From Local to Global: Lessons from the HEALTHWISE Consortium on Federating Secure Health Data Infrastructure

COLLECTION:
INTERNATIONAL DATA WEEK

PRACTICE PAPERS

TIM BECK
JASON FERRIS
ALAIN-DOMINIQUE GORSE
NEERJA KARNANI
SUMIR PANJI
CLAIR SULLIVAN
SIMON THOMPSON
ESMOND URWIN
XING YI WOO
PHILIP R. QUINLAN

*Author affiliations can be found in the back matter of this article

ABSTRACT

Trusted Research Environments and related secure data platforms are becoming the primary means by which sensitive health and biomedical data are made accessible for research without transferring data. However, the technical and legal demands on these infrastructures are considerable, and the challenge of enabling them to federate and interoperate across institutions, nations, and continents remains largely unresolved. This paper describes the current state of the art in secure health data infrastructure across four countries, drawing on the experience of members of the HEALTHWISE consortium (Healthcare Enhancement and Learning Through Worldwide Interoperable Systems and Equity). Taken together, the case studies illuminate a set of globally recurring tensions among what is technically possible, what is publicly acceptable, and what is operationally useful which reflect deeper structural challenges in the science of data. We argue that resolving these tensions requires not only continued technical innovation but a more deliberate approach to governance, standards adoption, training, and community building at a global scale.

CORRESPONDING AUTHOR:
Tim Beck
Centre for Health Informatics, School of Medicine, University of Nottingham, United Kingdom; NIHR Nottingham Biomedical Research Centre, University of Nottingham, Nottingham, United Kingdom
tim.beck@nottingham.ac.uk

KEYWORDS:
Trusted Research Environments; Federated data analysis; Health data governance

TO CITE THIS ARTICLE:
Beck, T., Ferris, J., Gorse, A.-D., Karnani, N., Panji, S., Sullivan, C., Thompson, S., Urwin, E., Woo, X.Y. and Quinlan, P.R. 2026 From Local to Global: Lessons from the HEALTHWISE Consortium on Federating Secure Health Data Infrastructure. *Data Science Journal*, 25: 41, pp. 1–11. DOI: https://doi.org/10.5334/dsj-2026-041

1. INTRODUCTION

The responsible use of sensitive health data for research, such as electronic health records, genomic and phenotypic data, and administrative datasets, is one of the defining data science challenges of our era. On one hand, the potential value of such data is enormous. Linked, longitudinal health records can accelerate drug discovery, improve clinical decision-making and address longstanding inequities in population health. However, the technical and legal demands imposed by the need to protect individual privacy are substantial, and these demands intensify when data must cross institutional, state, or international boundaries.

The dominant response to this challenge has been the development of Trusted Research Environments (TREs), also termed Secure Data Environments, Safe Havens, Data Enclaves, or Secure e-Research Platforms depending on jurisdiction. These are highly governed, technically secured systems in which data remain under the control of the data custodian while authorised researchers bring their analysis to the data, rather than moving data to the researcher, an approach described as data visitation or visiting (Weise *et al.*, 2022). The Five Safes Framework, defining Safe Projects, Safe People, Safe Settings, Safe Data, and Safe Outputs, has emerged as a widely adopted conceptual basis for evaluating and communicating the safety of such arrangements (Ritchie, 2017).

However, TREs have not converged on a single model. They have evolved in parallel across different national contexts, shaped by different legal frameworks, funding structures, clinical data systems, and public expectations. The result is a landscape of infrastructures that share common principles but are difficult to federate in practice. The HEALTHWISE consortium (Healthcare Enhancement and Learning Through Worldwide Interoperable Systems and Equity) was formed to address this challenge, bringing together TRE providers from the United Kingdom, Australia, Japan, Singapore, South Africa, and Taiwan to develop shared approaches to transnational and transcontinental federation. The consortium has proposed a maturity model framework for evaluating the governance, operational, and technical readiness of federated TRE networks (Luong *et al.*, 2026), which complements the country-level perspectives presented here.

This paper draws on the experiences of five HEALTHWISE member organisations to provide a comparative account of secure data infrastructure across four countries. Section 1.1 describes our approach, and Section 2 presents the five case studies. Section 3 draws out the cross-cutting themes these experiences reveal and reflects on what they tell us about the broader science of data, and Section 4 concludes the paper.

1.1. APPROACH

This is a practice paper grounded in the direct experience of those who design, operate, and use the infrastructures described, rather than a systematic review. The five organisations were selected because they presented their infrastructures at the HEALTHWISE session at International Data Week 2025. Together they span four countries on four continents and a range of infrastructure models, from a statewide integrated electronic medical record supported by a university TRE to a pan-African federated data science platform. Each case study was written by authors directly involved in developing, operating, or using the infrastructure concerned, drawing on the material presented at the session and supplemented by published literature and project documentation.

To enable comparison, each case study describes the same six dimensions: the context in which the infrastructure developed, its governance model, the data standards it has adopted, its technical architecture, its approach to federation, and the principal challenges encountered. These dimensions are summarised for all five organisations in Table 1. Cross-cutting themes were identified by comparing the case studies across these dimensions and noting issues that recurred in two or more organisations. The tripartite framing presented in Section 3.1 emerged from this comparison and from the discussion at the International Data Week session. Because the authors are practitioners reporting on their own infrastructures, the account reflects an insider perspective, and the five organisations are not intended to be representative of TREs globally.

