Secure data, scalable research

Regulating health data research platforms at a national-scale

Industry Reflections on National-Level
Trusted Research Environment Accreditation Policies

lifebit
lifebit.ai

Publication date: May 2023
Authors: Ingrid Knarston, Amanda White, Hadley Sheppard, Hannah Gaimster (Lifebit)
Contributors: Mark Avery, Kieran Raine (Eastern AHSN)

Contents

Introduction

Linked health and multi-omics data generated by hospitals, healthcare providers, research organisations and clinical trials holds vast untapped potential, with the ability to advance research and innovation, improve patient care and accelerate drug development.

Beyond providing critical patient care improvements, this data holds potential to create large economic benefits. An EY report published in 2019 valued the UK's National Health Service (NHS) data at over £9 billion. The life sciences industry can play a key role in helping the UK to realise this economic potential.

Despite this promise, the process of accessing health data is slow, inefficient and fragmented. Today, the majority of the world's global data resources are siloed across distributed projects, organisations and platforms.[1,2] This means that to enable research access, data is frequently duplicated or moved, which is not only an inefficient and expensive means of sharing data but also increases data privacy risk.[3] Critically, this data is out of reach to the researchers that need access to advance health R&D, so much so that the World Economic Forum states that 97% of all hospital data goes unused.

A solution to enabling secure data access for research is facilitated by Trusted Research Environments (TREs). Also known as "Secure Data Environments" (SDEs) or "Data Safe Havens", TREs are highly secure, controlled computing environments that allow researchers to gain access to data in a safe way. These are deployed typically at the data custodian/data controllers site, either on-premise or on the cloud.

TREs support the highest level of data governance for work on sensitive personal data, removing the need to share data physically among researchers and organisations. Instead, data remains in a secure environment where it can be analysed in situ by authorised researchers, using tools available in the TRE.

Introduction

In addition to improving access, interoperability and efficiency for health data research, the use of TREs is also an important step forward in fostering trust from the broader public. With strict data governance guardrails and incorporation of patient and public engagement in TRE design, TREs can protect the best interests of the public and patients for research with health data.

This new model for data access has already proved invaluable for providing authorised researchers with rapid access to data for life-saving research - with examples including Genomics England's TRE, NHS Digital's TRE, Northern Ireland's Honest Brokerage Service, the Scottish National Safe Haven and the Welsh population biobank, SAIL Databank.

To ensure the safe use of data in TREs as they propagate across the industry, work must be done to ensure that industry standards and best practices are being met across cyber security, technical capability and information governance. Several accreditation frameworks that audit and certify TREs/SDEs to such standards have been published to support this work, including the NHS Secure Data Environment and Our Future Health TRE accreditation processes.

In this whitepaper, Lifebit provides an industry perspective on the challenges and opportunities for the use of TREs/SDEs in the UK's health data ecosystem, with a recap of updates in health data policy. The paper provides a summary of new frameworks for the safe use of data within TREs/SDEs, and highlights recent accreditation processes in health data programmes. Additionally, it dissects the implications of TRE/SDE accreditation processes for public and private sector organisations across the health data ecosystem and highlights examples of exemplary work in support of these frameworks.

The current state

There remain significant barriers to accessing data across the healthcare and life sciences industry.[1,2] These include cost, time required to gain access to data, constraints for private sector data access, interoperability of the data itself and difficulties in identifying suitable datasets to answer specific research questions. These systemic barriers can only start to be addressed with the implementation of a secure national infrastructure for data access.

Across the UK life sciences sector, TREs are proliferating as a way of achieving both data accessibility and security. In the health sector, the NHS has recommended that SDEs should be the default way to access health and social care data for R&D going forward.

In 2020, NHS England, in partnership with Health Data Research UK (HDR UK), developed a TRE supporting academic research access to cardiovascular and cancer data for COVID-19 research. The TRE provides linked, nationally collated electronic health records for around 95% of the English population.[3]

In biobanking, TREs have been in use as early as 2007, when the Secure Anonymised Information Linkage (SAIL) Databank was established to provide researchers secure remote access to anonymised health and social care data records for the population of Wales.

