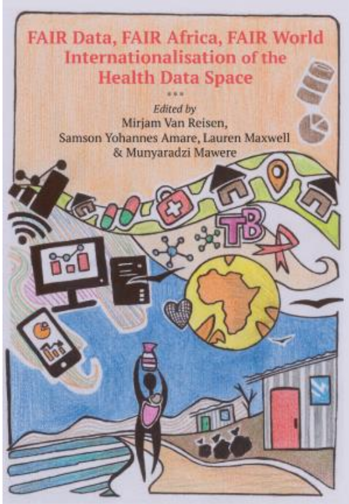
FAIR Data, FAIR Africa, FAIR World: The Internationalisation of the Health Data Space

Mirjam van Reisen, Samson Yohannes Amare, Lauren Maxwell & Munyaradzi Mawere

Chapter in:
Fair Data Fair Africa Fair World:
Internationalisation of the Health Data Space



Cite as: Van Reisen, M., Amare, S. Y., Maxwell, L. & Mawere, M. (2025). FAIR data, FAIR Africa, FAIR world: Internationalisation of the Health Data Space. https://doi.org/10.5281/zenodo.15382720. In: Van Reisen, M., Amare, S. Y., Maxwell, L. & Mawere, M. (Eds). *FAIR data, FAIR Africa, FAIR world: Internationalisation of the Health Data Space.* (pp. 1-24). Bamenda: Langaa. URL: https://www.researchgate.net/publication/391750151_FAIR_Data_FAIR_Africa_FAIR_World_The_Internationalisation_of_the_Health_Data_Space

The About the Authors note can be found here: https://raee.eu/wp-content/uploads/2025/05/About-the-Authors-and-Editors.pdf
The list of figures and tables can be found here: https://raee.eu/wp-content/uploads/2025/05/List-of-Figures-and-Tables.pdf

Contents

Chapter 1

FAIR Data, FAIR Africa, FAIR World: The Internationalisation of the Health Data Space

Mirjam van Reisen, Samson Yohannes Amare, Lauren Maxwell & Munyaradzi Mawere

Abstract

This book explores the FAIR health data space as an ethical and inclusive environment that fosters scientific integrity, social progress, and data sovereignty. It proposes data architectures that balance autonomy with the ability to share knowledge while addressing societal needs. As we look ahead, the focus shifts to solving complex challenges: ensuring that FAIR health data spaces are secure, privacy-respecting, and governed by transparent structures. This requires drawing on the lessons of Solid Privacy-Protecting Pods, Linked Data Spaces, and FAIR-OLR Data. The FAIR movement began with a powerful call to action: GO BUILD, GO CHANGE, GO TRAIN. In this book, we report on VODAN's implementation of these steps to support indigenous data reuse approaches, focusing on exploratory research related to the African Health Data Space. GO BUILD demonstrates the feasibility of novel federated architectures for FAIR data spaces. GO CHANGE shows how the federated reuse of FAIR data can be transformative to enable insights across diverse contexts. GO TRAIN forms the foundation for new models of data reuse. Trust, understanding, and competence arise from education and engagement. Training demystifies data processes, enables responsible data reuse, and ensures the participation of a new generation in a FAIR Internet of Data and Services.

Keywords: FAIR-OLR guidelines, FAIR data, Africa Health Data Space

From digital colonialism to data sovereignty

Ending helicopter research, where researchers from high-income countries conduct studies in low-to-middle-income countries without the engagement or benefit of local researchers, is a priority for global health researchers and journals (Dutta et al., 2023; McIntosh et al., 2023; Odeny & Bosurgi, 2022) and part of the move to decolonise global health. Helicopter research maintains power imbalances, shifts the focus and benefits of research from low-and-middle income countries (LMICs) to high-income countries (HICs) and limits investment in training and research infrastructure in LMICs. Controlling the collection and flow of data can be considered part of helicopter research. Data imperialism is a new form of exploitation where data is collected and controlled by HICs and LMICs must request or otherwise depend on HICs for access to the data collected in their country. When this happens with health systems or epidemic response data, countries lose their autonomy to prevent, detect, and address outbreaks or to maintain and improve their populations' health.

The FAIR (Findable, Accessible, Interoperable, and Reusable) principles (Wilkinson et al., 2016) are revolutionising how we structure and reuse health data. Understanding who data reuse benefits, should be a central consideration for data reuse architectures. The principles of FAIR-Ownership express this in Locale under Regulatory Compliance (OLR). This supports federated infrastructures for data reuse. The FAIR-OLR principles associate well with the CARE principles for indigenous data governance. The CARE principles (Carroll et al., 2020) stand for ethical principles of (i) Collective Benefit the design and function of data ecosystems enables Indigenous Peoples to derive benefit from the data; (ii) Authority to Control – the rights and interests of Indigenous Peoples' and Indigenous Data must be recognised as well as their authority to control such data; (iii) Responsibility – the use of Indigenous Data will support Indigenous Peoples' self-determination and collective benefit; and (iv) Ethics - the data life cycle and the data ecosystem will respect Indigenous Peoples' rights and wellbeing (Ramhit, 2024).

Federated reuse of FAIR health data ensures that data generators, including health facilities, retain control of their data. Federated data reuse means that data can be reused without moving or directly accessing it, ensuring data privacy and compliance with local data protection regulations through its privacy-by-design approach. FAIR-OLR is central to ensuring that data generators maintain control of their data and are the chief beneficiaries of data reuse.