	NOTTINGHAM (UK)	QUEENSLAND (AUSTRALIA)	CAPE TOWN (SOUTH AFRICA)	A*STAR (SINGAPORE)	SWANSEA (UK)
Context and drivers	National investment through HDR UK and DARE UK; NHS England requires re-use of patient data without specific consent to take place within a SDE	Statewide single electronic medical record (iEMR) providing a unified source of routinely collected health data	Fragmented regulation across multiple jurisdictions; historical data extraction without local benefit; DS-I Africa initiative funded by the US National Institutes of Health	City-state of approximately 6.2 million people with three health clusters; regional hub for cross-border biomedical research	Over a decade providing secure access to anonymised health and administrative data, predominantly from Wales
Governance model	Federation defined as independent organisations retaining control of their own data, services and platforms; national SDE policy	SMART Team hierarchical oversight with eHealth Queensland; conjoint appointments; Five Safes Framework implemented in UQTRE	GA4GH Passports for federated authentication and authorisation	TRUST hosted by A*STAR in partnership with the National Research Foundation and Ministry of Health	Five Safes Framework underpins SeRP and SAIL Databank governance; governance defined by data and operational context; governed output checking
Data standards	OMOP CDM; open-source tools for concept mapping and rule generation	OMOP CDM applied across thousands of iEMR source tables	GA4GH standards: Data Repository Service, Data Connect, Workflow Execution Service, and Passports	OMOP CDM through the National OMOP Working Group	GA4GH Task Execution Service standard, implemented in Five Safes TES
Technical architecture	OMOP mapping tooling available to the national SDE Network	UQTRE with segregated project workspaces, audit logging and governed output release	Cloud-agnostic federated open data science platform	TRUST platform; Biomedical Data Hub; PET Zoo microservices	SeRP platform with multi-modality support, output checking and high-performance computing; deployed in Canada and Australia
Approach to federation	Federation of TREs and SDEs through agreed standards and protocols	NINA: federated analytics and machine learning across seven sites; Australian Health Data Evidence Network (OHDSI-style)	GA4GH Starter Kit proof of concept across four African nodes	PETs: federated learning, secure multi-party computation, trusted execution environments, and homomorphic encryption	DARE UK blueprint components: TELEPORT and Five Safes TES
Principal challenges	Data transformation burden on custodians; researcher unfamiliarity with OMOP; ensuring SDEs are useful to researchers	Governance, access, and standardisation of highly complex clinical data	Learning curve for GA4GH standards; uneven research and technical capacity	Legal and regulatory fragmentation; interoperability gaps; accountability for federated analysis	Serving diverse technical and governance requirements at scale
Most relevant themes in Section 3	3.1 Tripartite tension 3.2 Standards as infrastructure	3.2 Standards as infrastructure 3.3 Governance as an enabler	3.2 Standards as infrastructure 3.4 Equity	3.2 Standards as infrastructure 3.3 Governance as an enabler	3.2 Standards as infrastructure 3.3 Governance as an enabler

Table 1 Comparison of the five HEALTHWISE infrastructures across the common dimensions used to structure the case studies, and the cross-cutting themes each case study most directly informs.

2. SECURE DATA INFRASTRUCTURE: FIVE ORGANISATIONAL PERSPECTIVES

Each of the following case studies is structured around the six dimensions set out in Section 1.1. Table 1 summarises these dimensions across the five organisations and indicates where each case study informs the cross-cutting themes discussed in Section 3.

2.1. UNIVERSITY OF NOTTINGHAM, UNITED KINGDOM: FEDERATION OF TRES AND NATIONAL HEALTH SERVICE SECURE DATA ENVIRONMENTS

The United Kingdom has made substantial national investments in federated health data infrastructure through programmes including Health Data Research UK (HDR UK) and Data and Analytics Research Environments UK (DARE UK). The University of Nottingham's Centre for Health Informatics plays a leading role spanning adoption of common data models, open-source tooling, and federation architectures.

The UK policy environment has been shaped by a significant shift. National Health Service (NHS) England has moved towards a position in which re-use of patient data without specific consent must take place only within a Secure Data Environment (SDE). This has driven rapid investment in SDE infrastructure, while creating pressure to ensure the resulting systems are genuinely useful to researchers, requiring engagement with the research community as much as technical build.

The Observational Medical Outcomes Partnership Common Data Model (OMOP CDM, or OMOP for short) has been adopted across the NHS Research SDE Network as a key standardisation layer to facilitate federation. Maintained by the Observational Health Data Sciences and Informatics (OHDSI) community, the OMOP CDM is a person-centred relational data model that organises clinical data into a common set of tables, such as conditions, drug exposures, procedures, and measurements, and maps local source codes to standardised vocabularies including Systematized Nomenclature of Medicine - Clinical Terms (SNOMED CT) and Logical Observation Identifiers Names and Codes (LOINC) (Hripcsak *et al.*, 2015). Because each site holds its data in the same structure and uses the same concepts, an analysis can be written once and executed at every site without the underlying data being shared. Common data models make federation easier by providing a shared vocabulary, but they also introduce friction: data custodians face complex extraction, transformation, and loading processes, and researchers may be unfamiliar with or uncertain about standardised representations of clinical data. Open-source tooling, including tools to assist with OMOP concept mapping and rule generation (Cox *et al.*, 2025), has been developed to lower this barrier, alongside integration with the national SDE Network.