In the research and university sector, TREs are have been adopted by organisations including the University of Cambridge/NIHR Cambridge Biomedical Research Centre, which is implementing a secure TRE infrastructure, CYNAPSE, that will allow authorised academic researchers to efficiently use linked health and multi-omics data.

The current state

Yet to evolve a fully interoperable and sustainable network of TREs in the UK and abroad, there remains a wide number of challenges and considerations that must be addressed. We explore some of these below:

Fragmentation of Resources and Systems

The current network of data centres, platforms, resources and databases within the UK is fragmented, with many instances of duplicated data, systems and efforts. In order to derive real value from data, there is a need for more collaborative approaches to implementing standardised tools and data formats. Given that researchers require access to increasingly larger cohorts, this means also having datasets and platforms that are fully interoperable with one-another.

Initiatives such as the HDR Innovation Gateway and European initiatives such as the Beyond 1 Million Genomes project and ELIXIR represent a step towards developing a network of databases/platforms. However this must be supported with the appropriate infrastructure, systems and governance models to bring large-scale connected data to fruition.

Interoperability

Health data comes from a wide range of sources, including clinical trials, surveys and electronic health records. This variability in how data is sourced results in a wide variation in how data is described and stored, ie the data models and ontologies used. This creates challenges for researchers seeking to analyse this data, with estimates reporting that data scientists spend 80% of their time cleaning and organising data.[4] Effective harmonisation of health data into common data models (eg Fast Healthcare Interoperability Resources (FHIR), Observational Medical Outcomes Partnership (OMOP) and Study Data Tabulation Model (SDTM)) is critical for researchers to query data resources efficiently.

Interoperability also applies at the platform/system-level. In order for data sharing to extend across a UK-wide and global research network, TREs must have the ability to interact with one another via industry-recognised standards, such as authentication systems (eg Okta) and APIs (eg GA4GH Beacon).

The current state

03

Security and Privacy

The sensitivity of health data poses inherent security risk. Despite processes such as data anonymisation or pseudonymisation, there is always a risk of re-identification and misuse of data.

The scale and sensitivity of linked health data bring unique challenges for data sharing and access. TREs must be compliant with industry-wide standards that go beyond GDPR, for example, ISO 27001, the UK government-backed scheme, Cyber Essentials Plus, and NHS England's standards for clinical risk management. To achieve this level of certification, platform providers need systematic approaches to managing and protecting health data, including regular external audits of the full platform.

04

Public and Participant Trust

Patient and Public Involvement and Engagement (PPIE) is critical to build trust in data use for research.[5] Behind each datapoint is the individual represented, and it is critical that their privacy is respected and protected when using health data in research.There should be a clear discussion with members of the public to explore and gain feedback on the purpose of TREs, what data is included in a TRE, what data will be used for, who will have access to data and the potential risks involved.

To earn public trust, TREs must be designed in collaboration with members of the public, so they can have input to how, why and where their data is used. Transparency is needed around the governance and security of TREs, as is publicly-available information about the studies being undertaken, or completed, within the platform. A secondary benefit of this is it allows researchers to identify similar work in other TREs, which aids collaboration and reduces duplication.

Health policy converges on the adoption of TREs/SDEs

A significant theme across the recent UK policy landscape has been the secondary use of NHS data and, specifically, how it can be more effectively used for research.[5–9] At the heart of these discussions is the issue of balancing security and usability of data. In setting up data research infrastructure, it must successfully hold the dual role of protecting the privacy of individuals' data, while allowing for powerful insights.

Across the policy landscape there has been a clear shift to move away from a centralised model of data sharing to a federated data access model, with a network of accredited TREs (Table 1). Conventional approaches to collaborating across distributed data have relied heavily on the data sharing model, characterised by the distribution (ie copying, movement or download) of data resources across multiple users or sites - either to a centralised repository or a local computing environment. This model results in significant data ingress-egress cost and exposes data to security and privacy risks each time it is copied or moved outside of its original location. Recognising the many inefficiencies and security risks that this model poses, there is a growing trend towards a federated data access model.