Facilitating data reuse is especially important for maternal and child populations, which are often excluded from health research due to ethical concerns, regulatory complexities, and perceived vulnerability. The lack of available data and research limits the evidence base for comprehensive health surveillance, safe and effective interventions, and improving routine care. Maternal and child health data is often siloed, preventing researchers from understanding how pre-pregnancy and pregnancy exposures, including infectious diseases - affect maternal, foetal, infant, and child development. Data sovereignty is essential for respecting the needs and wishes of vulnerable populations while maximising the utility of limited data. As such, we focus our inference particularly, but not exclusively, on data collected by health facilities in the context of antenatal care.

While the Internet is the global infrastructure that connects computers and networks, the World Wide Web is the system of information sharing built on that infrastructure, designed to enable open access, inclusivity, and connectivity among users across the globe. The Mundaneum, founded in 1895, was rooted in the same values that underpin the World Wide Web: the belief that universal access to knowledge fosters peace, equality, and human progress. As a precursor to the Web, the Mundaneum envisioned a global, interconnected system to make all human knowledge discoverable and shareable. This ethos connects to Social Linked Data (Solid) Pods and the emerging idea of personal data stores and collectives. Solid Pods, like FAIR Data Points, aim to return information control to individuals while enabling secure, federated collaboration in a modern-day implementation of the Mundaneum's humanist dream of an inclusive, accessible global knowledge commons.

The World Health Organization's (WHO) Global Strategy on Digital Health 2020–2025 (WHO, 2021) calls on researchers and health systems to consider health data as a global public good that should advance health equity and outcomes. Global public goods create benefits across borders, serving multiple generations and multiple populations. Local leadership of federated approaches to health data reuse means that health facilities can respond to the WHO call for quality, granular data to inform public health surveillance and policy while addressing individual and local priorities and regulations for data reuse. FAIR Data Points and Solid Pods respond directly to the WHO call by enabling the secure reuse of granular data to benefit individuals, healthcare providers, and health systems at the local, national, and cross-national levels.

The building blocks of knowledge: Semantic data

If information - and the knowledge that emerges from it - is dissected, the resource of it is ‘data’ and ‘inference’. A data instance can be defined as the most minor component, a single, specific occurrence of an instance within a set of values for a given structure or schema. Inference is the process of deducing information or knowledge from data instances. A digital data instance is the most granular occurrence of a data instance in a digital format. Digital inference relates to the processes of digital workflows to create information (and knowledge) from the digital data analysis, which can be through machine-learning or artificial intelligence processes.

In an analogue setting, data instances and semantics are intertwined. A data instance always occurs at a specific time, in a particular place and is observed by a certain being, all of which is context that adds meaning to the data instance. The inference process is often implicit, in which the linkages of the data instance to place, time and observer are accepted as a natural given. In a digital setting, the data instance and semantics are not necessarily given or might not necessarily be meaningful. A cell number ‘7’ has no meaning unless it can be connected to labels, such as ‘daughter’ and ‘age’, which could lead to the inference that this concerns someone with a daughter who is 7 years old. However, unless other semantic information is provided on the source of the information, we would not know whether this information is old information and the daughter has now become an adult, or current, or information on the future. To interpret ‘7’, the contextual information of the catalogue in which this data instance is incorporated is equally important. If we know that this concerns a register of a school in 1963, we would know that today (in 2025), this person is 69 years old.

Until recently, the orientation of the Internet has been to serve human thought processes with information and knowledge. Much less attention has been given to the substantive components of this information, the data instances, and supporting the inference with machine-actionable computational processes. This process of creating data instances as semantic objects is mimicked in FAIR data. In a structured way, each data instance is linked to other instances in object–predicate–subject triples, e.g., 'My daughter is 7'. Each statement is part of a data set: 'a school-registration in 1963' and a catalogue: 'St. Mary's College in Stratford', and a FAIR Data Point that makes the catalogues discoverable: 'Stratford primary schools'. In this vertical hierarchy of information, coupled with horizontal overlap in triples, information that can be deduced quickly builds up. This is a relevant strategy to deal with large volumes of information and make information discoverable and accessible.

Solid Pods, edge computing and trusted sources of truth

Sir Timothy Berners-Lee, the creator of the World Wide Web, was among the first to identify that the Internet was broken and required a different structure. Solid Pods are decentralised personal data stores that return control over how data is accessed and used to the individual (Sambra et al., 2016). Sir Timothy Berners-Lee developed Solid Pods to restore individual ownership and control over personal data in response to large tech platforms' growing centralisation and exploitation of information. Unlike traditional applications that retain user data in isolated silos, Solid Pods allow users to grant and revoke access to their data across multiple services without relinquishing ownership. As such, Solid Pods enable application-independent, consent-based data reuse and enhance privacy and data sovereignty. Solid Pods support secure, federated data reuse in health research and care by allowing patients to share selected health information with providers and researchers while maintaining control over their personal data.

The semantically linked, indexed catalogues and coded data enable the operations to be carried out in the containers for approved processes and applications. The separation of platform and data container allows for a new model in which the services in the platforms can be separated from the handling of the data. This also allows for the data sources to be assessed in terms of trust—is the data trustworthy, of good quality, and relevant?

Like the Solid approach, the FAIR Data Point (FDP) indexes the data and its quality, allowing visitors to discover the data source and determine whether it is relevant and trusted. Data visiting is the essential change of directionality needed to create autonomy over data; rather than centralising the data, it directs operations outwards to the data containers by allowing the computation to travel towards the pods to collect the computed aggregates and assemble the federated outcomes for further processing. This reversal of directionality allows the data handler, the owner of the data pod, to determine access control and grant permission. This process returns power to the owner to whom the data pertains.