A useful framing for understanding the UK approach is to define federation as independent organisations working together through agreed-upon standards and protocols while retaining control over their own data, services, and platforms. This definition foregrounds autonomy alongside interoperability, a balance easily lost when federation is treated as a purely engineering problem. Data federation has been estimated to be worth up to £319 billion to the UK economy by 2050, and many critical research questions cannot be answered without greater interoperability between existing but unlinked datasets and infrastructures (DARE UK, 2024b).

2.2. UNIVERSITY OF QUEENSLAND, AUSTRALIA: QUEENSLAND HEALTH DATA AND THE SMART TEAM/UQTRE

Queensland, Australia, holds a distinctive asset in its Integrated Electronic Medical Record (iEMR), a single clinical information system deployed statewide, providing a unified source of routinely collected health data. Realising its research value has required overcoming substantial challenges around governance, data access, standardisation, and secure analysis of sensitive clinical information. To address these, the University of Queensland (UQ), coordinated through its Data Science Collaborative Research Platform, brings together two complementary infrastructures: the SMART Team, supporting data access governance, extraction, and standardisation, and the UQ Trusted Research Environment (UQTRE), a governed secure digital platform for analysing sensitive data.

The SMART Team provides the data governance, operational coordination, and clinical informatics expertise required to translate routinely collected health data into research and

service improvement activities. Its governance model operates through a hierarchical oversight structure spanning an eHealth Queensland Executive Committee, project control groups, and specialty user groups covering pharmacy, laboratory, medical, care delivery, and health information management, established through years of contractual negotiation and conjoint appointments embedding SMART Team staff within eHealth Queensland. With the SMART Team now in place, the resource constraints to accessing iEMR have been dramatically reduced (Austin *et al.*, 2024).

UQTRE implements the Five Safes Framework through a combination of technical and procedural safeguards, including project-specific fine-grained access controls, multi-factor authentication, secure remote desktop access, comprehensive audit logging, and governance-led approval processes for data ingress and output release. Research projects are provisioned within fully segregated workspaces, ensuring strict isolation of data, users, analytical environments, and outputs between projects. The utilisation of UQTRE means the data custodians (e.g., Queensland Health) can have reassurance regarding the safety of the data.

While these two infrastructures mean iEMR data are now safely accessible, the UQ team have also implemented a 'transformative' work package to ready iEMR data for research, transforming Queensland Health data into the OMOP CDM and standardising a highly complex clinical environment of thousands of source tables into a research-ready common analytical structure. This has enabled participation in the Australian Health Data Evidence Network (AHDEN), an OHDSI-style federated research network that allows institutions to retain control of their data while contributing to large-scale collaborative analyses using shared data standards, analytical approaches, and open-source tools. Through AHDEN, Queensland is now aligned with a growing national and international community built around the OMOP CDM and the OHDSI ecosystem.

Building on this foundation, UQ is contributing to a new generation of federated research initiatives in which data remain within local governance boundaries while analyses can be executed across multiple sites. These include the National Infrastructure for federated learNing in digitAl health (NINA), which is developing privacy-preserving federated analytics and machine learning capabilities across four disease areas and seven sites involving 23 partners nationally.

2.3. UNIVERSITY OF CAPE TOWN, SOUTH AFRICA: ELWAZI AND THE DATA SCIENCE FOR HEALTH DISCOVERY IN AFRICA

Across the African continent, data governance is shaped by a highly fragmented regulatory landscape, with differing legal frameworks on data sovereignty across multiple jurisdictions. Research and technical capacity to build and operate TRE-equivalent platforms is unevenly distributed, and historical patterns of data extraction, in which African populations contribute data without benefitting from the resulting science, have generated well-founded concerns about research equity.

The eLwazi Open Data Science Platform (ODSP), developed at the University of Cape Town, addresses these challenges within the Data Science for Health Discovery and Innovation in Africa (DS-I Africa) initiative, funded by the US National Institutes of Health. DS-I Africa aims to leverage data science technologies to transform biomedical and public health research and develop solutions that lead to improved health for individuals and populations across the continent.

eLwazi provides a federated open data science platform built on Global Alliance for Genomics and Health (GA4GH) open standards, including the Data Repository Service for cloud-agnostic data access; Data Connect, which enables federated data querying across heterogeneous data stores; the Workflow Execution Service, allowing the same analytical workflow to run reproducibly across multiple computing environments; and Passports for federated authentication and authorisation. By building on open, community-adopted standards rather than proprietary solutions, eLwazi ODSP positions African health data infrastructure to interoperate with platforms in Europe, North America, and Asia.

