IMPERIAL

Institute of Global Health Innovation

Building the Future of UK Health Data: A Blueprint for the Health Data Research Service

Suggested citation

Ghafur S, Waldock WJ, Leis M, Acharya A, Howitt P, O'Shaughnessy J, Darzi A. Building the Future of UK Health Data: A Blueprint for the Health Data Research Service. Imperial College London (2025) doi.org/10.25560/126419

Contents

Acknowledgements

We would like to thank the following people who contributed to this document and have agreed to be acknowledged. We conducted a series of one-to-one interviews and held two roundtable discussions with individuals with a strong interest in this topic. Participants included representatives from government, the National Health Service (NHS), academia, the technology and life sciences industries, research institutions, and data privacy organisations. The contributions of the individuals listed below have been invaluable to the development of the report.

Raghib Ali, Our Future Health

Dominique Allwood, Imperial College Health Partners

Prabhu Arumugam, Amazon Web Services

Hutan Ashrafian, Flagship Pioneering

Junaid Bajwa, Flagship Pioneering

Deeph Chana, Imperial Business School

Matthew Chisambi, Imperial College Health Partners

Rebecca Cosgriff, Life Arc

Omar Din, Bourne Health

Darryn Hale, DAC Beachcroft

Nicola Hamilton, Our Future Health

Indra Joshi, Optum UK

Jonathan Loukes, Novartis

Michael Macdonell, Doccla

Lucy Mackillop, Optum UK

Adam Manhi, Flatiron

Nicole Mather, IBM

Ed Middleton, MHRA

Tamsin Morris, AstraZeneca

Ana Luisa Neves, Global Digital Health Unit, Imperial College London

Mehul Patel, Prova Health

Charlotte Refsum, Tony Blair Institute of Global Change

Philip Russmeyer, FITFILE

Samin Saeed, GSK

Nadeem Sarwar, Novo Nordisk and UK Dementia Mission

Haris Shuaib, Newton's Tree

Harpreet Sood, Skai Health and The Hurley Group

Cathie Sudlow, Usher Institute

Lydia Torne, Latham and Watkins

Janet Valentine, ABPI

Phil Waywell, Yorkshire & Humber Secure Data Environment

Jacob West, Microsoft

The report synthesises a range of insights and opinions and, as such, responsibility for the final version rests with the authors.

We would like to thank the following people and organisations for their help in producing this report:

Lydia Torne from Latham and Watkins for her expert review of the legal and value-sharing sections.

Imperial College Health Partners for sponsorship of the first roundtable.

FITFILE for their funding to support the second roundtable.

Newmarket Strategy, in particular Ele Harwich and Will Knight, for their help in running the events.

Dr Ghafur, Lord O'Shaughnessy, and Professor the Lord Darzi

The UK has established a robust health-data ecosystem that drives medical research, enhances patient care and accelerates life sciences innovation. Our globally respected health institutions have delivered breakthrough advances in genomic medicine and integrated real-world evidence into clinical trials, demonstrating clear leadership on the international stage.

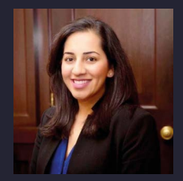
The digital healthcare landscape has changed dramatically since our last report in 2023.[1] Generative AI and Agentic systems are rapidly changing data utilisation, diagnostic capabilities and care delivery across the healthcare sector. The recent £600 million investment from the Government and The Wellcome Trust in the Health Data Research Service (HDRS) reflects this and other technological shifts, and signals a commitment to building a more efficient, data-driven and patient-centred system aligned with the NHS's *Fit for the Future: 10 Year Health Plan for England*.[2,3] However, the full potential of health data to drive better health outcomes and a stronger NHS alongside life sciences competitiveness and global innovation excellence can only be realised through comprehensive integration of primary care data. This is where most patient interactions, longitudinal health records and population-level insights reside but where the largest interoperability and data governance challenges have historically been.

Maximising the full value of NHS health data requires addressing fundamental challenges, including building on substantial public engagement to understand patient and public views. Public trust between the NHS and citizens forms the foundation of a successful health-data strategy. NHS organisations' data-sharing agreements require demonstrable citizen benefit and complete adherence to legal, regulatory, privacy and security obligations. Success must involve the creation of substantial value for patients, clinicians, taxpayers and the economy. Failure predictably erodes public support.