This is a transformative and consequential reversal of directionality. Empowering the data producers to determine data uses encourages equitable data reuse while building better quality, more granular source data. Solid Pods and FAIR data points move beyond commercial data reuse to enable individuals' reuse of their own data, health practitioners' reuse of patient data, and health facility reuse of routine health care data. Moving towards secure, ethical, and equitable reuse of granular data to build better health and health care at the individual, health centre, sub-national, national, and global levels.

Autonomy, FAIR data and the choreography of access

Data is a material resource that pertains to a data subject, which evokes ethical qualities concerning control over the data and autonomy in determining the boundaries of its use and re-use—which can and should be subject to access arrangements. FAIR data support both data reuse and the imposition of access conditions and controls related to its reuse. FAIR data is inherently federated artificial intelligence (AI) ready and ready to be used in a federated approach. Data rather than application-centric approaches to data reuse facilitate the specification of granular conditions for reuse, enabling data protection legislation and individual decision-making about who should use data and for what purposes. Solid Pods and FAIR Data Points enable data collaboratives where data reuse results in shared benefits rather than representing an extractive process.

This technical transformation from application-centric data reuse to data-visiting of FAIR data objects supports the Internet of FAIR Data and Services (IFDS), where access is strictly controlled under local access conditions. FAIR-ORL supports the shift in the choreography of access and ensures data generators have data sovereignty. This change is particularly pertinent for personal, sensitive data, like health data, which should be used to build individual and local health under locally determined access conditions.

The African Health Data Space

The Africa Health Data Space is a collaborative initiative to improve health data management and analytics across Africa. It focuses on generating continuous, real-time, high-velocity clinical observational patient data from resource-limited communities that have not been well-represented in digital health data. The initiative involves deploying FAIR Data Points in health facilities across multiple African countries, including Uganda, Kenya, Tanzania, Ethiopia, Nigeria, Rwanda, Namibia, Liberia, Guinea Bissau, Ghana, Tunisia and South Africa, with plans to expand further. The data remains securely stored in the health facilities, ensuring local ownership and compliance with regulatory frameworks.

The federated Africa Health Data Space aims to enable ethical and efficient data sharing and analysis, supporting local and global health research and outbreak response. This initiative is part of a broader grassroots effort to strengthen national capacities for health data analytics and promote the use of health data at the point of care in Africa and beyond. The African Health Data Space responds to the pressing need for indigenous data collection and reuse approaches rooted in equity and data sovereignty. While global health policy and practice are meant to build population health, extractive practices in global health research represent a form of digital colonialism. HIV data from Africa have long been collected and stored by external agencies such as the U.S. President's Emergency Plan for AIDS Relief (PEPFAR). The continent's HIV data mainly resided on servers outside the continent, managed with limited involvement from local health authorities. While this data was instrumental in shaping global health policies and funding decisions, African countries often lacked access to the complete datasets, restricting their ability to analyse trends or set priorities independently. Similarly, during the 2013 Ebola outbreaks in West Africa, data was rapidly extracted and shared internationally, sometimes without proper ethical oversight or benefit to the local health systems. This reinforces a model of extractive data use in emergency response. Many of these datasets are no longer available in the country where the data was collected and are no longer discoverable.

Like the Ebola data, data taken out of context lose their interpretability and value. Poorly curated data and data lacking contextual metadata present a huge problem for health researchers and policymakers. Callard's (2025) investigation of stillbirths through birth registry data in Namibia demonstrates how disparate, heterogeneous data creates challenges for researchers investigating the causes of stillbirth and other adverse birth outcomes. The African Health Data Space responds to the shared need to use local data for local health while providing quality and granularity of data and the engagement of local expertise needed to interpret data and related findings. As such, the African Health Data Space moves from extractive data use in emergency response or global health research to ethical and equitable data reuse through data sovereignty.

The Africa Health Data Space is built on three layers. The first operational layer is where semantic health data curation, indexing and catalogues happen, enabling the data instances to be usable and interoperable for multiple operations. The services layer identifies the access control and permissions, data security and how the data is exposed. The intelligence later offers the operations, through data visiting, over limited data in computations carried out in federated format, allowing insights. Such insights can assist the health workers within a health facility. If access is granted under defined conditions, the insights from the data can also assist the planning of clinical trials, monitoring and surveillance of diseases, and pharmacovigilance. The

Africa Health Data Space is the brain connecting the different containers of information through which algorithms can travel. If a node provides permission for access, this node can participate in the analytical computation carried out.

VODAN: The Value-driven Ownership of Data and Accessibility Network

The research presented in this volume results from collaboration in the Value-driven Ownership of Data and Accessibility Network (VODAN). This is a collaborative research initiative aimed at promoting the ethical and efficient use of data. Initially focused on combating the novel coronavirus (COVID-19) by enabling access to critical health data in Africa, VODAN has evolved to address broader issues of data ownership, accessibility, and privacy. The network emphasises the creation of FAIR (Findable, Accessible, Interoperable, Reusable) data under well-defined conditions, ensuring that data is used responsibly and benefits local communities. VODAN's architecture includes measures to protect privacy and comply with regulatory frameworks, fostering trust and innovation in data-driven applications. The initiative has expanded to include various African countries and collaborates with international organisations to build a resilient and equitable data ecosystem.