The GA4GH Starter Kit has been central to making these standards accessible, providing a proof of concept for data discovery and federated analysis between four nodes within Africa. Implementing community-developed GA4GH standards presents a significant learning curve for research software engineers without prior exposure to them. Investment in training and expert support is treated as a prerequisite for successful deployment rather than an optional extra, a distinctive feature of the eLwazi approach alongside its technical infrastructure.

2.4. A*STAR, SINGAPORE: SECURE AND CONNECTED DATA SCIENCE

Singapore is a highly connected city-state with a multi-ethnic population of approximately 6.2 million, and with three major health clusters comprising eleven public hospitals, engaged in ambitious national and cross-border data science programmes.

Singapore has established complementary national platforms and capabilities that together strengthen its secure, interoperable health data ecosystem. Trusted Research and real world-data Utilisation and Sharing Tech (TRUST), a national platform hosted by A*STAR (Agency for Science, Technology and Research) in partnership with the National Research Foundation and Ministry of Health, enables anonymised real-world data and research data to be brought together, accessed, and used securely to advance health data partnerships, working closely with the Biomedical Data Hub, which curates and maintains strategic deidentified research datasets and cohorts. Building on this national infrastructure, the A*STAR Bioinformatics Institute leads the National OMOP Working Group, supporting data harmonisation and interoperability standards alongside computational research groups with disease-domain expertise in multimodal data integration, analytics, and synthetic data generation. In parallel, the A*STAR Institute of Advanced Intelligence and Computing is developing Privacy Enhancing Technologies (PETs), including federated learning, secure multi-party computation, trusted execution environments, and homomorphic encryption, enabling insights without data leaving its jurisdiction of origin. Its PET Zoo makes these technologies available as microservices, including PENGUIN, a Privacy Enhanced Guarded Intelligence Environment for analysis; PENTHER, for Privacy Enhanced Trusted Homomorphic Encryption; and PENDA, for Privacy Enhanced Data Aggregation. Together these support secure AI development, cross-institutional collaboration, and trusted data use.

Singapore has identified a set of core challenges in cross-border data science: legal and regulatory fragmentation; privacy and ethical concerns; interoperability gaps between institutional systems; accountability mechanisms for federated analysis; and technical infrastructure limitations. These challenges are not unique to Singapore, but Singapore's position as a regional hub for biomedical research, emphasis on translation, and its active participation in networks including HEALTHWISE make it a productive site for developing and evaluating practical responses to them.

2.5. SWANSEA UNIVERSITY, UNITED KINGDOM: SeRP AND SAIL, SCALE, DIVERSITY, AND THE EVOLVING TRE

Swansea University's Secure e-Research Platform (SeRP) and the Secure Anonymised Information Linkage (SAIL) Databank represent one of the most operationally established TRE infrastructures globally, with over a decade of experience providing secure access to anonymised health and administrative data, predominantly from Wales, to academic, government, and health organisation users.

The two are distinct but interdependent. SAIL is a population-scale databank of anonymised, individual-level records linked across health and administrative sources for the population of Wales, and it brings its own extensive, independent information governance framework, which determines on legal and ethical grounds who may access the data it holds. SeRP provides the platform within which those data are hosted and analysed, supplying the secure environment, services, and operational controls. SeRP has supported landmark studies, including analyses informing the UK Scientific Advisory Group for Emergencies during the COVID-19 pandemic, and now operates internationally, supporting multiple large-scale TREs with diverse technical and governance requirements.

SeRP is understood as a full infrastructure platform defined by governance, data, and operational context, with governance for both SeRP and the SAIL Databank structured around the Five Safes Framework. It supports a diverse range of data types and analytical requirements, providing multi-modality support, data acquisition management systems, output checking, and high-performance computing environments. A 2026 roadmap centres on re-engineered data management, enhanced multi-modality support, and the development of collaborative spaces for cross-institutional research. The SeRP model has also been deployed internationally, including in Canada and Australia, demonstrating that TRE architectures developed in one context can be adapted and extended elsewhere.

Drawing on this operational experience, Swansea has been central in DARE UK efforts to develop proof-of-principle implementations for federation, and these now feature as components of DARE UK's federated architecture blueprint (DARE UK, 2024a). The blueprint sets TELEPORT and Five Safes Task Execution Service (TES), which appears in the blueprint under its earlier name TRE-FX, alongside further components. TELEPORT enables direct query federation, sending a query to multiple TREs with results returned without data leaving each environment (Orton *et al.*, 2023). Five Safes TES builds on the GA4GH TES standard to enable indirect, agent-mediated analysis, where analytical agents running within TREs return results. Pop-up, project-specific TREs can be instantiated rapidly, maintaining consistent governance while reducing the overhead of persistent infrastructure; federated learning, exchanging model parameters rather than data, offers a longer-term route to cross-boundary analysis. Taken together, these components describe how an established single-site platform such as SeRP becomes a node in a network, and how population data held in a databank such as SAIL can be analysed alongside data held elsewhere while remaining subject to the governance framework under which they are held. The 'Multi TRE: Single Pane of Glass' vision, querying multiple TREs as if a single database, encapsulates the field's ambition of preserving each TRE's sovereignty and governance while making their collective analytical potential accessible as a unified resource.