A federated data access model is where data is not stored centrally, instead they remain in place and technology is used to link users with data, ie the compute and analytics are taken to the data. Federation is quickly emerging as a key technology to facilitate secure data sharing without needing to physically move data from its organisational or jurisdictional boundaries.[10]

Supporting these shifts in data access models, numerous health data policies in the past two years have converged on the recommendation that TREs/SDEs should be the default way to access NHS-related data, which is summarised in Table 1. This has been further bolstered by large scale investment. Within the NHS there have been key investments in the NHS National SDE, 11 NHS sub-national SDEs and the NHS Federated Data Platform - totalling up to £200M. In the public sector research ecosystem, Our Future Health, set to be the UK's largest biobanking initiative, has received significant investment from government, charity and pharma in establishing its TRE.

Health policy converges on the adoption of TREs/SDEs

DARE UK

Another organisation which is currently producing recommendations on TREs is DARE UK (Data and Analytics Research Environments UK).

In April 2023, DARE UK produced an initial draft architecture blueprint for a proposed federated network of TREs for public review and comment. This blueprint highlights the crucial balancing act TREs must fulfill in order to maintain the high levels of security and public trust necessary for research whilst providing a complete analytical ecosystem required by researchers to gain maximum insights from sensitive data.

Consistent with this DARE UK report and summarising recent policy guidelines, there is a shift towards establishing a federated network of inter-linked data repositories that sit within safe havens, or TREs/SDEs. Within these, authorised researchers can gain access to data on a project-by-project basis. It is further evident that there needs to be additional regulation around the secondary use of health data for research. With a proliferation of informatics platforms used for analysing health data, this important infrastructure must gain and maintain wider public support.[11] Establishing a clear and comprehensive process of TRE/SDE accreditation will help to ensure that there is appropriate oversight and assurance of secondary health data use.

Health policy converges on the adoption of TREs/SDEs

Table 1. Convergence of UK policy on the use of TREs/SDEs for health data research

Policy/ Paper	Organisation	Date	Key takeaways on TREs/SDEs
Secure Data Environments for NHS and social care data	UK Government: Department for Health and Social Care	Sept 2022	Policy guidelines for how secure data environments will be used to access NHS health and social care data. Broadly, SDEs should align with the Five Safes framework, widely regarded as representing best practice in data protection.
Life Sciences Vision	UK Government: Office for Life Sciences	July 2021	The government and the life science sector's plan to build a thriving sector. The following commitments were included: • Providing innovators with smoother, faster access to high quality 'real world', clinical and genomic data to support more effective and efficient clinical trials and allow more accurate evaluation of new innovations/technologies. • Accrediting a handful of TREs to become the default route for accessing large-scale NHS data. • TREs will be built to be highly secure and interoperable while protecting the public interest.
Building Trusted Research Environments - Principles and Best Practices; Towards TRE ecosystems	UK Health Data Research Alliance	Dec 2021	This whitepaper acts as a guide for principles, best practices and guidelines for data sharing and linkage practice within TREs. The guidelines are structured around the Five Safes framework and highlighted key points including: • Commitment to data access approach based primarily around TREs with robust and independent accreditation, monitoring and auditing. • Highlighted the potential to maximise the potential and security of TREs with federated analysis, using common agreed specifications and systems to simplify data access processes for research.
Better, Broader, Safer: Using Health Data for Research and Analysis 'The Goldacre Review'	UK Government: Department for Health and Social Care	Apr 2022	Professor Ben Goldacre's review on the efficient and safe use of health data for research and analysis. Recommendations included: • Building a small number of secure analytics platforms - shared TREs for all analysis of NHS patient records data by academics, NHS analysts, and innovators, with a culture of openness and re-use of code and platforms. • Use TREs as a strategic opportunity to drive modern, efficient, open, collaborative approaches to data science. • Make use of the enhanced privacy protections of TREs to create new, faster access rules and processes for safe users of NHS data. • Ensure all TREs publish logs of all activity, to build public trust.