The relationship between VODAN and Leiden University Medical Centre (LUMC) is centred around collaborating to promote FAIR data principles, particularly in health data management and research. LUMC is a key partner in the VODAN network, which aims to enable access to critical health data while ensuring that the data remains securely stored in its original location. This partnership focuses on deploying FDPs in various health facilities, including LUMC, to facilitate data sharing and analysis without relocating the data. The collaboration has led to significant milestones, such as the successful execution of data visiting between African health facilities and LUMC, demonstrating the feasibility of the FAIR data approach. Additionally, LUMC researchers and faculty are actively involved in training and supporting the implementation of FAIR data practices within the VODAN network, the UNA Europa Network, and other collaborations.

The Leiden University Institute of Advanced Computer Science (LIACS) curriculum includes the course Data Science in Practice. This course was particularly productive in promoting innovative solutions using FAIR-linked data. A second course, Regulatory Governance in Data Science, has supported students in investigating the regulatory and ethical frameworks for approaching data science. The quality of data curation determining the quality and representativity of AI is studied as an emerging problem. In this book, the last chapter introduces an undergraduate academic framework for teaching data science, emphasising advanced skills in cultural competence, representativity and social responsibility.

VODAN is a subsidiary of the Africa University Network on FAIR Open Science (AUN-FOS), which promotes open science principles across African universities. It focuses on addressing global inequalities in access to, and the benefits from participation in, innovation and research. The network emphasises the importance of making data Findable, Accessible, Interoperable, and Reusable (FAIR) to support scientific advancements and community resilience. It has established programmes like the Digital Innovation and Skills Hub (DISH), which provides pre-university courses and skill development for vulnerable youth, drop-outs, and refugees, with a strong emphasis on data science and mental health support. The Globalisation, Accountability, Innovation and Care (GAIC) Research Lab drives the network, which explores how innovation can contribute to or disrupt communities and aims to create a more equitable and inclusive research environment in Africa and elsewhere.

FAIR Implementation Network: GO TRAIN, GO BUILD and GO CHANGE

This book is the result of an original meeting held in 2019 in Leiden, when the GO FAIR Implementation Network Africa was established. FAIR Africa of GO FAIR aims to promote adopting and implementing FAIR data principles across Africa. The essence of this network is to foster a collaborative environment that brings together African expertise, both within the continent and in the diaspora, to build an Internet of FAIR Data and Services. This initiative seeks to address the challenges of data governance, ownership, and accessibility within the African context, ensuring that data is used for the common good and contributes to local and global research and innovation.

Key objectives of the Implementation Network FAIR Africa include:

- *Promoting FAIR data principles*: Encouraging the adoption of FAIR data principles to enhance the quality, accessibility, and usability of data in various domains, including health, agriculture, and environmental science.
- *Building capacity and expertise*: Strengthening the capacity of African researchers, policymakers, and practitioners to implement FAIR data practices through training, workshops, and collaborative projects.
- *Fostering collaboration*: Facilitating collaboration between African institutions and international partners to share best practices, tools, and resources for FAIR data implementation.
- *Addressing data governance challenges:* Ensuring that data governance frameworks in Africa align with FAIR data principles, promoting ethical data sharing and use while respecting local regulations and contexts.
- *Supporting Open Science:* Advancing open science initiatives in Africa by making research data more accessible and reusable, thereby enhancing the impact and reach of African research.

The network aims to learn from other GO FAIR Implementation Networks and contribute to the global FAIR data community by sharing insights and findings from African implementations. This collaborative approach is essential for addressing global challenges and ensuring that Africa plays a significant role in the international data ecosystem.

The Implementation Network FAIR Ambassadors was also established in 2019 to strengthen researchers' implementation of FAIR data principles across new countries and sectors. Here are the key objectives of the Implementation Network FAIR Ambassadors:

- Promote FAIR data principles to enhance the visibility and discoverability of research data across various disciplines and countries.

- Foster the reuse of FAIR data to improve the efficiency and effectiveness of research.
- Involve researchers, students, and data stewards from different regions and communities to expand the reach and impact of FAIR data practices.

The network targets PhD students, master's students, bachelor's students, diploma students, data stewards, and seasoned researchers interested in exploring the relevance of FAIR data principles in their specific contexts and new fields.

Ambassadors coordinate with each other and the GO FAIR International Support and Coordination Office to promote the FAIR data movement and GO FAIR approach in their countries. Participants in the network receive training through the GO TRAIN Pillar and the GO FAIR community, exchanging experiences and educating each other on FAIR data practices. The network aims to increase the involvement of representatives from different countries and communities, fostering a global community of FAIR data advocates.

The Implementation Network FAIR Ambassadors play a crucial role in advancing the adoption of FAIR data principles worldwide. By engaging a diverse range of researchers and providing training and support, the network contributes to creating a global ecosystem where data is findable, accessible, interoperable, and reusable, ultimately enhancing the impact and reach of scientific research.

The GO FAIR Pillars 'GO TRAIN', 'GO BUILD', and 'GO CHANGE' are pivotal in advancing the FAIR African Health Data Space, as they align with the growing need to enhance the capacity for managing health data across the continent. GO TRAIN and GO BUILD highlight the importance of equipping individuals, institutions, and governments with the necessary skills and infrastructure to develop and manage privacy-preserving, safe, robust, data-driven, interoperable health systems. By fostering training in data science, digital health tools, and the FAIR data principles, GO TRAIN and GO BUILD help establish the capacity to ensure that the health data generated in Africa is not only accessible but also accurately and efficiently utilised to improve healthcare delivery. GO CHANGE, on the other hand, focuses on transforming healthcare systems and policies to integrate open, interoperable data solutions, encouraging collaboration across countries, sectors, and organisations, and ensuring that systems respond to specific context situations. This approach aims to help create a unified and sustainable health data ecosystem that enables better decision-making and equitable access to healthcare while preserving the privacy and secure reuse of data for research, ultimately driving meaningful change in Africa's health outcomes.

