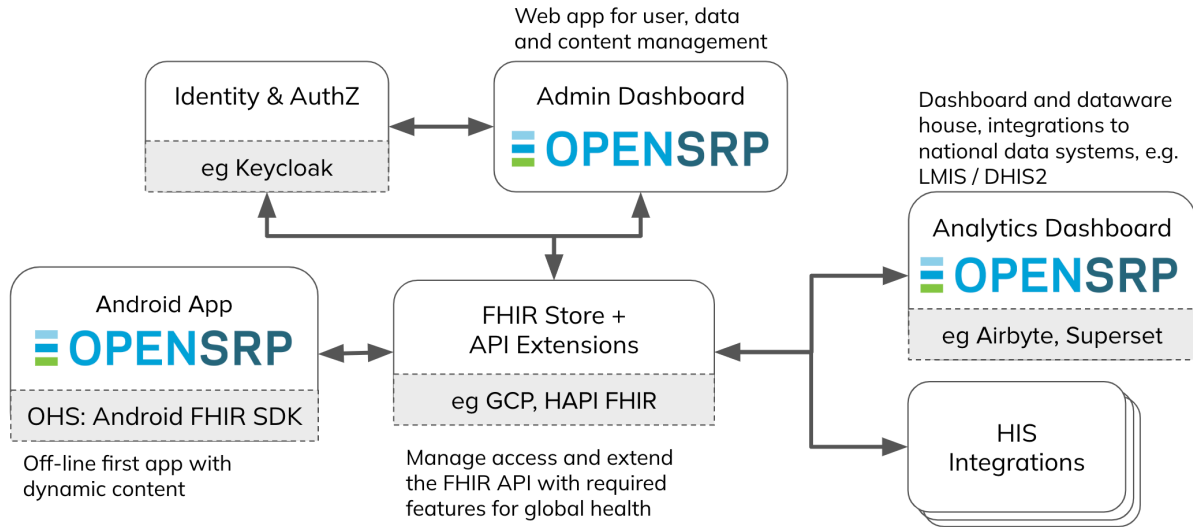
---

[8]https://developers.google.com/open-health-stack

Figure 5: Overview of OpenSRP2 open source framework for building clinical administration apps. HIS: health information systems. Image source: https://docs.opensrp.io/.

complemented and integrated with generic open source datawarehousing technologies such as Clickhouse[9] and dbt.[10]

All in all, we see that in the context of LMICs, the standardization of the three domains put forward by Tsafnat merge into one. The SHR, as the key component within the OpenHIE specification, serves as the back-end of the system-of-record and provides a transactional, persistent storage engine for information exchange. Downstream longitudinal data stores continue to use FHIR as the common data model for analytical purposes. One could argue that it is in fact advantageous to converge to just one standard, thereby reducing complexity and cost of the total system. Such a perspective ties in with the hourglass model of layered systems architecture which has been used in the design of the Internet and Unix and has enabled viral adoption and deployment scalibility (Estrin and Sim 2010; Beck 2019). The hourglass mode is "... an approach to design that seeks to support a great diversity of applications (at the top of the hourglass) and allow implementation using a great diversity of supporting services (at the bottom)." (Beck 2019) The center of the hourglass - the waist or also called the spanning layer in the information systems parlance - is defined by a set of minimal standards which mediates all interactions between the higher and lower layers. In case of the Internet, the spanning layer is defined by the TCP/IP protocol, which is supported by a variety of supporting services (many different physical networks) and used in building various applications (email, videoconferencing etc.).

Within the context of LMICs, we believe that FHIR can act as the spanning layer within open health data system at large. Because FHIR is inherently designed to make optimal use

---

[9]https://getdbt.com

[10]https://www.getdbt.com/

of internet standards, such as the json file format and REST APIs, it is very modular and developer friendly. The many components that make up the FHIR allows the standard to be used effectively to implement subsystems, such as a facility registry or a health worker registry. In comparison, OMOP and openEHR are less modular in their design and are thereby less suitable as a standard to implement the subsystems defined, for example, in the OpenHIE specification.