Health policy converges on the adoption of TREs/SDEs

Policy/ Paper	Organisation	Date	Key takeaways on TREs/SDEs
Data saves lives: reshaping health and social care with data	UK Government: Department for Health and Social Care	June 2022	Policy paper that lays out the strategy for how data will be used to improve the health and care of the population in a secure, trusted and transparent way. Commitments included: • Mandating the use of secure data environments for NHS data, noting this would be the default route for NHS and adult social care organisations to provide access to their de-identified data for research and analysis. • Providing support for SDE mandate with: a public and policy guidelines, a robust accreditation regime, full technical specification and a comprehensive roadmap to ensure partners know how to implement the framework. • Presented 11 draft guidelines for SDEs, based on the ONS Five Safes Framework. • Introduction of new privacy enhancing technologies to keep personal data safe, including federated analytics and homomorphic encryption.
Genome UK: 2021 to 2022 implementation plan	UK Government: Dept of Health & Social Care, Dept for Business, Energy & Industrial Strategy, & Office for Life Sciences	May 2021	Policy paper defining a series of commitments for how the UK Government will work towards the vision laid out in Genome UK. • Commitment to make progress on Genomics England's new, next-generation Trusted Research Environment which provides improved, authorised access to genomic data and other linked data to researchers from across the sector.
Data: a new direction - government response to consultation	UK Government: Department for Digital, Culture, Media & Sport	June 2022	A government consultation to inform development and reform of the UK's data protection laws, aimed at creating a pro-growth and trusted data regime as part of the UK's National Data Strategy. • Highlighted key challenges in the UK including "data infrastructure that is not interoperable, legal and cultural barriers to data sharing and inconsistent data capability in the workforce." • Committed to creation of a "joined-up and interoperable data ecosystem for the public sector across the whole of the UK, that will address the limitations outlined above, whilst ensuring high levels of public trust."
A plan for digital health and social care	UK Government: Department for Health and Social Care	June 2022	This policy paper laid out a number of key commitments for enabling secure, transformative data-sharing, including commitments to: • Transition a data infrastructure underpinned by "convergence, standards and APIs, national technology platforms, and federated secure data environments." • An England-wide federated network of TREs by March 2025 that would allow researchers access to secure, high-quality, linked data sets to support research using data generated from across the NHS, including genomics, imaging and pathology.
The UK Digital Strategy	UK Government: Department for Digital, Culture, Media & Sport	July 2022	The UK government's vision for harnessing digital transformation and building a more inclusive, competitive and innovative digital economy. Commitments included: • Continued support for data-led decision-making through the Integrated Data Service, consisting of a cloud-based platform and TRE. The service will enable analysts and researchers to access, link, analyse and disseminate a range of data to help inform policy decisions. • To ensure the highest levels of privacy, patient safety and responsible use of data, the health and care system will use SDEs, including TREs, to provide researchers and analysts with secure access to data.

TRE principles and accreditation processes

The implementation of national-scale data research infrastructure needs to be supplemented with robust regulatory frameworks. A key recommendation from the Goldacre Review was the need for a single accreditation framework to identify that individuals and suppliers have the appropriate credentials and training to work safely with patient data, mirroring the accredited Researcher and Processor schemes run by the Office for National Statistics under the Digital Economy Act (2017).[8]

These frameworks would represent an oversight, credibility and assurance for TREs/SDEs, to ensure that they meet the necessary standards of cyber security, data governance, operation and technology. Accreditation frameworks would further provide data custodians/controllers with a formal process to evaluate whether a TRE supplier is equipped to securely host and fully make use of their data for life-saving research.