FAIR Data, FAIR Africa, FAIR World: The internationalisation of the Health Data Space

The title 'FAIR Data, FAIR Africa, FAIR World – The Internationalisation of the Health Data Space' was first used for a conference held in August 2024 and forms the basis of this book. The title highlights the significance of adopting FAIR data principles globally, focusing on Africa. This conference emphasised the importance of internationalising health data spaces to ensure data is managed and shared ethically and efficiently. The event aimed to discuss the advancements in FAIR open engineering and linked data technologies, fostering a more integrated and equitable global health data ecosystem. By promoting the FAIR data principles, the conference sought to enhance data stewardship, improve health outcomes, and support research and innovation in healthcare. The internationalisation of the health data space is crucial for addressing global health challenges and ensuring that all regions, including Africa, can benefit from and contribute to advancements in health data management.

The next step was expanding the interaction between FAIR communities, SoLID communities and the Linked Data Spaces communities. These aim to propose data spaces based on distinct containers, in which linked data enable interoperability between the data sources. The SoSy2025 conference, held in Leiden on 24 April 2025, will focus on issues related to the internationalisation of health data spaces, the adoption of FAIR data principles, and the importance of data management in healthcare. The conference will include discussions on how FAIR data principles can be applied to improve healthcare systems, enhance data sharing, and promote global collaboration in health data management and cross-domain interoperability.

Introduction to the chapters

Chapter 2, by Amare, entitled 'Bridging Borders with FAIR Data: Transforming Digital Ecosystems for Maternal Health and Public Health Surveillance in Africa', discusses how a federated data architecture can be implemented to manage patient data across multiple African countries while ensuring FAIR data principles and regulatory compliance. The researcher found that deploying a FAIR-compliant federated system enabled localised, semantically enriched data management, improving clinical decision-making. The study demonstrates that the FAIR framework supports Ownership, Localisation, and Regulatory Compliance (OLR), establishing a viable digital ecosystem for patient data in low-resource settings.

In Chapter 3, titled 'Introducing Data Sovereignty over Patient Data: Patient Data Ownership in Residence of Health Facilities in Kenya', Nalugala, Jati, Amare, Omare, Wairimu, Kahiro, Nandwa, Okeyo, Kinoti, Aktau, Mulingwa, and van Reisen identify a lack of data integration in digital patient care systems and the challenge of ensuring data sovereignty. They argue for federated data handling with localised machine-actionable metadata. The research supports frameworks for patient data control and advocates for legislation for inclusive health data, highlighting the potential of the Africa Health Data Space.

Chapter 4 by Aktau, Amare, van Reisen, Taye, Gebremeskel, Jati, and Plug is called 'GO TRAIN: A Protocol for Metadata Creation for the FAIRification of Patient Data Health Records'. The chapter explores the components required for creating FAIR-compliant data architectures in patient health systems. The conclusion emphasises that most requirements for FAIR data are feasible, although some platforms are unsuitable due to data quality and usability issues. GO TRAIN's role in building capacity for FAIR implementation is crucial.

In Chapter 5, 'De Novo FAIRification: A Literature Review', Lin investigates the existing literature on *De Novo* FAIRification, which automates FAIR data principles during data collection. It contrasts this with *post-hoc* FAIRification, which is labour-intensive and inefficient. The conclusion highlights that while *De Novo* FAIRification offers significant benefits, its practical implementation remains limited, necessitating further research to address existing gaps.

In Chapter 6, 'Federating Tools for FAIR Patient Data: Strengthening Maternal Health and Infectious Disease Surveillance from Clinics to Global Systems', Amare, Taye, Plug, Medhanyie, and van Reisen study the adaptation of the CEDAR Workbench for federated deployment in resource-constrained African environments, ensuring FAIR-OLR compliance. The conclusion demonstrates the feasibility of federated microservices for data sovereignty, even in offline settings, enabling health facilities to act as single sources of truth while enhancing public health data infrastructure.

Chapter 7 by Jati, Amare, Kawu, Nandwa, Taye, and van Reisen, entitled 'GO CHANGE: Adoption of FAIR-OLR Architectures to Support Insights from Patient Health Data Records in Africa', investigates how contextual factors influence the adoption of a FAIR-OLR-compliant patient information system. Using the Unified Theory of Acceptance and Use of Technology (UTAUT) analysis across five African countries, it found strong convergence towards FAIR-OLR principles, legislative changes, and the critical role of dashboards. War experiences emphasised federated data needs, highlighting the necessity of adaptable architectures.

In Chapter 8, 'A Critical Incident Assessment of a FAIR Implementation Study on Patient Health Data in Africa', Kawu, Jati, Schrijver, van Reisen, Tadele, O'Sullivan, and Hederman research how critical incidents impact the implementation of FAIR-compliant health data systems, identifying the facilitators and barriers to their adoption in resource-constrained settings. The authors found that FAIR implementation improves health data accuracy and decision-making but faces challenges like infrastructure deficits. Investments in power solutions, usability enhancements, and capacity-building are essential for success. Stakeholder collaboration and adaptive system design are crucial for realising FAIR's transformative potential in African digital health initiatives.