In that context, the HDRS should be positioned as a critical national infrastructure for UK life sciences that aligns public investment, research capability and industrial strategy to attract global partnerships and high-value research and development. By connecting NHS data assets through a secure, trusted, and interoperable framework, the UK can offer unparalleled real-world insights, supporting both domestic innovation and international competitiveness.

This report presents a strategic vision for the NHS in England to lead in life sciences and innovation through ethical, sustainable health data use aligned with core national health service values. We possess the capacity to shape a future where health data drives scientific breakthroughs, strengthens the NHS, and delivers equitable public benefit. This vision demands clearer ambition, bolder commitments and sustained focus on maintaining life sciences and innovation leadership in a significantly more competitive global landscape. The UK stands positioned to establish global benchmarks for responsible, transparent, and impactful health data use that benefit the healthcare system, patients, and communities.

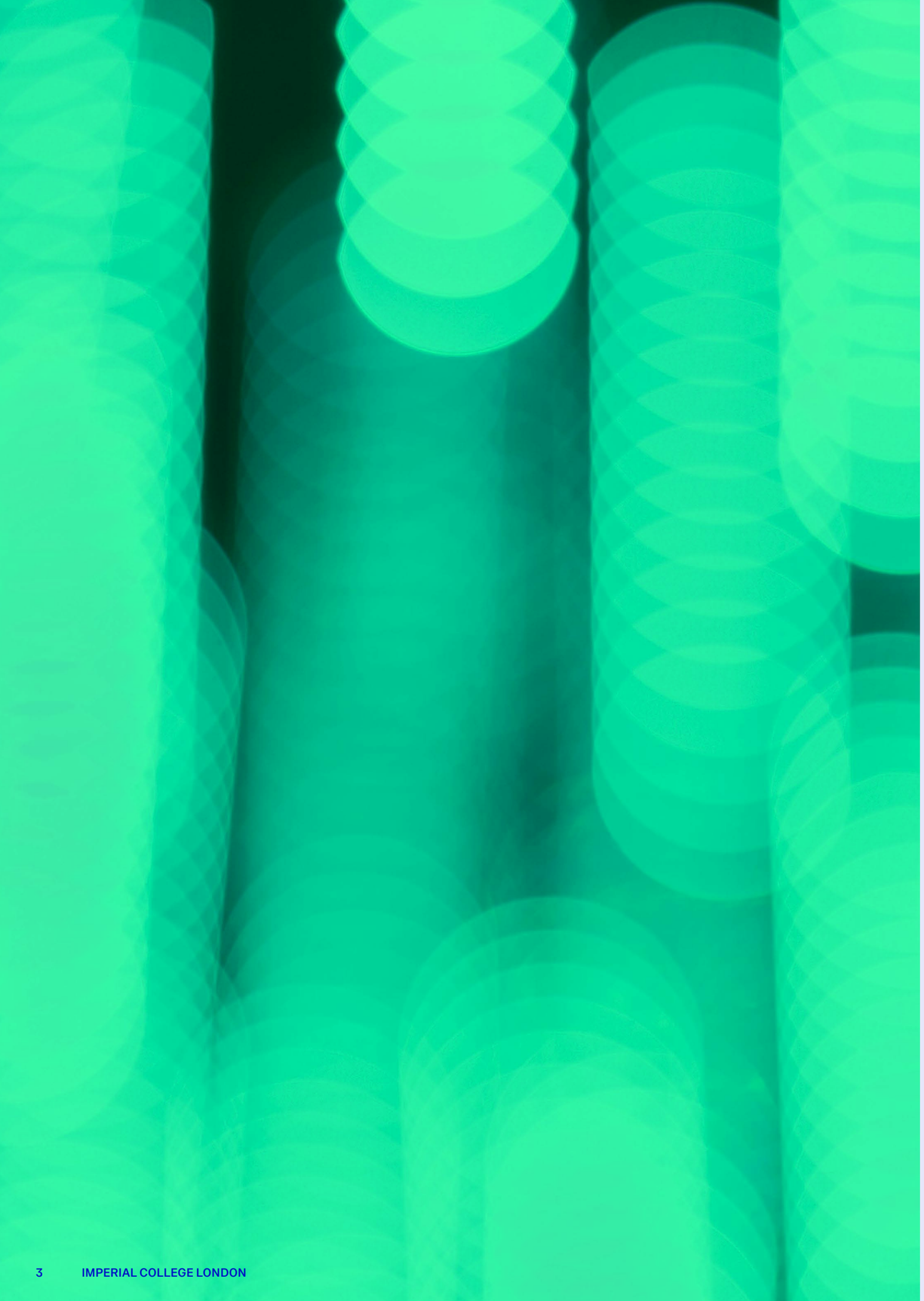
This vision depends on trust, transparency and shared benefit. Public confidence in how data is used must remain the cornerstone of progress. When done responsibly, health-data integration can deliver a triple dividend: improving care for patients, strengthening the NHS and driving sustainable economic growth across the UK.



