# Beyond the paradox of interoperability in open health data standards

Daniel Kapitan[a,b,c], Melle Sieswerda[d,e], Andre Dekker[e]

[a] *Dutch Hospital Data, Utrecht, the Netherlands,*

[b] *PharmAccess Foundation, Amsterdam, the Netherlands,*

[c] *Eindhoven University of Technology, Eindhoven, the Netherlands,*

[d] *Integral Cancer Registry Netherlands, Eindhoven, the Netherlands,*

[e] *Maastricht University, Maastricht, the Netherlands,*

**Abstract**

In response to the proposal of Tsafnat et al. to converge towards three open health data standards, 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). Based on our ongoing work in both areas, we provide a critical reflection on the proposed alignment of using OpenEHR, FHIR and OMOP as the default standard for the three domains of clinical care and administration, data exchange and longitudinal analysis, respectively. We find that for the two contexts considered there, this trichotomy does not do justice to details that are crucial in real-world implementations. This perspetive describes more specific design principles and implementation choices for these two types of health data sharing. In both case, we observe a strong convergence towards FHIR, while OMOP is used as an alternative or in combination with FHIR for federated learning.

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

## 1. Looking beyond the paradox of 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

*Corresponding author
*Email address:* `daniel@kapitan.net` (Daniel Kapitan)

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 the 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), data platforms (de Reuver et al., 2022) 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; and ii) health data platforms for low and middle income countries (LMICs). We conclude with ...

## 2. The paradox of open for health data standards

Besides the paradox of interoperability put forward by Tsafnat et al., we argue that although 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 platform and associated ecosystem. 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 et al., 2018). Only when these aspects of digital platforms are fullfilled 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 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. 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. A canonical example in the birth of the GSM standard, which was mandated by European legislation.[1] Reflecting on this perspective on openness, we observer 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 (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 systems of record.

These external incentives have resulted in a large boost in both commercial and open source development activities in the FHIR ecosystem. One such example 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). 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, the ecosystem of OMOP and OpenEHR is have not yet profited from external incentives to grow their community …

- smaller than that of FHIR
- The majority of OMOP components are run by Observational Health Data Science and Informatics (OHDSI) …Say something that OMOP still has a relatively large ecosystem, with R libraries, pyOMOP etc.
- The OpenEHR, however, seems subscritical. We are not judging the con-

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

tent and approach of OpenEHR, but there are just so few implementations.

Hence, we stress that beyond the evaluating the intrinic structure of an open standard (which one is most suitable for a given domain) 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 on the whole ecosystem around a standard, FHIR stands out as having the diverse and rich ecosystem because it has been mandated in certain jurisdictions. This is relevant when comparing these standards in real-world implementations.

## 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 prevented from reaching its full potential and, ultimately, from making the transition from research to clinical practice. Data like this 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 exchanging 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. To this purpose, the 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, has been taken as the starting point (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)).

The principle of FL has been actively researched and developed within the Dutch healthcare community. The first proof-of-concept of the so-called Personal Health Train (van Soest et al., 2018) ultimately led to the open source vantage6 platform (Smits et al., 2022). One of the prerequisites of this architecture is that organizations that participate in a federation of 'data stations' use the same common data model (CDM) to make the data Findable, Accessible, Interoperable and Resusable. 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. FHIR has been chosen as the common data model quite early in the development because of its practicality, provenance of RESTful APIs out of the box, support of many healthcare terminologies and flexibility through the profiling mechanims (Choudhury et al., 2020).

This solution design has gained traction since then. The CODA platform, which aims to implement a FL infrastructure in Canada, compared OMOP