In Chapter 9, Aktau discusses 'The Potential of Adoption of FAIR Guidelines in Digital Healthcare in Kazakhstan', extending the relevance of the Health Data Space to Asia. She looks at whether the FAIR Guidelines are aligned with Kazakhstan's regulatory and policy framework for health data and how their implementation can support digital health in the country. Aktau concludes that FAIR implementation in Kazakhstan can address technical and policy challenges, enhancing digital healthcare integration and personalised medicine. However, successful adoption requires capacity building, particularly among administrative and political leaders, and adaptations to align FAIR architecture with Kazakhstan's specific data integration needs.

In Chapter 10, titled 'Harmonising Antenatal Care Records Across Four African Countries: A Comparative Analysis and Development of a Standardised Data Model', Amare, Weldu, and Gebremeskel investigate the challenge of inconsistent antenatal care (ANC) data collection across 12 health facilities in Ethiopia, Kenya, Nigeria, and Uganda. Their comparative analysis of DHIS2 templates found that only 4 of 228 variables were captured in all four countries. Authors subsequently developed a FAIR electronic Case Report Form (eCRF) to improve data standardisation and semantic consistency, supported by a harmonised ontology published in BioPortal. The study concludes that DHIS2-based reporting is insufficient for cross-country interoperability and advocates for EMR-based, FAIR-compliant systems that support granular patient data and federated data reuse in support of an African Health Data Space.

In Chapter 11, titled 'Complexities of Recording and Reporting Perinatal Data: A Case Study from Namibia, ' Beatrix Callard maps the processes and challenges of reporting stillbirths and neonatal deaths across five public and private hospitals in Windhoek. The study identifies fragmented data flows, duplicate reporting, and burdensome manual recordkeeping as key barriers to reliable perinatal mortality data. Piloting a low-cost, interoperable dashboard aligned with FAIR principles, the study demonstrates data completeness, accuracy, and timeliness improvements. Key interventions—such as daily event verification, Z-lining of forms, standardised naming conventions, and cross-checking across data streams—reduced duplication and empowered clerical staff to support data management. The chapter argues that local data ownership, simplification of systems, and routine verification are critical for strengthening maternal and neonatal health surveillance in low-resource settings.

In an implementation study in Indonesia, in Chapter 12, Jati, Lasroha and van Reisen investigate 'Enhancing Data Privacy and Availability in Indonesia's SATUSEHAT Platform through FAIR-OLR Principles', Indonesia's health data-exchange platform. The main question is how the FAIR-OLR framework can be adapted to improve patient data privacy, consent, and regulatory compliance in the platform. The researchers conclude that integrating FAIR-OLR into SATUSEHAT addresses gaps in data privacy and patient control. A case study in a local health clinic demonstrated positive results, suggesting that FAIR-OLR can enhance regulatory compliance and patient-centric data exchange within Indonesia's digital health system.

In Chapter 13, Aktau and van Reisen discuss the 'Creation of a FAIR Data Point for a Clinical Trial: The CoHSI2 Dataset'. This study looks at the setup of a FAIR Data Point for clinical trial data. The CoHHSI2 data concerns a clinical trial for a vaccine to protect against schistosomiasis, a neglected disease. The trial is conducted on female participants. The question investigated is how the FAIRification of the CoHSI2 dataset could be achieved using a FAIR-by-increment approach and what challenges arise in deploying a FAIR Data Point. The study successfully implemented a FAIR Data Point through a stepwise FAIRification process, enhancing data verification and reusability. However, challenges remain concerning interoperability testing, access control, and security for sensitive data. Embedding FAIR data principles from the outset is recommended to standardise practices and improve academic transparency.

In Chapter 14, 'Implementation of De Novo FAIRification in Relational Legacy Systems: The Case of the Electronic Medical Record System for Maternal Health in Afya.ke.', Lin discusses whether De Novo FAIRification can be applied to relational database healthcare data, and how it can improve interoperability and data management in healthcare systems. The study demonstrates that De Novo FAIRification can successfully convert relational healthcare data into FAIR-compliant formats, enhancing interoperability and usability. The solution is scalable, efficient, and applicable to low- and middle-income countries, improving healthcare data management, supporting maternal health, and enabling data reuse for various health applications. The study shows that De Novo FAIRification effectively transforms healthcare data into interoperable formats, improving data management for maternal health and infectious diseases. In future, this could help support better surveillance, clinical trials, and pharmaco-vigilance, offering scalable solutions for healthcare systems in low- and middle-income countries.

In Chapter 15, Jati and van Reisen discuss 'Narratives in Public Agenda-setting for FAIR Data and Health Data Management in Africa: Enhancing Maternal Health and Infectious Disease Outcomes'. This study examines how the FAIRification of patient health records in low- and middle-income countries can improve patient care, particularly in maternal health and infectious disease management, and overcome challenges such as data ownership, interoperability, and infrastructure constraints. The study demonstrates that applying FAIR data principles to patient health data can enhance maternal health and infectious disease outcomes. By using Kingdon's model, the intervention led to policy adoption, including in Kenya, where patient data was recognised as strategic and the Afya.ke EMR system was implemented, improving data accessibility and control.

Studying this question further, in Chapter 16, Amare, Weldu, Medhanyie and van Reisen identify how a dynamic, granular approach to patient data can improve the management of infectious diseases in maternal healthcare, particularly in detecting syphilis during pregnancy, and support broader public health applications. In this chapter, 'Testing the Cross-Border Africa Health Data Space: Monitoring, Tracking and Prevention of Mother to Child Syphilis Transmission' the study demonstrates that real-time, patient-level data analysis through a knowledge graph enhances maternal healthcare by enabling cross-facility and cross-country analytics. This approach improves syphilis detection, supports infection surveillance, and enhances maternal and neonatal health outcomes.