Dr Saira Ghafur
Director of Digital Health,
Institute of Global Health Innovation
Imperial College London

Lord James O'Shaughnessy
Visiting Professor,
Institute of Global Health Innovation
Imperial College London

Lord Ara Darzi
Co-Director,
Institute of Global Health Innovation
Imperial College London



Executive Summary

The HDRS represents a £600 million investment mandate from the UK Government (£500m) and The Wellcome Trust (£100m). This is a once-in-a-generation opportunity to integrate and scale the UK's health-data infrastructure and ensure that the UK remains at the forefront of innovation in the healthcare and life sciences sphere. However, questions remain regarding optimal approaches to the implementation of this endeavour. England's health-data landscape has historically been characterised by fragmentation and missed opportunities. Previous initiatives, including Care.data, the General Practice Data for Planning and Research programme (GPDPR) and the National Programme for IT, have encountered significant challenges, contributing to public scepticism and professional caution. The rapid advancement of artificial intelligence since 2023 has underscored the potential value of health-data infrastructure, whilst simultaneously highlighting the importance of robust ethical frameworks. England's comprehensive NHS longitudinal patient records could represent a distinctive asset for research, provided appropriate coordination mechanisms can be established.
By bringing together insights from a broad range of experts, this white paper makes actionable recommendations to aid HDRS leadership in turning that mandate into a reality. This report synthesises insights from experts across healthcare, research, policy, industry and public engagement to offer recommendations for operational reality.

The HDRS seeks to develop a unified and secure infrastructure that enables responsible innovation whilst maintaining ethical standards. Three interconnected objectives are central to this vision:

1. streamlining data access and governance through clear pathways that preserve necessary safeguards;
2. facilitating cross-sector collaboration amongst researchers, industry, NHS and policymakers; and
3. supporting innovation while protecting privacy and security through appropriate technical and governance frameworks.

This report examines six themes, outlined overleaf, that are essential to HDRS's success. Integration and technical requirements to address the challenges of bringing together disparate systems and to establish interoperability standards. The integration strategy is essential for effective engagement with industry. There is a need for consistent access across distributed data sources, coupled with a shared analytics toolkit enabling standardised methods and reproducible research. The data landscape must be mapped to consider the optimal vehicles to build the core of the service; this may not necessarily be the Secure Data Environments (SDEs), given their recent introduction although their emerging infrastructure may be usefully built upon. New skills and capabilities will be essential early objectives for the new leadership, most likely achieved through seconding an interim executive team by Q2 2026.

Several dependencies warrant attention. Regulatory and legal reform may be needed to clarify data-controller responsibilities and streamline governance processes. Workforce skills must be developed, including essential skills in service design, commercial product design and delivery, data engineering and information governance. Policy consensus-building across government departments, NHS organisations and professional bodies is required to prevent fragmentation. The particularly complex challenge of primary care data access requires very careful consideration given its essential role in enabling better population health. Legislative pathways, compensation models and risk-allocation frameworks remain areas where further specification is required. A single monolithic data stack should not be pursued but a virtualised, federated architecture may be advisable, delivering a 'network of nodes' that build upon existing infrastructure.

England possesses substantial assets to enable the success of the HDRS: comprehensive health data, world-leading research institutions, a thriving life sciences industry, and a public that has historically supported the NHS and health research. What has been lacking is coordinated and enduring support and finance for the infrastructure to connect these elements. The HDRS represents an opportunity to address this gap, though success will rely on careful consideration of strategic decisions regarding architecture, governance and stakeholder engagement.