3. CROSS-CUTTING THEMES AND WIDER SIGNIFICANCE

3.1. THE TRIPARTITE TENSION: POSSIBLE, ACCEPTABLE, USEFUL

Perhaps the most generalisable insight from HEALTHWISE members' collective experience is that building effective data infrastructure requires simultaneous navigation of three distinct domains: what is technically possible, what is publicly and legally acceptable, and what is operationally useful to researchers. These domains do not automatically align, and progress in one can create friction in another. A technically sophisticated federation architecture may not be acceptable to data custodians concerned about sovereignty, a concern that shapes the fragmented regulatory landscape in which eLwazi operates (Section 2.3); a data model that satisfies governance requirements may not be useful to researchers who prefer to work with raw clinical and genomics data, which is the friction that accompanies OMOP adoption in the UK and Queensland (Sections 2.1 and 2.2); and a policy environment that restricts data movement may outpace the availability of technical alternatives such as federated learning, as the NHS England Secure Data Environment requirement and Singapore's cross-border constraints both illustrate (Sections 2.1 and 2.4).

This tripartite framing has broad applicability as an analytical tool for the design and evaluation of any data infrastructure (Figure 1). Where only two of the three conditions are met, predictable failure modes emerge: possible and publicly acceptable but not useful infrastructure remains governed but unused; possible and useful but not acceptable infrastructure represents ungoverned innovation; and acceptable and useful but not yet possible infrastructure remains aspirational but unfeasible. These conditions are best understood as applying to a given capability at a given moment rather than to an organisation as a whole. All five HEALTHWISE members satisfy the three conditions for their established services, which is why those services are in operational use, and each is also working towards capabilities that do not yet satisfy all three.

Which condition is the limiting factor differs by setting. For federated analysis across African jurisdictions, the limiting factor has been what is technically possible without standards expertise in place, which is why eLwazi treats training and expert support as prerequisites (Section 2.3). For cross-border analysis from Singapore, the limiting factor is acceptability under fragmented legal and regulatory regimes, which PETs are intended to address (Section 2.4). For OMOP adoption in the UK and Queensland, the limiting factor is usefulness, since standardised representations impose effort on custodians and can be unfamiliar to researchers, prompting investment in open-source mapping tools and transformation work (Sections 2.1 and 2.2). Identifying which of the three conditions is currently limiting progress is therefore more informative than classifying an infrastructure as a whole. It suggests that data science, as a discipline, needs methods not only for building systems but for assessing and communicating their position across all three dimensions, and for identifying where the limiting factor lies at any given stage of development.

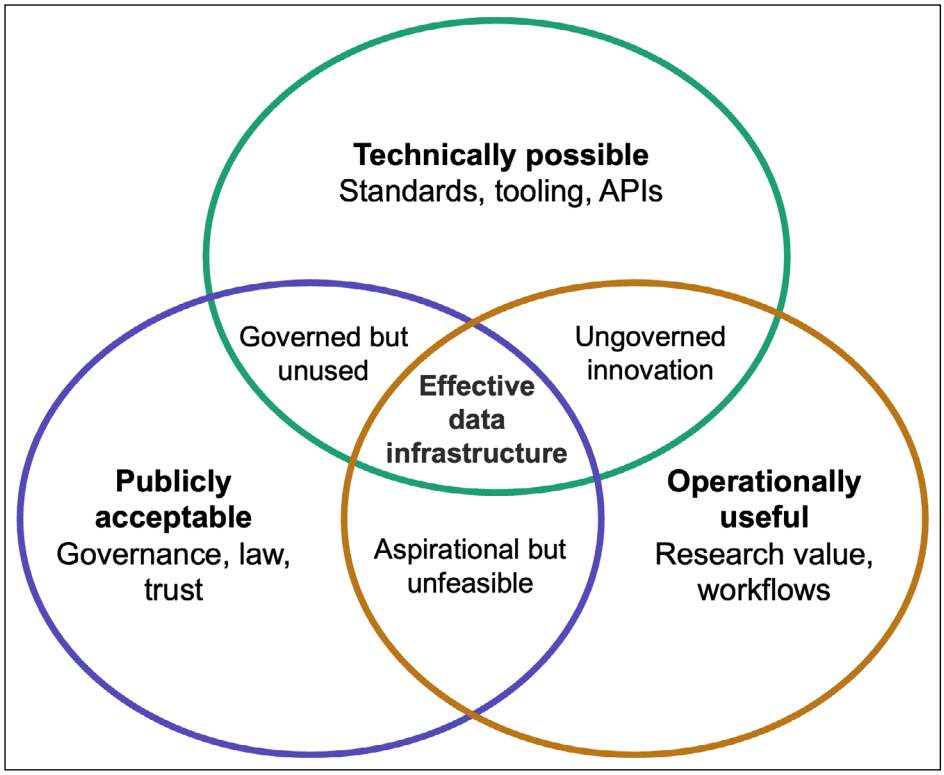


Figure 1 The tripartite tension in health data infrastructure. Effective infrastructure requires alignment across all three domains.