In Chapter 17, ‘GO FAIR: Ontology Development of Health Semantics with Cultural Specificity: Traditional Health Practices using Tsebel in Conflict Zones,‘ Kahsay, van Reisen and Lin discuss the response in maternal health management in regions that have suffered from war and destruction of health services. The authors ask the question: How can traditional health practices in conflict-affected areas, particularly for infectious diseases and maternal health, be integrated into digital health systems, and how can FAIR-compliant ontologies support culturally informed healthcare solutions? The study highlights that integrating traditional health practices into FAIR-compliant ontologies can improve maternal health and infectious disease management by enhancing data visibility and interoperability while respecting cultural knowledge and ensuring secure, ethical health interventions in crisis-affected communities.

Turning to the application of FAIR data to administration data handling, in Chapter 18, Utami and van Reisen investigate the question of how FAIR data principles can enhance Indonesia’s One Data Policy to improve cross-government data exchange. The study shows that applying FAIR data principles to Indonesia’s One Data Policy can improve data accessibility, interoperability, and reusability across government agencies, addressing challenges like differing data standards and silos, thereby enhancing data-driven decision-making. The chapter is entitled ‘Satu Data Indonesia and FAIR Data: Advancing Coherent Data Management in Government Administration’.

The Health Data Space is also relevant in the Netherlands. In Chapter 19, Landa-Figueroa and Schrijver discuss the relevance of FAIR guidelines for the elderly care sector. Their chapter is entitled: ‘Integrating the Personal Health Train Methodology into Healthcare Systems for Enhanced Elderly Care in the Netherlands’. The question that the authors address is how the Personal Health Train (PHT) methodology, using the FAIRification process, can enhance data integration in healthcare systems to improve care for elderly patients, particularly those with conditions like dementia. The study concludes that the PHT methodology, by adhering to FAIR data principles, can improve data integration across healthcare organisations, alleviating the workload of elderly care professionals and improving care outcomes for elderly patients with dementia, while maintaining privacy and ethical considerations.

In Chapter 20, 'FAIR Data Implementation for Analysis of Research Data in Human Trafficking and Migration', Smits, Upase, Pandey, Bala and Van Reisen discuss the relevance of data space in the human trafficking domain. The authors ask how the FAIRification process can be applied to a dataset on human trafficking to enhance data analysis and provide new insights into trafficking networks. The study found that the FAIRification process allows the conversion of survivor narratives into machine-actionable data, revealing critical nodes and transit routes within trafficking networks. It enhances scalability, ethical data handling, and interoperability, ultimately advancing human trafficking research by improving data analysis and safeguarding sensitive information. The chapter shows the potential for other domains and cross domain research.

The last chapter, Chapter 21, by Folorunso, Oladipo, and Van Reisen, presents 'A Higher Education Curriculum for Cultural Competence, Representation, and Social Responsibility in AI and FAIR Data practices'. The chapter proposes a curriculum designed to advance FAIR data practices in academia. It emphasises blending theoretical foundations with practical skills, including addressing biases in data and AI, fostering inclusivity, and empowering learners to advocate for ethical, transparent data practices.

Conclusion

The historical patterns of colonial resource extraction are mirrored in contemporary data practices, where data generated in LMICs is collected and controlled by businesses, NGOs, and governments in HICs. This new era of digital colonialism leads to economic and epistemic dependencies, limiting governments' autonomy and agency to respond to their populations' needs. The overlapping challenges of helicopter research and digital colonialism underscore the need for equitable data governance frameworks that prioritise local agency, ensure fair benefit-sharing, and promote collaborative, equitable research practices.

In the context of Africa's emerging Health Data Space, the aspirations of the Mundaneum resonate strongly. As we face increasingly complex health challenges, including emerging infectious disease, climate change, and an increasing chronic disease prevalence, local communities' work to build inclusive, equitable access to high-quality, representative data must be connected and supported. Health data, information, and knowledge should be used to advance health, especially for understudied populations, like pregnant women and their infants and children or migrant and refugee populations.

When coupled with OLR principles through federated data reuse infrastructures, the principles of FAIR data serve as the foundation for creating a truly equitable global healthy ecosystem. In the FAIR African Health Data Space context, these principles are not merely technical standards. They are a call to action for inclusivity, transparency, and collaboration. FAIR data practices can empower communities by ensuring that health data is accessible to all, especially those whose voices have been marginalised. By embracing FAIR-OLR principles, we can bridge divides, foster global partnerships, and strengthen our understanding of health challenges, no matter where they occur. The vision of a FAIR Africa and a FAIR World is one where borders do not confine data and knowledge; rather, data sovereign approaches to data reuse build better individual health, healthcare, and health policy, uniting people and populations in support of the common good. This movement is rooted in the belief that, through secure, localised reuse of interoperable data, we can build a future in which data can build health for every individual and community.

In aggregate, this book's chapters demonstrate the VODAN team's remarkable progress in enabling FAIR-OLR architectures for health data reuse over the last five years. In 2020, when we began our work, we did not understand how to move forward, what workflows to use, what tools were available, what problems could be expected and what the result of the FAIRification would be. As we reflect now, in 2025, these elements are clear. We have plenty of tools to choose from, understand the advantages and limitations of available workflows, and have many use cases and partners. This book maps out the path we took, the way forward, and the challenges ahead. Looking back, we can report that the visionaries who expected the Internet of FAIR Data and Services to be advanced by 2025 were correct, and they were also right to think that from this time onwards, efforts should focus on user-friendly applications and tools.