Conclusion

We agree with Tsafnat et al. that there is a dire need to converge to open data standards in healthcare, and support the proposal to focus on openEHR, FHIR and OMOP in healthcare informatics going forward. However, open standards are a necessary but not sufficient condition for the convergence of health data standardization. The availability of open source implementations and complementary technologies are as important when choosing which open standard to use. Furthermore, we find that the proposed trichotomy is not always relevant. As an alternative, we find that the full-STAC approach described by Mehl et al. more comprehensive. In the case of FL, we see a convergence towards OMOP and FHIR, which can be used interchangeably. In the case of LMICs, we think that FHIR as the potential of acting as the spanning layer within the open health data system at large, thereby enabling much wider standardization and adoption. We strongly support ongoing developments to increase the availibility of open source implementations as digital public goods ("Digital Public Goods Alliance" 2024) and integration projects such as Instant OpenHIE[11], through which we have a fighting chance to move the needle in health data standardization for LMICs.

References

Armbrust, Michael, Ali Ghodsi, Reynold Xin, and Matei Zaharia. 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)*, 8.

Beck, Micah. 2019. "On the Hourglass Model." *Communications of the ACM* 62 (7): 48–57. https://doi.org/10.1145/3274770.

Choudhury, Ananya, Johan van Soest, Stuti Nayak, and Andre Dekker. 2020. "Personal Health Train on FHIR: A Privacy Preserving Federated Approach for Analyzing FAIR Data in Healthcare." In *Machine Learning, Image Processing, Network Security and Data Sciences*, edited by Arup Bhattacharjee, Samir Kr. Borgohain, Badal Soni, Gyanendra Verma, and Xiao-Zhi Gao, 85–95. Communications in Computer and Information Science. Singapore: Springer. https://doi.org/10.1007/978-981-15-6315-7_7.

---

[11]https://jembi.gitbook.io/instant-v2

Cremonesi, Francesco, Vincent Planat, Varvara Kalokyri, Haridimos Kondylakis, Tiziana Sanavia, Victor Miguel Mateos Resinas, Babita Singh, and Silvia Uribe. 2023. "The Need for Multimodal Health Data Modeling: A Practical Approach for a Federated-Learning Healthcare Platform." *Journal of Biomedical Informatics* 141 (May): 104338. https://doi.org/10.1016/j.jbi.2023.104338.

Dalhatu, Ibrahim, Chinedu Aniekwe, Adebobola Bashorun, Alhassan Abdulkadir, Emilio Dirlikov, Stephen Ohakanu, Oluwasanmi Adedokun, et al. 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 (03/04): 130–39. https://doi.org/10.1055/s-0043-1768711.

de Reuver, Mark, Carsten Sørensen, and Rahul C. Basole. 2018. "The Digital Platform: A Research Agenda." *Journal of Information Technology* 33 (2): 124–35. https://doi.org/10.1057/s41265-016-0033-3.

"Digital Public Goods Alliance." 2024. *Digital Public Goods Alliance - Promoting Digital Public Goods to Create a More Equitable World.* https://digitalpublicgoods.net/.

Estrin, Deborah, and Ida Sim. 2010. "Health Care Delivery. Open mHealth Architecture: An Engine for Health Care Innovation." *Science (New York, N.Y.)* 330 (6005): 759–60. https://doi.org/10.1126/science.1196187.

Firely. 2023. "FHIR in US Healthcare Regulations."

Gruendner, Julian, Thorsten Schwachhofer, Phillip Sippl, Nicolas Wolf, Marcel Erpenbeck, Christian Gulden, Lorenz A. Kapsner, et al. 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 (10): e0223010. https://doi.org/10.1371/journal.pone.0223010.

Hai, Rihan, Christos Koutras, Christoph Quix, and Matthias Jarke. 2023. "Data Lakes: A Survey of Functions and Systems." *IEEE Transactions on Knowledge and Data Engineering* 35 (12): 12571–90. https://doi.org/10.1109/TKDE.2023.3270101.

Harby, Ahmed A., and Farhana Zulkernine. 2022. "From Data Warehouse to Lakehouse: A Comparative Review." In *2022 IEEE International Conference on Big Data (Big Data)*, 389–95. Osaka, Japan: IEEE. https://doi.org/10.1109/BigData55660.2022.10020719.

———. 2024. "Data Lakehouse: A Survey and Experimental Study." {{SSRN Scholarly Paper}}. Rochester, NY. https://doi.org/10.2139/ssrn.4765588.

"HCX Protocol V0.9." 2023.

Health-RI. 2024. "Agreements on the National Health Data Infrastructure for Research, Policy and Innovation - Health-RI Nationale Gezondheidsdata-infrastructuur - Confluence." Wiki. https://health-ri.atlassian.net/wiki/spaces/HNG/pages/249073646/Agreements+on+the+National+Health

Jain, Paras, Peter Kraft, Conor Power, Tathagata Das, Ion Stoica, and Matei Zaharia. 2023. "Analyzing and Comparing Lakehouse Storage Systems."