We have outlined the recommendations from the six main themes of report overleaf:

Theme:

Recommendation:

1 Public Trust and Engagement

1.1 Systematic Public Engagement
Embed systematic public engagement through ring-fenced patient and public involvement and engagement (PPIE) budgets with expenditure deadlines, federated engagement models connecting local to national teams, and democratic co-design processes that move beyond consultation.

1.2 Governance and Accountability
Establish clear data-controller roles and accountability mechanisms with patient and advocacy group participation in governance structures, ensuring transparent communication of data-subject rights and compliance.

---

2 Service Design and User Needs

2.1 Define User Requirements
Define and prioritise use cases across researchers, industry partners, policymakers, clinicians and public health teams. Specify service models, dataset and depth requirements, access modalities, analytical capabilities, compute requirements, governance and training needs, payment models and willingness to pay before developing service architecture.

2.2 Clarify Technical Model
Clarify if the HDRS will consolidate health datasets into a unified platform or operate as a marketplace of users; this will determine user experience, governance models, pricing structures, and technical infrastructure.

2.3 Ensure Sustainable Infrastructure
Establish robust technical capabilities and infrastructure with long-term staffing plans and sustainable funding models, drawing on lessons from initiatives such as *Our Future Health.*

---

3 Legal and Risk Frameworks

3.1 Clarify Risk Allocation
Establish clear risk-allocation frameworks and related legal documentation, ensuring data controllers, processors, and users understand respective responsibilities, liabilities, and accountability structures throughout the data lifecycle. This can be supported by legal advice that supports researchers to establish appropriate permissions for their work.

3.2 The Data (Use and Access) Act [hereafter The Data Act]
The HDRS governance structures should be designed to assess whether proposed research meets the 'reasonably described as scientific and conducted in the public interest' standard, providing a clear decision-making framework aligned with the legal basis.

---

4 Integration and Technical Requirements

4.1 Balance Anonymisation and Analytical Value
Balance anonymisation benefits (simpler legal pathway, reduced regulatory burden) against analytical limitations for precision medicine and longitudinal research through a tiered-access model matching data sensitivity to the extent that granularity is required.

4.2 Phased Federated Architecture
A single monolithic data stack should not be pursued. Adopt a phased, federated architecture building upon proven initiatives through shared standards and application programming interfaces (APIs), prioritising interoperability over consolidation whilst establishing transparent development timelines.

4.3 Governance Beyond Individual Leadership
Develop governance structures that transcend individual leadership tenures, balancing political opportunities with institutional resilience whilst maintaining public-sector interests in technology partnerships.

4.4 Set up the HDRS as a Government-owned Company (GovCo)
This scenario offers the right balance between autonomy and accountability and will mitigate the risk of a protracted set-up time.

---

5 Primary Care Data

5.1 Legal Liability Reform
Consider the appropriate legal frameworks for transferring data from GP practices to a national body for secondary use (following models such as Finland's Findata). This will limit GP practices' liability for the acts and omissions of HDRS's use of the data, whilst preserving practice accountability for data quality and patient opt-outs.

5.2 Technical Infrastructure and Fair Compensation
Explore implementing automated data extraction using GP system APIs with minimal practice burden, coupled with appropriate financial compensation and performance tools supporting patient care. Specific GP burden concerns and fair reimbursement rates require further specification.

5.3 Build Healthcare Professional Confidence
Address GP liability concerns through the HDRS or the Secretary of State for Health and Social Care via the Department of Health and Social Care (DHSC) potentially assuming data-controller status. Developing case studies that demonstrate how linked data improves clinical outcomes may help secure professional trust.

---

6 Value Generation

6.1 Value Prioritisation
Prioritise patient benefits through improved clinical outcomes whilst establishing tiered licensing fees for commercial users to generate sustainable NHS revenue streams. Establish a range of data services and environments, including enabling onshore investment and business growth via data-product development.

6.2 Measurement and Accountability
Establish baseline measurements at inception with intermediate milestones aligned to implementation phases, tracking progress across economic impact, research excellence, system performance, and health outcomes.

6.3 Industry Collaboration Frameworks
Create frameworks for industry collaboration that balance mutual benefit with public-interest protection.