3.2. STANDARDS AS INFRASTRUCTURE

All five case studies demonstrate sustained engagement with open international standards as the primary mechanism for enabling interoperability: the OMOP CDM in Nottingham, Queensland, and Singapore (Sections 2.1, 2.2 and 2.4), the Five Safes Framework in the UK and Australian environments (Sections 2.2 and 2.5), and GA4GH standards in both eLwazi and Swansea's Five Safes TES (Sections 2.3 and 2.5). This represents a significant shift from earlier approaches in which individual institutions built bespoke solutions. As well as being technical specifications, standards also help to encode agreed-upon ways of working, create common vocabularies that enable communication across institutional and national boundaries, and lower the barrier to participation for new entrants through adoption of existing standards.

Published reference architectures play a similar role at the level of the whole infrastructure. The Open Source Secure Data Infrastructure and Processes (OSSDIP) blueprint, for example, documents the architecture, roles, and processes of a secure data infrastructure together with an open-source reference implementation, and adopts data visiting, in which analysts work with data inside a controlled environment rather than receiving a copy, as the default access mechanism (Weise *et al.*, 2022). At the federation level, the DARE UK Federated Architecture Blueprint (DARE UK, 2024a) provides an equivalent reference point for connecting multiple TREs. Blueprints of this kind give new entrants a tested starting point rather than requiring each organisation to design its infrastructure from first principles, and they supply a common vocabulary of components, roles, and processes against which existing TREs can describe and compare themselves. Among HEALTHWISE members, the deployment of the SeRP model in Canada and Australia and the use of the GA4GH Starter Kit to establish federated nodes within eLwazi show how reusable reference implementations lower the barrier to entry in practice.

At the same time, there are costs associated with adopting standards. Learning curves are steep, as eLwazi found in implementing GA4GH standards with research software engineers new to them (Section 2.3). Mapping existing data to standardised models is laborious, as in the transformation of thousands of iEMR source tables in Queensland (Section 2.2). The value of standards may also not be immediately apparent to data custodians, who bear the implementation cost without necessarily seeing the federation benefit, which is the friction the UK open-source concept mapping tools are intended to reduce (Section 2.1). The data science community has work to do in developing better tools, training resources, and incentive structures to make standards adoption more manageable and demonstrating through concrete examples what becomes possible when standards are in place.

3.3. GOVERNANCE AS AN ENABLER, NOT A CONSTRAINT

Robust governance, instead of being seen as an obstacle to data access, is what makes data access possible at all. The SMART Team's multi-layered oversight structure (Austin et al., 2024; Section 2.2), SeRP's role in national COVID-19 policy analysis (Section 2.5), and Singapore's national biomedical data framework (Section 2.4) all illustrate that trust in a data infrastructure depends on the visibility and rigour of its governance as much as on its technical security. The Five Safes Framework has proved durable precisely because it provides a language for governance that is comprehensible to data custodians, researchers, regulators, and the public alike (Green & Ritchie, 2023).

The challenge is that governance frameworks developed in one jurisdiction do not automatically align to another, as the highly fragmented data law landscape across African countries illustrates (Munung *et al.*, 2024). Other contrasting examples include the UK's NHS-centric governance (Section 2.1) and Singapore's national biomedical framework (Section 2.4). Progress on transnational federation will require not only technical interoperability but governance interoperability with shared frameworks for assessing and communicating trustworthiness across jurisdictions. The HEALTHWISE consortium has begun to address this need directly, proposing a six-domain maturity model, spanning governance, operations, data management and security, infrastructure, research alignment, and outreach, for assessing the readiness of TREs joining federated networks (Luong *et al.*, 2026).

3.4. EQUITY AND THE GLOBAL DISTRIBUTION OF DATA INFRASTRUCTURE CAPACITY

The HEALTHWISE consortium's membership makes visible a dimension of the data infrastructure challenge that is easily overlooked in discussions centred on the experiences of high-income countries, namely the highly uneven global distribution of capacity to build and operate secure data platforms. The eLwazi platform addresses this directly through knowledge transfer, training investment, and the deliberate adoption of open standards that can be implemented without proprietary dependencies (Section 2.3). Capacity is not the only dimension of equity. Where data are collected from communities that have historically had little control over their use, the question is also who governs those data and who benefits from them. The Collective Benefit, Authority to Control, Responsibility, Ethics (CARE) Principles for Indigenous Data Governance address this directly, extending ethical frameworks beyond privacy to encompass Indigenous data sovereignty and the rights of data communities to govern and benefit from research conducted using their data (Engstrom *et al.*, 2024; Watego *et al.*, 2025). Applied alongside the Findable, Accessible, Interoperable, Reusable (FAIR) data principles (Wilkinson *et al.*, 2016), they bear on infrastructure design as much as on research ethics, since decisions about where data are held, who may query them, and how access is authorised determine in practice whether such rights can be exercised.

These equity considerations are central to the science of data. Any global health data infrastructure in which the capacity to generate, store, govern, and analyse data that is concentrated in a small number of wealthy countries will produce systematically biased science. The HEALTHWISE consortium's explicitly global composition is one response to this and the deliberate adoption of open standards and open-source tooling across all member institutions, summarised in Table 1, is another.

4. CONCLUSION

The successes achieved by HEALTHWISE member organisations are substantial. Populations are linked across whole health systems, federated queries are executed across national TRE networks, and open standards are enabling researchers across multiple continents to work with shared analytical frameworks. The challenges surrounding these successes include governance processes that require years of negotiation, standards that impose significant implementation burdens, and legal fragmentation that prevents data from crossing borders even when all parties would benefit.