In the UK, the UK Health Data Research Alliance has been instrumental in defining the key overarching principles and best practices for TREs, with a 2021 white paper serving as a guide for implementation across the health sector (Table 1). Recent driver projects funded through Data and Analytics Research Environments UK (DARE UK) are further defining standardised architecture and governance models for TRE, to support delivery of a more coordinated national data research infrastructure.

Recently, two large-scale health data controllers have announced accreditation processes for TREs/SDEs, including England's public health service, **NHS England**, and **Our Future Health.**

In 2022, **NHS England** published a set of 12 guidelines outlining the expectations for how SDEs are to be used to access NHS health and social care data.[10] These guidelines build upon the Five Safes framework developed by the ONS. They further committed to establishing an accreditation process and an organisation to ensure compliance. This is set to be published alongside further technical guidance and information governance requirements in the coming months.

In 2023, **Our Future Health**, the UK's largest ever health research programme, published its TRE accreditation process for research organisations wishing to host consented, de-identified Our Future Health data (genomic and health-related data) in their own TRE. The process is based on well-established frameworks, such as the Five Safes framework, UK GDPR, and cyber security standard ISO 27001.

Needing to move at pace, Our Future Health's accreditation process was published separately to that of NHS England, however there is an intention to align the two processes in the future.

The overarching design and governance principles for secure and robust TREs/SDEs are summarised in Table 2. At a high-level, these reflect the information governance processes necessary to ensure that public and patients' best interests are maintained and the security measures needed to support the highest level of security and privacy over the data. They further indicate the required functionality to support researchers in performing collaborative, reproducible analyses over large-scale sensitive data.

TRE principles and accreditation processes

Key security, governance and technology capabilities underpinning an accredited TRE/SDE. Summarised from the NHS SDE policy guidelines and Our Future Health TRE accreditation process.

Security

- Cybersecurity standards
  - ISO 27001
  - NHS Digital cloud security good practice guide
  - NHS Digital Data Security and protection Toolkit
  - Cyber Essentials Plus
  - The Security of Network and Information Systems Regulations 2018
  - Data Centre Alliance Class 3 Facility European Code of Conduct
- Effective logging and monitoring of TRE, cloud, users, data
- Proactive patch and vulnerability management
- Active minimisation of unauthorised access
- Data encryption (in rest and in transit)
- Data minimisation techniques applied
- Data de-identification
- Data pseudonymisation
- Penetration and security testing by third party
- Firewalls between SDE/TRE and external websites/repositories
- Multi-factor authentication of users

Information Governance

- GDPR Compliance, including appointed Data Protection Officer
- Very strict controls on what data can be downloaded from the TRE: only non-disclosive results data may be exported
- Arrangements to either satisfy or set aside common law duty of confidentiality
- Document management and change management processes must be in place
- Accredited organisation is accountable for third-party suppliers
- Transparency of TRE/SDE security, design approach and data use, with publicly accessible information
- Patients and public actively
- involved in decision-making
- All uses of data must be
- for the public good
- Ability to audit the activity taking
- place within the TRE/SDE
- Processes that track individual consent
- Ability to delete data

Technology

- In the Our Future Health TRE, data and code must be ingressed through an Airlock, outputs egressed through Airlock
- Workspaces/compartments to carry out projects - keeping data, users and projects segregated
- Adopting common data model(s) to support data interoperability and other standardisation, eg. API/messaging standards
- Distinct areas for managing code versus data
- Metadata catalogue and metadata standards, including provenance
- Support for open-working and code-sharing (NHS Open Source Code Policy and Reproducible Analytical Pipelines Policy), drawing on open-source standards (Open Standards principles)
- Collaboration software, eg Git
- Support for diverse range of end-users
- Privacy Enhancing Technologies
- De-identified data linkage

Implications for the health data ecosystem

Data custodians/controllers

Accreditation frameworks will assign a greater level of assurance to TRE suppliers and data custodians. Further, the TRE accreditation frameworks proposed mean that data custodians will have a high degree of control overseeing the accreditation process and ongoing information governance of the TRE - to establish what can and can't be done with their datasets. In the example of Our Future Health, the data controller (Our Future Health) retains oversight of the full accreditation process and all ongoing interactions with the data, ie for each new data import, tool import or new project, there is an associated approval process that Our Future Health oversees.