Introduction

The UK stands at a pivotal moment to realise the transformative potential of healthcare data. We benefit from the NHS – one of the most trusted institutions in the world – a proven track record in health and life sciences innovation, and a dynamic technology sector anchored by world-class universities. The Covid-19 pandemic offered a powerful proof of concept when emergency legislation enabled comprehensive data-sharing. The results were remarkable for informing public-health interventions and accelerating life-saving vaccine development. Crucially, this success required collaboration across the entire economy with the NHS, academia, industry, and government working in close collaboration.

The reality is that we have not fully taken advantage of this progress. The English health-data landscape is characterised by fragmentation, overlapping systems, and under-utilised assets that collectively constrain the potential of the NHS's substantial data resources. The NHS confronts persistent challenges in data integration, governance, and public trust that impede effective utilisation of this strategic national asset. Since our last white paper on this topic[1] in 2023, the rapid rise of artificial intelligence (AI) has disrupted industries worldwide, and healthcare is no exception. AI-driven advancements are transforming the way we utilise data, diagnose conditions, and deliver more efficient care. In a bid to accelerate progress[2] to build a more efficient, data-driven system, the UK Government has recently invested £500 million, alongside £100 million from the Wellcome Trust, to establish the HDRS.[3] Its launch in 2026 will provide a single point of access to data for clinical research in a more efficient manner. This significant investment aims to strengthen research and development in life sciences, drive vital national investment, and accelerate clinical-trial development.

First and foremost, the HDRS must deliver value to patients and the public. The service has the potential to deliver the following core benefits to individual patients:

- enhanced clinical safety and personalised care through integrated health data
- accelerated access to healthcare innovation by facilitating research and development
- strengthened data governance through robust oversight mechanisms, transparent data management, and user-controlled settings that empower patients to exercise control over use of their data, while building public confidence in health data utilisation.

Recent Policy Initiatives: Summary of Policy Landscape

In the past 18 months there have been significant policy developments – including *Fit for the Future: 10 Year Health Plan for England*[4] – that are reshaping the health-data conversation. The UK Government's 2025 *Life Sciences Sector Plan*[5] subsequently positioned NHS health data as a strategic national asset aimed at driving economic growth and scientific advancement. Central to this strategy is HDRS. It is intended to be the backbone of a secure, AI-ready infrastructure that links hospital, genomic, and primary care data. The HDRS aims to establish the UK as a global destination for clinical trials and AI development whilst ensuring the NHS receives a fair share of resulting benefits. *Fit for the Future* requires that – by 2027 – 95% of people with complex needs have an agreed personal-care plan, with neighbourhood health teams using these plans to proactively manage multiple conditions and deliver personalised, coordinated care across community-based services.[4,6]

The Government's *AI Opportunities Action Plan*[7] provided a comprehensive list of recommendations to grow the UK's AI sector and drive adoption across the economy to boost growth and improve public services. *The Sudlow Review*, published in 2024, recommended the set-up of the HDRS.[8] The review highlighted that national health data constitutes critical national infrastructure, including the Data for Research and Development [R&D] Programme,[9] that can underpin the health of the nation, with principles directly aligning with NHS efforts to maximise health-data value whilst maintaining ethical governance and public trust. The pending *Data Act* has updated data protection laws to promote innovation and growth alongside protecting public rights. It has been forecast to deliver an estimated £10 billion economic boost over 10 years.[10]

As the NHS moves from an analogue to a digital service,[11] it needs to build on reliable, robust, clean, and curated data repository. The NHS can leverage its reputation as a global leader in responsible health-data management, driving innovations benefiting both the UK and the international healthcare community. This opportunity has gained significant momentum and coincides with critical updates to the NHS's data infrastructure, notably the introduction of the Secure Data Environment (SDE) Network, the Federated Data Platform (FDP) and continued evolution of the NHS app. The NHS app provides an accessible portal for patients to engage with the heath-data architecture, and the plan is for the app to be the 'digital front door' to all NHS services. The FDP has potential for applications including improving patient flow, supporting elective care recovery, and optimising resource allocation. The SDE network represents a decentralised yet standardised approach to data access and research. Rather than relying on traditional data-sharing methods involving data duplication and transfer, SDEs function as controlled-access environments where authorised users can analyse data without removing it from secure locations.[12]