What the collective experience of HEALTHWISE members reveals is that these are not simply engineering problems awaiting better technical solutions. They reflect deep structural features of how knowledge, power, and trust are distributed in global health research. Addressing them

requires data science that is as attentive to governance, policy, and equity as it is to algorithms and architecture, and a community of practice that is genuinely international in composition and concerns. The HEALTHWISE consortium represents one attempt to build such a community; the broader challenge is embedding these principles in the institutions, funding structures, and training programmes that shape the next generation of data infrastructure.

ACKNOWLEDGEMENTS

The authors thank the participants of the International Data Week 2025 session 'The Highs and Lows of Providing Digital Infrastructure to Enable Safe Access to Sensitive Data for Research' for their contributions to the discussions that informed this paper. We are grateful to all members of the HEALTHWISE consortium for their collective effort in devising the content of the session and for their ongoing collaboration in advancing secure and federated health data infrastructure internationally.

FUNDING INFORMATION

TB, ST, and PQ are supported by Health Data Research UK (HDR-23002), which is funded by the Medical Research Council (UKRI), the National Institute for Health Research, the British Heart Foundation, Cancer Research UK, the Economic and Social Research Council (UKRI), the Engineering and Physical Sciences Research Council (UKRI), Health and Care Research Wales, Chief Scientist Office of the Scottish Government Health and Social Care Directorates, and Health and Social Care Research and Development Division (Public Health Agency, Northern Ireland). CS is supported by Medical Research Future Fund (MRFF). XYW is supported by Singapore Translational Cancer Consortium, Industry Alignment Fund – Industry Collaboration Project (IAF-ICP Grant No. H24-MCP0058), National Medical Research Council Open Fund – Large Collaborative Grant (OF-LCG Grant No. OFLCG23may-0039, OFLCG23may-0031, OFLCG23may-0040), National Medical Research Council Open Fund – Individual Research Grant (OF-IRG Grant No. OFIRG25jul-0164), Human Health and Potential (HHP) Industry Alignment Fund Pre-Positioning Programme (IAF-PP Grant No. H25J6a0007, H25J6a0026, H25J6a0030, H25J6a0034, H25J6a0148), and Human Health and Potential (HHP) Prenatal/Early Childhood Grant (Grant No. H24P2M0005). SP is supported by grant number U2CEB032224, funded by the NIH and the National Institute of Biomedical Imaging and Bioengineering (NIBIB). NK is supported by Wellcome Discovery grant (2025, G1251481).

COMPETING INTERESTS

The authors have no competing interests to declare.

AUTHOR AFFILIATIONS

Tim Beck orcid.org/0000-0002-0292-7972
Centre for Health Informatics, School of Medicine, University of Nottingham, United Kingdom; NIHR Nottingham Biomedical Research Centre, University of Nottingham, Nottingham, United Kingdom

Jason Ferris orcid.org/0000-0001-7474-0173
Queensland Digital Health Centre, University of Queensland, Australia; School of Public Health, University of Queensland, Australia

Alain-Dominique Gorse orcid.org/0000-0003-1230-4194
Queensland Digital Health Centre, University of Queensland, Australia; Data Science Collaborative Research Platform, University of Queensland, Australia

Neerja Karnani
Institute for Human Development and Potential (IHDP), Agency for Science, Technology and Research (A*STAR), Singapore; Bioinformatics Institute (BII), Agency for Science, Technology and Research (A*STAR), Singapore

Sumir Panji orcid.org/0000-0003-0447-0598
Computational Biology Division, Department of Integrative Biomedical Sciences, Faculty of Health Sciences, University of Cape Town, South Africa

Clair Sullivan orcid.org/0000-0003-2475-9989
Queensland Digital Health Centre, University of Queensland, Australia; School of Medicine, University of Queensland, Australia

Simon Thompson
Population Data Science, Swansea University, Swansea, United Kingdom

Esmond Urwin orcid.org/0000-0003-4626-2172
Centre for Health Informatics, School of Medicine, University of Nottingham, United Kingdom; NIHR Nottingham Biomedical Research Centre, University of Nottingham, Nottingham, United Kingdom

Xing Yi Woo orcid.org/0000-0002-5980-4383
Bioinformatics Institute (BII), Agency for Science, Technology and Research (A*STAR), Singapore

Philip R. Quinlan orcid.org/0000-0002-3012-6646
Centre for Health Informatics, School of Medicine, University of Nottingham, United Kingdom; NIHR Nottingham Biomedical Research Centre, University of Nottingham, Nottingham, United Kingdom

REFERENCES

Austin, J.A., Lobo, E.H., Samadbeik, M., Engstrom, T., Philip, R., Pole, J.D. and **Sullivan, C.M.** (2024) 'Decades in the Making: The Evolution of Digital Health Research Infrastructure Through Synthetic Data, Common Data Models, and Federated Learning', *J Med Internet Res*, 26, p. e58637. Available at: https://doi.org/10.2196/58637

Cox, S., Masood, E., Panagi, V., Macdonald, C., Milligan, G., Horban, S., Santos, R., Hall, C., Lea, D., Tarr, S., Mumtaz, S., Akashili, E., Rae, A., Urwin, E., Cole, C., Sheikh, A., Jefferson, E. and **Quinlan, P.R.** (2025) 'Conversion of Sensitive Data to the Observational Medical Outcomes Partnership Common Data Model: Protocol for the Development and Use of Carrot', *JMIR Res Protoc*, 14, p. e60917. Available at: https://doi.org/10.2196/60917

DARE UK. (2024a) *DARE UK Federated Architecture Blueprint*. Available at: https://zenodo.org/doi/10.5281/zenodo.14192786 (accessed 5 November 2026).