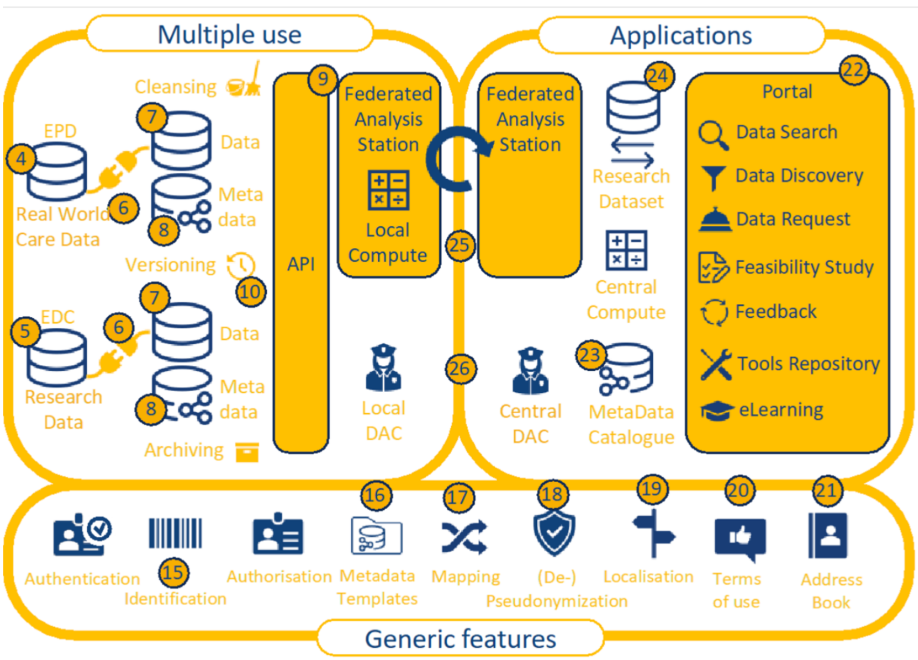


Figure 1: Reference architecture for the Dutch health data infrastructure for research and innovation (Health-RI, 2024)

and FHIR and chose the latter as it has been found to support more granular mappings required for analytics (Mullie et al., 2023). 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).

Using FHIR as the CDM in a FL architecture for persistent storage for longitudinal analysis may strike many as counterintuitive. The fact that some recent studies the practical designs of such platforms only mention OMOP only attests to this misconception (Wong et al., 2022).

- we don't say you can't use OMOP for data stations, in fact the following projects do that:
  - OHDSI analytics (Khalid et al., 2021)
  - … (look for more)
- this distinction become irrelevant as OMOP will have FHIR interfaces, see https://omoponfhir.org
- however, FHIR is the CDM we choose because it is a superset, allowing for example also administrative and claims data or claims data can't fit in OMOP
- Principle of FHIR Profiles can be tied to principle of late binding: allow ingest of widely different sources, and gradually but 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 having very detailed data
- Add more arguments here …

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

The OpenHIE framework (ope, 2022) has been adopted by many sub-Saharan African countries (Mamuye et al., 2022) as the architectural blueprint for implementing nation-wide health information exchanges (HIE), including Nigeria (Dalhatu et al., 2023), Kenya (Thaiya et al., 2021) and Tanzania (Nsaghurwe et al., 2021). Conceptually, the OpenHIE framework constitutes a framework for an open digital platform, that mostly focuses on transactional exchange of data, that is, primary data sharing. Given the need to also enable secondary data sharing for academic research, real-world evidence studies etc., African countries have, as a matter of course, extended the framework to include "data & analytics services" as an additional domain.

Based on our direct involvement in implementing and designing health data platforms in these countries, we observed that FHIR was practically the only viable solution. Although FHIR is originally intened for creating a longitudinal database, in fact open source FHIR implementations such as the HAPI FHIR server are the most widely used standards for realizing the so-called Shared Health Record within the OpenHIE architecture. not intended for

In relation to the conventional perspective on 'openness', which originally focused on open source and open standards as discussed above, has been superseded by "… conflicts about privacy, economic value extraction, the emergence of artificial intelligence, and the destabilizing effects of dominant platforms on (democratic) societies. Instead of access to information, the control of personal data has emerged in the age of platforms as the critical contention" (Keller and Tarkowski, 2021). These conflicts are particularly salient in the healthcare domain, where people are generally willing to share their health data to receive the best care (primary use), while the attitude towards secondary use of health data varies greatly depending on the type and context (Cascini et al., 2024). Case studies on digital platforms in healthcare point to an emerging pattern where the focus shifts from the digital platform with its defining software and hardware components, to the data as the primary object of interest in and of itself (Ozalp et al., 2022; Alaimo and Kallinikos, 2022). This observation ties into with the proposed research agenda by de Reuver et al. to consider data platforms as a phenomenon distinct from digital platforms (de Reuver et al., 2018, 2022).

Essentially, "… this paradox is that openness of data is both a challenge to and an enabler of concentrations of power. The ideas of open access and free reuse of information goods continue to be some of the most powerful challenges to the exclusive control by corporations and states over information goods. Yet making such resources open also exposes them to the imbalances of power that shape these societies – and in the worst cases serves to strengthen these imbalances." Put differently, having open standards is no guarantee that we will be able to breakthrough the current status quo with huge asymmetries in terms of bargaining power, access control and development resources related to health data sharing. We posit that having open data standards is a necessary condition in improving the current state of affairs, yet it is not a sufficient condition.