Previous Failed Data Initiatives

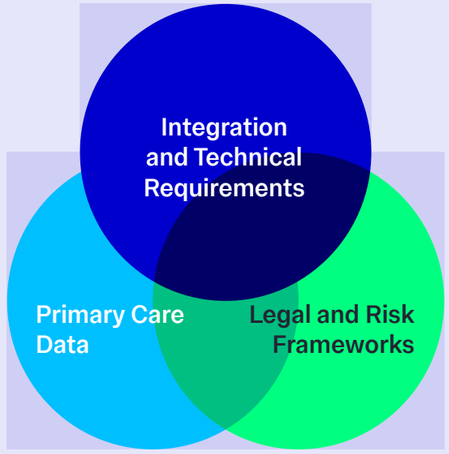
NHS England has the historical legacy of several unsuccessful data endeavours. Since the collapse of the General Practice Data for Planning and Research (GPDPR) programme[13,14] in 2021, a notable gap has emerged in primary care data initiatives. GPDPR's failure, largely attributed to insufficient public engagement, privacy concerns, and a lack of transparent opt-out mechanisms, has left policymakers cautious about launching similar large-scale initiatives. However, the absence of a replacement programme significantly undermines the NHS's ability to fully utilise primary care data, which constitutes one of the most valuable resources for preventative care, population health, chronic disease management, and early diagnosis. Presently, researchers can access Clinical Practice Research Datalink (CPRD) data and request linkage to Hospital Episodes Statistics (HES) data, though this process is often time-consuming, costly and delivering insufficient data depth. Without comprehensive, standardised access to primary care data, opportunities for data-driven research, population health strategies, predictive analytics, and AI-driven healthcare improvements remain under-served. A renewed focus on the utilisation of primary care data must prioritise robust public and professional engagement, transparent governance structures, and granular opt-out frameworks building on the National Data Opt-Out that already exists.

Objectives of the HDRS

The HDRS presents an exciting opportunity for England to enable a cohesive, ethical, and scalable health data research service to realise the potential of NHS data assets whilst maintaining public confidence and professional engagement necessary for long-term success. Alongside these objectives, the HDRS must represent value for money to taxpayers.

The HDRS must address these challenges through three primary objectives:

1. Streamline access and governance of health data by establishing clear pathways that balance efficiency with appropriate safeguards.
2. Facilitate cross-sector collaboration by creating mechanisms enabling researchers, policymakers, industry partners, clinicians, and public health teams to work effectively with health data.
3. Support innovation whilst protecting privacy through technical and governance frameworks that enable valuable research without compromising individual rights or public trust.



Three interdependent themes essential to the success of the HDRS.

This document outlines key considerations leading to an implementation roadmap for the HDRS to successfully meet these objectives. The following sections examine three interdependent themes that may be essential to HDRS success. Integration and technical requirements address challenges of bringing together disparate systems and establishing interoperability standards. The particularly complex challenge of primary care data access may require innovative solutions that respect GP autonomy whilst enabling population health research. Legislative pathways, compensation models, and risk-allocation frameworks remain areas where further specification would be valuable.

1. Service Design and Public Trust

Patient and Public Involvement and Engagement
Meaningful Patient and Public Involvement and Engagement (PPIE) must be embedded throughout HDRS development and operation, informing governance structures, use case prioritisation, and communication strategies. Patients and patient advocacy groups should participate directly in governance discussions rather than being consulted peripherally. Research charity partnerships may include Cancer Research UK, the British Heart Foundation and the Association of Medical Research Charities. Data subjects, healthcare professionals, researchers, and commercial partners must understand who acts as data controller in various scenarios, what rights individuals retain over their data, and how accountability mechanisms function. Ensuring awareness of, and compliance with, data-subject rights, including rights to access, rectification, erasure, and objection under data-protection legislation, proves critical.

Communicating HDRS's value proposition represents a significant challenge across multiple stakeholder groups. For the public, the benefits of linked health data are most readily apparent in direct clinical care contexts, such as clinicians accessing complete patient histories to inform treatment decisions. The value for research and population health improvement, whilst substantial, proves more difficult to communicate effectively. Developing and widely disseminating case studies is essential to showcase successful applications of linked data used at scale in research, thus demonstrating tangible benefits to patient outcomes, service delivery, and medical advancement.