Jones, James, Daniel Gottlieb, Joshua C Mandel, Vladimir Ignatov, Alyssa Ellis, Wayne Kubick, and Kenneth D Mandl. 2021. "A Landscape Survey of Planned SMART/HL7 Bulk FHIR Data Access API Implementations and Tools." *Journal of the American Medical Informatics Association* 28 (6): 1284–87. https://doi.org/10.1093/jamia/ocab028.

Karamagi, Humphrey C, Derrick Muneene, Benson Droti, Violet Jepchumba, Joseph C Okei-

bunor, Juliet Nabyonga, James Avoka Asamani, Moussa Traore, and Hillary Kipruto. 2022. "eHealth or e-Chaos: The Use of Digital Health Interventions for Health Systems Strengthening in Sub-Saharan Africa over the Last 10 Years: A Scoping Review." *Journal of Global Health* 12 (December): 04090. https://doi.org/10.7189/jogh.12.04090.

Keller, Paul, and Alek Tarkowski. 2021. "The Paradox of Open." *Open Future*, March.

Khalid, Sara, Cynthia Yang, Clair Blacketer, Talita Duarte-Salles, Sergio Fernández-Bertolín, Chungsoo Kim, Rae Woong Park, et al. 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 (November): 106394. https://doi.org/10.1016/j.cmpb.2021.106394.

Kroes, Johannes A., Aruna T. Bansal, Emmanuelle Berret, Nils Christian, Andreas Kremer, Anna Alloni, Matteo Gabetta, et al. 2022. "Blueprint for Harmonising Unstandardised Disease Registries to Allow Federated Data Analysis: Prepare for the Future." *ERJ Open Research* 8 (4). https://doi.org/10.1183/23120541.00168-2022.

Kurniawan, Kevin, Inraini FitriaSyah, Ahmad Rafi Jayakusuma, Resty Asmauryanah Armis, Yusran Lubis, Muhammad Abdi Haryono, Benyamin Harefa, and Anuraj Shankar. 2019. "Midwife Service Coverage, Quality of Work, and Client Health Improved After Deployment of an OpenSRP-driven Client Management Application in Indonesia." In *5th International Conference on Health Sciences (ICHS 2018)*, 155–62. Atlantis Press. https://doi.org/10.2991/ichs-18.2019.21.

Lee, Seongwon, Chungsoo Kim, Junyuk Chang, and Rae Woong Park. 2022. "FeederNet (Federated E-Health Big Data for Evidence Renovation Network) Platform in Korea – OHDSI."

Mamuye, Adane L., Tesfahun M. Yilma, Ahmad Abdulwahab, Sean Broomhead, Phumzule Zondo, Mercy Kyeng, Justin Maeda, Mohammed Abdulaziz, Tadesse Wuhib, and Binyam C. Tilahun. 2022. "Health Information Exchange Policy and Standards for Digital Health Systems in Africa: A Systematic Review." *PLOS Digital Health* 1 (10): e0000118. https://doi.org/10.1371/journal.pdig.0000118.

Mandl, Kenneth D., Daniel Gottlieb, Joshua C. Mandel, Vladimir Ignatov, Raheel Sayeed, Grahame Grieve, James Jones, Alyssa Ellis, and Adam Culbertson. 2020. "Push Button Population Health: The SMART/HL7 FHIR Bulk Data Access Application Programming Interface." *Npj Digital Medicine* 3 (1): 1–9. https://doi.org/10.1038/s41746-020-00358-4.

Mateus, Pedro, Justine Moonen, Magdalena Beran, Eva Jaarsma, Sophie M. van der Landen, Joost Heuvelink, Mahlet Birhanu, et al. 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 (July): 104661. https://doi.org/10.1016/j.jbi.2024.104661.

Mehl, Garrett. 2020. "Open Smart Register Platform (OpenSRP)." *mHealth Compendium* 5 (October): 42–43.

Mehl, Garrett L, Martin G Seneviratne, Matt L Berg, Suhel Bidani, Rebecca L Distler, Marelize Gorgens, Karin E Kallander, et al. 2023. "A Full-STAC Remedy for Global Digital Health Transformation: Open Standards, Technologies, Architectures and Content." *Oxford Open Digital Health* 1 (January): oqad018. https://doi.org/10.1093/oodh/oqad018.