These new accreditation frameworks mean that the TRE-data custodian relationship moves closer towards a partnership model, rather than complete relinquishing of the data to a platform provider.

TRE/SDE suppliers

To remain or become compliant with TRE/SDE best practices and accreditation frameworks (Table 2), TRE/SDE suppliers across the life sciences sector will need to remain vigilant and adaptive.

The technological capabilities for an accredited TRE will likely expand and evolve over time as new security vulnerabilities come to light or as user/researcher requirements grow. It will thus be important for TRE suppliers to be adaptive, innovative and agile in their design, development and delivery of TREs - industry SMEs are likely to be favoured in this model.[12]

The effective way to build public trust in the use of TREs is via transparency, accountability and public engagement and involvement. With a stronger focus on publicly-accessible documentation on TRE processes, design and governance, TRE suppliers will also need to display greater transparency in what is often highly guarded content. This applies to both TRE suppliers and the cloud/on-premise infrastructure providers on which they are deployed (eg Amazon Web Services, Microsoft Azure and Google Cloud Platform).

Implications for the health data ecosystem

Data consumers

For biotech and pharmaceutical companies who want to gain access to large cohorts of health-related data, the TRE/SDE model and associated accreditation frameworks will bring a notable shift away from the conventional model for data access (ie moving or copying data into a centralised or on-site computing environment). To gain access to population-level health datasets, such as from the NHS or Our Future Health, data consumers will need to adopt a TRE model for data access.

Indeed, some pharma companies have benefited from this technology already, with Boehringer Ingelheim currently working with TRE provider Lifebit to gain access to several global biobanks, via a federated TRE access model. Further, several private sector companies have invested in the Our Future Health biobank as founding industry members, which during the early years of the programme will be the only large commercial organisations that are eligible to apply to use Our Future Health resources for research. Our Future Health's industry partners will apply to their Access Board for approval to conduct research in the same way as other researchers. Note that researchers from SMEs and academia will be eligible to use the Our Future Health resource without having to be members.

Participants and the public

The TRE accreditation process will establish a robust framework for the regulation and audit of informatics platforms and data access in the life sciences industry - to ultimately support more productive and ethical use of health-related data.

Requirements for meaningful patient and public involvement and engagement in the process will ensure that views, priorities and concerns of people are taken into account and that people are involved in decisions about how data is used.

Further, the establishment and monitoring of TREs can bring vast potential benefits back to the wider public. They can both power research and innovation to support and enhance patient care, as well as drive greater economic value from this data. As an example, the NHS has the ability to leverage significant revenue from commercial vendors (eg big pharma) by making patient data securely available for authorised researchers on approved R&D projects. In a public health system that is currently facing a challenging funding climate, this represents an opportunity to enhance financial sustainability.

Adapting for the future

There has been significant innovation and work across the life sciences sector to build a more robust and transparent ecosystem of interoperable TREs. We highlight some of the advances across the sector that are supporting these changing policies and frameworks.

Eastern AHSN

Case study: Eastern Academic Health Science Network

Eastern Academic Health Science Network (AHSN) has been a key partner for several projects that are informing secure infrastructure for data research in the UK.

HDR UK Sprint Exemplar – Rare Diseases

Funded by HDR UK and led by Cambridge University Hospitals – proof of concept API-driven cloud TRE integrating electronic patient records and genomics data for participating patients with rare diseases at the NIHR BioResource.

Data access with Gut Reaction

Building on the learning from the Rare Diseases Sprint and funded by HDR UK and UK Research and Innovation, Gut Reaction, the Health Data Research Hub for Inflammatory Bowel Disease (IBD), is a secure data resource that builds on the high-quality phenotypic and genomic data from consented participants in the NIHR Bioresource longitudinal patient record data from NHS hospitals and the UK IBD registry. Eastern AHSN worked closely with the NIHR Bioresource and NHS partners to source and link the data, and with Crohn's & Colitis UK and the Gut Reaction Patient Advisory Committee to design and implement a new process for patient involvement in decision-making about data access applications.