Healthcare professionals, particularly GPs who serve as gatekeepers for primary care data, must play a central role in the set-up of the HDRS. Without primary care buy-in and advocacy, public trust will be difficult to achieve. Under current data protection law, individual GPs serve as data controllers, carrying personal and professional liability for any data breaches or misuse of their data. The perception that, if they share data with a third party, this liability will extend to the acts or omissions of the third party is a powerful disincentive to participate in data sharing initiatives, regardless of their potential research value or system benefits. This structural barrier has limited the use of primary care data for decades despite its transformative potential. The new HDRS must address this core issue, identify legal pathways to mitigate this risk (e.g. parties acting as data controllers in common with clear data sharing agreements, potentially with further contractual indemnification protection against breaches by the new data controller), and seek dialogue to build professional trust for GPs to be part of the solution.

Building and Maintaining Public Trust
The foundation of any successful health-data strategy is protecting the trust between the NHS and the public; it must never be taken for granted. Data-sharing arrangements must demonstrably serve citizens' interests while rigorously meeting all legal, regulatory, privacy and security obligations. Done poorly, we face predictable withdrawal of public support that will take years to rebuild. Trust must be actively built through transparency, demonstrated value, and engagement with public concerns regarding privacy, commercialisation, and secondary uses of health information. NHS England's 2024 *Public Attitudes to NHS Data*[15] reinforces these imperatives. After hearing from 2,200 individuals, the NHS remained the most trusted entity for data handling, with 72 per cent expressing trust, though cybersecurity threats constituted the top public concern.

Transparency
Engagement initiatives include deliberative forums hosting citizen panels to understand public expectations and concerns. A notably effective example has been the Powerful Moments, Powered by NHS Data initiative,[16] designed to highlight the vital role that securely shared and ethically used NHS data plays in transforming healthcare. By presenting compelling real-life examples, the initiative demonstrates how access to quality health data has led to breakthroughs in research, improved treatments, and ultimately better patient outcomes. During the Covid-19 pandemic, the importance of NHS data was highlighted through initiatives such as the RECOVERY trial,[17] with the public able to appreciate direct benefits from using NHS data. These findings had immediate and global impact, changing clinical practice and saving lives in real time. The Cohort 3 National Data Engagement strategy [18] recommends maintaining an opt-out system, while simplifying clearer, tailored choices and regular reminders of status. The main proposal is to remove the ability to opt out of data being used for NHS planning purposes, though safeguards against misuse and the ability to opt out of research uses would be maintained. [18,19]

Engagement
Engagement strategies must avoid tokenistic approaches and performative consultation exercises. Public deliberation remains sporadic and needs to be embedded systematically into policy-making processes. Practical mechanisms to ensure substantive engagement include ring-fencing PPIE budgets with mandatory expenditure deadlines. This requires sustained commitment from decision-makers to adopt a more democratic design process. Whilst this may lengthen delivery timelines, despite established structures, it ultimately ensures greater legitimacy and effectiveness in health data governance. The federated engagement model connecting local, regional, and national teams has been well received by the public, though room for improvement exists, particularly in innovating public engagement mechanisms. One London[20] remains a best-practice model, utilising deliberative methodologies to shape policy, whilst Citizen Jury processes in Yorkshire and the Humber,[21] Thames Valley,[22] and Surrey have effectively engaged marginalised communities to enrich insights.

The foundation of any successful health-data strategy is protecting the trust between the NHS and the public; it must never be taken for granted.

An illustrative example of large-scale public partnership and engagement is the *Our Future Health* [OFH][23] initiative. OFH aims to recruit up to five million volunteers across the UK to contribute health and genetic data for research purposes. As of July 2025, over two and a half million[24] individuals have enrolled, reflecting significant public engagement and a degree of trust in the initiative's governance framework. This milestone demonstrates that large-scale data initiatives can achieve substantial participation when they are supported by transparent communication and robust data protections and are aligned with public interest. There is a unique opportunity for the public to contribute to potentially ground-breaking medical research by providing their health data, while feeling empowered to benefit from the discoveries made through the programme. The initiative emphasises inclusivity by aiming to gather diverse data from across the UK population. OFH commercial partnerships have emerged as a potential blueprint for the industry by collaborating with a diverse group of industry partners to fund, design, and implement this ambitious research programme and by bringing together expertise from leading organisations in the life sciences sector. However, it is important to note that these partnerships are predicated on a period of exclusivity, designed to recognise their early investment in this platform. It is unlikely that HDRS would be able to offer this exclusivity as it would go against NHS data-sharing principles.[25]