Mullie, Louis, Jonathan Afilalo, Patrick Archambault, Rima Bouchakri, Kip Brown, David L Buckeridge, Yiorgos Alexandros Cavayas, et al. 2023. "CODA: An Open-Source Platform for Federated Analysis and Machine Learning on Distributed Healthcare Data." *Journal of the American Medical Informatics Association*, December, ocad235. https://doi.org/10.1093/jamia/ocad235.

"National Digital Health Mission." 2020. India National Health Authority.

Nsaghurwe, Alpha, Vikas Dwivedi, Walter Ndesanjo, Haji Bamsi, Moses Busiga, Edwin Nyella, Japhet Victor Massawe, et al. 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 (1): 139. https://doi.org/10.1186/s12911-021-01499-6.

"OpenHIE Framework V5.2-En." 2024. *OpenHIE*. https://ohie.org/.

Pedreira, Pedro, Orri Erling, Konstantinos Karanasos, Scott Schneider, Wes McKinney, Satya R Valluri, Mohamed Zait, and Jacques Nadeau. 2023. "The Composable Data Management System Manifesto." *Proceedings of the VLDB Endowment* 16 (10): 2679–85. https://doi.org/10.14778/3603581.3603604.

Peng, Yuan, Elisa Henke, Ines Reinecke, Michéle Zoch, Martin Sedlmayr, and Franziska Bathelt. 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 (January): 104925. https://doi.org/10.1016/j.ijmedinf.2022.104925.

Reynolds, Carl J., and Jeremy C. Wyatt. 2011. "Open Source, Open Standards, and Health Care Information Systems." *Journal of Medical Internet Research* 13 (1): e1521. https://doi.org/10.2196/jmir.1521.

Rieke, Nicola, Jonny Hancox, Wenqi Li, Fausto Milletarì, Holger R. Roth, Shadi Albarqouni, Spyridon Bakas, et al. 2020. "The Future of Digital Health with Federated Learning." *Npj Digital Medicine* 3 (1): 1–7. https://doi.org/10.1038/s41746-020-00323-1.

Sinaci, A. Anil, Mert Gencturk, Celia Alvarez-Romero, Gokce Banu Laleci Erturkmen, Alicia Martinez-Garcia, María José Escalona-Cuaresma, and Carlos Luis Parra-Calderon. 2024. "Privacy-Preserving Federated Machine Learning on FAIR Health Data: A Real-World Application." *Computational and Structural Biotechnology Journal* 24 (December): 136–45. https://doi.org/10.1016/j.csbj.2024.02.014.

Smits, Djura, Bart Van Beusekom, Frank Martin, Lourens Veen, Gijs Geleijnse, and Arturo Moncada-Torres. 2022. "An Improved Infrastructure for Privacy-Preserving Analysis of Patient Data." In *Studies in Health Technology and Informatics*, edited by John Mantas, Parisis Gallos, Emmanouil Zoulias, Arie Hasman, Mowafa S. Househ, Marianna Diomidous, Joseph Liaskos, and Martha Charalampidou. IOS Press. https://doi.org/10.3233/SHTI220682.

Summit Institute for Development. 2023. "BUNDA App." *Summit Institute*. https://www.sid-indonesia.org/post/bunda-app.

Syzdykova, Assel, André Malta, Maria Zolfo, Ermias Diro, and José Luis Oliveira. 2017. "Open-Source Electronic Health Record Systems for Low-Resource Settings: Systematic Review." *JMIR Medical Informatics* 5 (4): e44. https://doi.org/10.2196/medinform.8131.

Teo, Zhen Ling, Liyuan Jin, Nan Liu, Siqi Li, Di Miao, Xiaoman Zhang, Wei Yan Ng, et al. 2024. "Federated Machine Learning in Healthcare: A Systematic Review on Clinical Applications and Technical Architecture." *Cell Reports Medicine* 5 (2): 101419. https://doi.org/10.1016/j.xcrm.2024.101419.

Thaiya, Mbugua Samuel, Korongo Julia, Mutai Joram, Masese Benard, and Dr Alice Nambiro. 2021. "Adoption of ICT to Enhance Access to Healthcare in Kenya." *IOSR Journal of Computer Engineering* 23 (2): 45–50.

Tilahun, Binyam, Adane Mamuye, Tesfahun Yilma, and Yasser Shehata. 2023. "African Union Health Information Exchange Guidelines and Standards."

Tsafnat, G, R Dunscombe, D Gabriel, G Grieve, and C Reich. 2024. "Converge or Collide? Making Sense of a Plethora of Open Data Standards in Healthcare: An Editorial."