Gut Reaction was initially funded through Health Data Research UK and now continues as part of the NIHR BioResource, supported by Crohn's & Colitis UK.

Building a next generation TRE with CYNAPSE

Eastern AHSN supported the procurement of a software solution to deliver the Cambridge Biomedical Campus an improved research data infrastructure with more effective and secure data sharing. The next generation TRE, CYNAPSE has been developed to support all features and capabilities needed for a small initial set of research groups and is now being scaled to the wider NIHR Cambridge Biomedical Research Centre (BRC).

Eastern AHSN

Case study: Eastern Academic Health Science Network

Streamlining data access with multi-party TRE federation

In 2022, a DARE UK-funded consortium led by the NIHR Cambridge BRC and including Genomics England, Eastern AHSN, Cambridge University Health Partners and Lifebit, virtually linked the TREs of NIHR Cambridge BRC and Genomics England. The federated TRE infrastructure allowed researchers to seamlessly perform parallel queries and analyses over the Genomics England and NIHR Cambridge BRC datasets in what is believed to be the first demonstration of multi-party federation of genomic data in the UK between a higher education institute and public body.

Developing information governance models for cross-agency data linkage

Combining large scale datasets is particularly important in youth mental health research where there is a complex interplay of nature and nurture, and an urgent need for earlier identification of potential mental health problems using cross-agency data from health, education and social care records. There are significant governance challenges when finding solutions to sensitive health data sharing.

To address this, a DARE UK-funded consortium, led by Dr Anna Moore from the University of Cambridge, engaged in co-creation of technical and governance solutions with participants and the public. The team recruited a panel of almost 100 young people, parents and guardians to develop public-guided recommendations for how federated analysis technologies could be deployed and demonstrated this using synthetic data. The FAIR TREATMENT (Federated analytics and AI Research across TREs for AdolescenT MENTal health) consortium included: InterMine, AIMES, Information Governance Services (IGS), University of Birmingham, University of Essex, Anna Freud National Centre for Children and Families, Cambridgeshire County Council, Bitfount and Eastern AHSN.

Eastern AHSN

Case study: Eastern Academic Health Science Network

These important driver projects have culminated to inform a national-scale programme, where the Eastern AHSN team is now coordinating one of the UK's 12 sub-national SDEs. The East of England collaboration is led by Cambridge University Hospitals with Cambridge University Health Partners and Eastern AHSN, and involves Integrated Care Systems and Boards, NHS hospitals, three university medical schools, research institutes, patient groups and industry partners. Together, they are working to establish an SDE for the region's population of approximately seven million people.

Collating insights across these projects has enabled testing and refinement of the optimal architecture and best practises for the SDE program. Bringing broad perspectives from engagement across all stakeholders in the industry, Mark Avery, Eastern AHSN Director of Health Informatics commented that ...

Mark Avery
Eastern AHSN Director of Health Informatics

"Enabling academic and commercial organisations to access health and care data is essential if they are to innovate to provide the treatments of tomorrow. Patients and the public are generally supportive where there is a clear link to patient and public interest alongside assurance that appropriate permissions and safeguards are in place. Hospitals are aware that the complexity, depth, and volume of data they hold is of considerable value to academic and industry research in the public interest. However, the complexity of governance and commercial arrangement currently presents an insurmountable barrier for most.

Access to linked deidentified data through Secure Data Environments maximises the opportunity for health and care research and innovation in the public interest whilst protecting the privacy of individuals."

Conclusion

TREs and SDEs are quickly becoming established as essential entities to enable secure and powerful health data research at scale. Supporting this infrastructure with the implementation of accreditation frameworks and bodies that regulate the access to this data will help to foster trust from the wider public and ultimately build a safer and more sustainable TRE ecosystem.