The shift in perspective from digital platforms supporting primary data sharing toward data platforms supporting secondary data sharing is a contentious issue both in high income countries, with for example the ongoing efforts to establish a European Health Data Space (Otto et al., 2022), and low- and middle income countries that are aiming to deploy nationwide health information exchanges to support primary and secondary data use at once (Mamuye et al., 2022). A better understanding of openness is particularly relevant if we are to realize a solidarity-based approach to health data sharing that i) gives people a greater control over their data as active decision makers; ii) ensures that the value of data is harnessed for public good; and iii) moves society towards equity and justice by counteracting dynamics of data extraction (Kickbusch et al., 2021; Prainsack et al., 2022; Prainsack and El-Sayed, 2023).

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).

## 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
- MPC as next step up from FL

## References


, 2022. OpenHIE Framework v5.0. https://ohie.org/.

, 2023. HCX Protocol v0.9.

, 2024. Digital Public Goods Alliance. https://digitalpublicgoods.net/.

Alaimo, C., Kallinikos, J., 2022. Organizations Decentered: Data Objects, Technology and Knowledge. Organization Science 33, 19–37. doi:10.1287/orsc.2021.1552.

Cascini, F., Pantovic, A., Al-Ajlouni, Y.A., Puleo, V., De Maio, L., Ricciardi, W., 2024. Health data sharing attitudes towards primary and secondary use of data: A systematic review. eClinicalMedicine 71, 102551. doi:10.1016/j.eclinm.2024.102551.

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.

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.

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.

de Reuver, M., Ofe, H., Agahari, W., Abbas, A.E., Zuiderwijk, A., 2022. The openness of data platforms: A research agenda, in: Proceedings of the 1st International Workshop on Data Economy, Association for Computing Machinery, New York, NY, USA. pp. 34–41. doi:10.1145/3565011.3569056.

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.

Firely, 2023. FHIR in US Healthcare Regulations. Technical Report.

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.

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+Infrast

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.

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., Reps, 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.

Kickbusch, I., Piselli, D., Agrawal, A., Balicer, R., Banner, O., Adelhardt, M., Capobianco, E., Fabian, C., Singh Gill, A., Lupton, D., Medhora, R.P., Ndili, N., Ryś, A., Sambuli, N., Settle, D., Swaminathan, S., Morales, J.V., Wolpert, M., Wyckoff, A.W., Xue, L., Bytyqi, A., Franz, C., Gray, W., Holly, L., Neumann, M., Panda, L., Smith, R.D., Georges Stevens, E.A., Wong, B.L.H., 2021. The Lancet and Financial Times Commission on governing health futures 2030: Growing up in a digital world. The Lancet 398, 1727–1776. doi:10.1016/S0140-6736(21)01824-9.

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.

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.

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.

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.

Otto, B., Ten Hompel, M., Wrobel, S. (Eds.), 2022. Designing Data Spaces: The Ecosystem Approach to Competitive Advantage. Springer International Publishing, Cham. doi:10.1007/978-3-030-93975-5.

Ozalp, H., Ozcan, P., Dinckol, D., Zachariadis, M., Gawer, A., 2022. "Digital Colonization" of Highly Regulated Industries: An Analysis of Big Tech Platforms' Entry into Health Care and Education. California Management Review 64, 78–107. doi:10.1177/00081256221094307.

Prainsack, B., El-Sayed, S., 2023. Beyond Individual Rights: How Data Solidarity Gives People Meaningful Control over Data. The American Journal of Bioethics 23, 36–39. doi:10.1080/15265161.2023.2256267.

Prainsack, B., El-Sayed, S., Forgó, N., Szoszkiewicz, Ł., Baumer, P., 2022. Data solidarity: A blueprint for governing health futures. The Lancet Digital Health 4, e773–e774. doi:10.1016/S2589-7500(22)00189-3.

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.

Rieke, N., Hancox, J., Li, W., Milletarì, 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.

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.

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.

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.

van Soest, J., Sun, C., Mussmann, O., Puts, M., Malic, A., van OPPEN, C., Towend, D., Dekker, A., Dumontier, M., 2018. Using the Personal Health Train for Automated and Privacy-Preserving Analytics on Vertically Partitioned Data .

Wong, J., Prieto-Alhambra, D., Rijnbeek, P.R., Desai, R.J., Reps, J.M., Toh, S., 2022. Applying Machine Learning in Distributed Data Networks for Pharmacoepidemiologic and Pharmacovigilance Studies: Opportunities, Challenges, and Considerations. Drug Safety 45, 493–510. doi:10.1007/s40264-022-01158-3.