In the work ahead, we must bring together and learn from the experiences of Solid Pods, Linked Data Spaces, FAIR Data Points, and FAIR-OLR data reuse. Through collected action, we can ensure the security of data spaces based on federated FAIR data architectures, respond to emerging data privacy legislation, connect data siloes, and address other wicked problems needed to build investment and trust in health data capture and reuse.

When the VODAN FAIR data community started its journey five years ago, we were led by the mantra of GO BUILD, GO CHANGE and GO TRAIN. GO BUILD has assured us that a FAIR Data Space is feasible, desirable, and doable. GO CHANGE has given us confidence that GO FAIR is context-sensitive, adaptable to specific places and times, and capable of structuring science and allowing new phenomena to be made visible. GO TRAIN then remains perhaps the most important of all. Training provides the basis for trust, for understanding, for accountability. Training allows for the right questions to be asked; it removes the mist of data-handling and makes it a transparent process. Training allows the proliferation of FAIR data by increasing the number of competent participants. It ensures that data handling integrates seamlessly with machine learning (ML) and AI and increases understanding that FAIR data curation is, in fact, a prerequisite for these things. Training will encourage the participation of a new generation in the Internet of FAIR data and Services. As we move into the next five years of this decade, keeping emphasis on GO BUILD, GO CHANGE and GO TRAIN will guarantee that we will arrive in a space where FAIR Data will be normalised, FAIR Africa will be a natural concept, and a FAIR World will be in reach.

References

Abrahams, L., & Mbanaso, U. (2017, May 2–4). State of Internet security and policy in Africa [Conference presentation]. The First African Academic Network on Internet Policy, International Institute of Tropical Agriculture (IITA) Resort, Ibadan, Nigeria.

Africa Health Data Space. (2024). *The African Health Data Space*. https://www.africanhealthdataspace.org/

Callard, B. (2025) Complexities of recording and reporting perinatal data: A Case Study from Namibia. In Van Reisen, M., Amare, S. Y., Maxwell, L. & Mawere, M. (Eds.), *FAIR data, FAIR Africa, FAIR world: Internationalisation of the Health Data Space*. Bamenda: Langaa

Carroll, S. R., Garba, I., Figueroa-Rodríguez, O. L., Holbrook, J., Lovett, R., Materechera, S., Parsons, M., Raseroka, K., Rodriguez-Lonebear, D., & Rowe, R. (2020). The CARE Principles for Indigenous Data Governance. *Data Science Journal*, *18*(1), 43. https://doi.org/http://doi.org/10.5334/dsj-2020-043

Dutta, M., del Pilar-Labarda, M., Kpokiri, E., Thwaites, L., & Clark, J. (2023). How to ensure equitable research partnerships in global health. In (Vol. 381): British Medical Journal Publishing Group.

Kaul, I., Conceição, P., Le Goulven, K., & Mendoza, R. (Eds.). (2002). *25 questions and answers*. In *Providing global public goods: Managing globalization* (pp. 1–19). Oxford University Press.

McIntosh, K., Messin, L., Jin, P., & Mullan, Z. (2023). Countering helicopter research with equitable partnerships. *The Lancet Global Health*, *11*(7), e1007-e1008.

Mundaneum. (2023). *Mundaneum. Machine to think the world.* Mundaneum. Available at: https://mundaneum.org/en/activite/mundaneum-machine-to-think-the-world/ (Accessed 5 April 2025)

Odeny, B., & Bosurgi, R. (2022). Time to end parachute science. In (Vol. 19, pp. e1004099): Public Library of Science San Francisco, CA USA.

Otlet, P. (2015). *Le livre sur le livre: Traité de documentation*. Facsimile of the original 1934 edition. Les Impressions Nouvelles.

Peeters, B. (2015). Les deux vies de Paul Otlet. In: Otlet, P. (2015). *Le livre sur le livre: Traité de documentation*. Facsimile of the original 1934 edition. Les Impressions Nouvelles.

Sambra, A. V., Mansour, E., Hawke, S., Zereba, M., Greco, N., Ghanem, A., Zagidulin, D., Aboulnaga, A., & Berners-Lee, T. (2016). Solid: a platform for decentralized social applications based on linked data. *MIT CSAIL & Qatar Computing Research Institute, Tech. Rep., 2016*.

Ramhit, K. (2024) FAIR & CARE Principles: How Persistent Identifiers (PIDs) Interconnect. ORCID US Community. Available at: https://orcidus.lyrasis.org/2024/10/23/fair-care-principles-how-persistent-identifiers-pids-interconnect/

Van Reisen, M., Yohannes, S., Plug, R., Tadele, G., Gebremeskel, T., Kawu, A.A., Smits, K., Woldu, L.M., Stocker, J., Heddema, F., Folorunso, S., Kievit, R., Medhanyie, A.A. (2024). Curation of Federated Patient Data: A Proposed Landscape for the Africa Health Data Space. In: Agbotiname Lucky Imoize, Professor Mohammad S. Obaidat, and Professor Houbing Song (eds.) (2024). *Federated Learning for Digital Healthcare Systems.* Series: Intelligent Data-Centric System. Amsterdam: Elsevier. https://doi.org/10.1016/B978-0-443-13897-3.00013-8

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., . . . Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, *3*(1), 160018. https://doi.org/10.1038/sdata.2016.18

World Health Organization (WHO). (2021). *Global strategy on digital health 2020–2025*. World Health Organization. https://www.who.int/publications/i/item/9789240020924