This secure, national-level infrastructure can accelerate the UK's R&D ambitions to become a life sciences superpower. More importantly, it will allow healthcare providers and researchers to fully leverage health data in a more transparent and trusted way, driving scientific breakthroughs and advances in patient care.

References

1. Denton, N. et al. Data silos are undermining drug development and failing rare disease patients. Orphanet J. Rare Dis. 16, 161 (2021).

2. Koutkias, V. From Data Silos to Standardized, Linked, and FAIR Data for Pharmacovigilance: Current Advances and Challenges with Observational Healthcare Data. Drug Saf. 42, 583–586 (2019).

3. Wood, A. *et al.* Linked electronic health records for research on a nationwide cohort of more than 54 million people in England: data resource. *BMJ* **373**, n826 (2021).

4. Cleaning Big Data: Most Time-Consuming, Least Enjoyable Data Science Task, Survey Says. Forbes (2016).

5. Horn, R. et al. Managing expectations, rights, and duties in large-scale genomics initiatives: a European comparison. Eur. J. Hum. Genet. 31, 142–147 (2023).

6. UK Health Data Research Alliance & NHSX. Building Trusted Research Environments - Principles and Best Practices; Towards TRE ecosystems. https://zenodo.org/record/5767586 (2021) doi:10.5281/ZENODO.5767586.

7. Department of Health and Social Care. Better, broader, safer: using health data for research and analysis. https://www.gov.uk/government/publications/better-broader-safer-using-health-data-for-research-and-analysis (2022).

8. Department of Health and Social Care. Data saves lives: reshaping health and social care with data. https://www.gov.uk/government/publications/data-saves-lives-reshaping-health-and-social-care-with-data (2022).

9. Department of Health & Social Care. Secure data environment for NHS health and social care data - policy guidelines. https://www.gov.uk/government/publications/secure-data-environment-policy-guidelines/secure-data-environment-for-nhs-health-and-social-care-data-policy-guidelines#:~:text=Secure%20data%20environments%20are%20data,the%20data%20leaving%20the%20environment (2022).

10. Alvarellos, M. et al. Democratizing clinical-genomic data: How federated platforms can promote benefits sharing in genomics. Front. Genet. 13, 1045450 (2023).

11. Evers, M., Pérez, L., Robke, L. & Smietana, K. Better data for better therapies: The case for building health data platforms. https://www.mckinsey.com/industries/life-sciences/our-insights/better-data-for-better-therapies-the-case-for-building-health-data-platforms (2022).

12. Troise, C., Corvello, V., Ghobadian, A. & O'Regan, N. How can SMEs successfully navigate VUCA environment: The role of agility in the digital transformation era. Technol. Forecast. Soc. Change 174, 121227 (2022).

13. van Rijsbergen, D. & Stulp, F. Effective data management: lessons to learn from Big Pharma. European Pharmaceutical Review (2022).

14. Riknbeek, P. EHDEN adds 22 Data Partners from final Call #7 in 13 countries, 5 in collaboration with HDR UK, and overall network grows to 187, mapping ~830 million records to the OMOP Common Data Model. EHDEN https://www.ehden.eu/ehden-adds-22-data-partners-from-final-call-7-in-13-countries-5-in-collaboration-with-hdr-uk-and-overall-network-grows-to-187-mapping-830-million-records-to-the-omop-common-data-model/ (2022).

15. Gliadkovskaya, A. Providers, payers need to move beyond compliance with FHIR interoperability, experts say. Here's why. Fierce Healthcare (2022).

lifebit

lifebit.ai

Get in touch:

Request a Platform Demo

Email us at hello@lifebit.ai

At Lifebit we provide data standardisation and offer secure, federated, end-to-end TRE solutions for clients including Genomics England, Cambridge Biomedical Research Centre, Danish National Genome Centre and Boehringer Ingelheim to help researchers turn data into discoveries.

Genomics England

DANISH NATIONAL GENOME CENTER

UNIVERSITY OF CAMBRIDGE

Boehringer Ingelheim