DARE UK. (2024b) *Scientific use cases for cross-domain sensitive data research in the UK*. Available at: https://zenodo.org/doi/10.5281/zenodo.14025303 (accessed 5 November 2026).

Engstrom, T., Lobo, E.H., Watego, K., Nelson, C., Wang, J., Wong, H., Kim, S.L., Oh, S.I., Lawley, M., Gorse, A.D., Ward, J. and **Sullivan, C.** (2024) 'Indigenous data governance approaches applied in research using routinely collected health data: a scoping review', *NPJ Digit Med*, 7(1), p. 68. Available at: https://doi.org/10.1038/s41746-024-01070-3

Green, E. and **Ritchie, F.** (2023) 'The present and future of the Five Safes framework', *Journal of Privacy and Confidentiality*, 13(2). Available at: https://doi.org/10.29012/jpc.831

Hripcsak, G., Duke, J.D., Shah, N.H., Reich, C.G., Huser, V., Schuemie, M.J., Suchard, M.A., Park, R.W., Wong, I.C.K. and **Rijnbeek, P.R.** (2015) 'Observational Health Data Sciences and Informatics (OHDSI): opportunities for observational researchers', *Studies in health technology and informatics*, 216, p. 574. Available at: https://doi.org/10.3233/978-1-61499-564-7-574

Luong, J.H.C., Beggs, J., Ferris, J., Furuta, K., Gorse, A.D., Jefferson, E., Aung, K.M.M., Sullivan, C., Woo, X.Y., HEALTHWISE Consortium, Cheong, I.H., Kozlakidis, Z., Quinlan, P.R. and **Beck, T.** (2026) 'A maturity model framework for federated networks of trusted research environments', *Front Digit Health*, 8, p. 1699125. Available at: https://doi.org/10.3389/fdgth.2026.1699125

Munung, N.S., Staunton, C., Mazibuko, O., Wall, P.J. and **Wonkam, A.** (2024) 'Data protection legislation in Africa and pathways for enhancing compliance in big data health research', *Health Res Policy Syst*, 22(1), p. 145. Available at: https://doi.org/10.1186/s12961-024-01230-7

Orton, C., Thompson, S., Lee, A., Whittle, J., Clark, L., Healy, J., Jackson, M., Kavoussanakis, K., Morris, C., Parsons, M., Prodan, B., Scobbie, D. and **Vragkos, D.** (2023) *TELEPORT: Connecting researchers to big data at light speed*. Available at: https://zenodo.org/doi/10.5281/zenodo.10055358 (accessed 5 November 2026).

Ritchie, F. (2017) 'The 'Five Safes': a framework for planning, designing and evaluating data access solutions', *Data for Policy*, 10. Available at: https://doi.org/10.5281/zenodo.897821

Watego, K., Brodhurst-Hill, M., Murray, C., Mangoyana, C., Emery, A., Woon, J., Williams, K., Reid, N., Engstrom, T., Nelson, C., Sullivan, C. and **Setchell, J.** (2025) 'Indigenous people's perspectives on sharing health data for service delivery purposes: an inquiry using Indigenous methodologies', *Lancet Reg Health West Pac*, 65, p. 101753. Available at: https://doi.org/10.1016/j.lanwpc.2025.101753

Weise, M., Kovacevic, F., Popper, N. and **Rauber, A.** (2022) 'OSSDIP: open source secure data infrastructure and processes supporting data visiting', *Data Science Journal*, 21(1), p. 4. Available at: https://doi.org/10.5334/dsj-2022-004

Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L.B. and **Bourne, P.E.** (2016) 'The FAIR Guiding Principles for scientific data management and stewardship', *Scientific data*, 3(1), pp. 1–9. Available at: https://doi.org/10.1038/sdata.2016.18

TO CITE THIS ARTICLE:
Beck, T., Ferris, J., Gorse, A.-D., Karnani, N., Panji, S., Sullivan, C., Thompson, S., Urwin, E., Woo, X.Y. and Quinlan, P.R. 2026 From Local to Global: Lessons from the HEALTHWISE Consortium on Federating Secure Health Data Infrastructure. *Data Science Journal*, 25: 41, pp. 1–11. DOI: https://doi.org/10.5334/dsj-2026-041

Submitted: 30 July 2026
Accepted: 30 October 2026
Published: 19 November 2026

COPYRIGHT:
© 2026 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See https://creativecommons.org/licenses/by/4.0/.

Data Science Journal is a peer-reviewed open access journal published by Ubiquity Press.