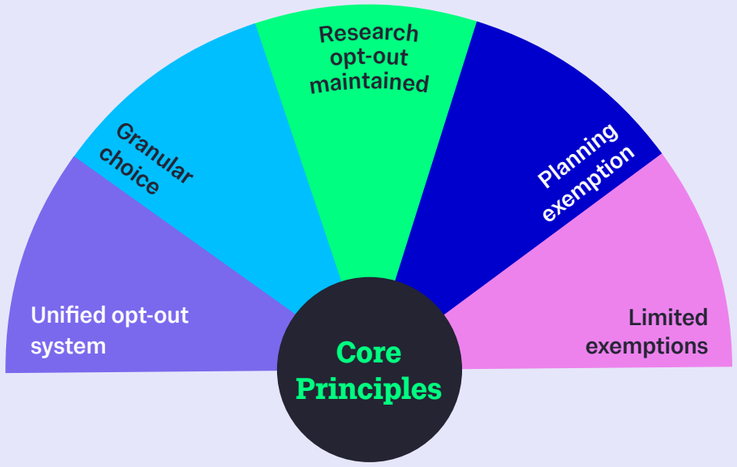
We propose that the HDRS includes a transparent, accessible, opt-out system that empowers citizens whilst enabling essential health-system planning and neighbourhood health delivery:

Core Principles

- **Unified opt-out system:** single, simplified process replacing fragmented existing mechanisms
- **Granular choice:** citizens can tailor opt-out preferences or make a single comprehensive choice
- **Research opt-out maintained:** full ability to opt out of data use for research purposes
- **Planning exemption:** data use for NHS planning and neighbourhood health service design cannot be opted out of, with strict safeguards against misuse
- **Limited exemptions:** clearly explained justifications with ongoing public consultation

Transparency and Awareness

- **Proactive communication:** Regular reminders of opt-out status and ongoing public awareness
- **Educational initiatives:** Clear explanation of how secondary data use supports neighbourhood health services, early intervention, and care closer to home
- **Safeguard visibility:** Enhanced awareness of robust protections, cybersecurity measures, and governance frameworks. Public-facing materials explicitly detail what is and is not covered by opt-out provisions.



2. Defining the Service Offering

The service must be defined and set up to appeal to different user groups and use cases (for example, observational research or clinical trials) who will ultimately determine the success of the HDRS. Additionally, use cases will help shape the technical requirements. If the service does not meet the needs of the end-users, it will likely be underutilised and will be unable to attract further funding and the partnerships needed for sustained growth and innovation. These needs will be determined by both the user groups and the use cases. **Table 2** outlines the main intended user groups, and their likely requirements. **Table 3** outlines the potential use cases.

USER GROUP	SCALE	DATA REQUIREMENTS	ANALYTICAL NEEDS	PAYMENT CAPACITY	VALUE PROPOSITION
Academic researchers	Individual principal investigators to large research consortia	Modest datasets for focused studies to comprehensive longitudinal data	Advanced analytical environments, significant compute resources, statistical modelling	Limited – constrained by grant funding	Research outputs, scientific discovery, publication potential
Industry partners – large pharma	Large pharmaceutical companies	Targeted, specialised, fit-for-purpose datasets for focused studies, extensive multi-year, population-level data-sets and post-market safety monitoring	Sophisticated statistical environments and analytical infrastructure	High – willing to pay substantial/ premium fees for comprehensive access	Drug discovery, development, safety monitoring
Industry partners – small biotech	Small biotech startups, digital health startups	Targeted, specialised, fit-for-purpose datasets for focused studies	Flexible analytical requirements	Low to moderate – constrained budgets requiring flexible pricing structures	Focused research questions, drug discovery, targeted product development
Policy makers and NHS analysts	Local authority, regional, and national decision-makers	Aggregated, up-to-date data, population-level indicators	Dashboards, visualisations, indicators rather than complex statistical modelling	Minimal, direct payment capacity	High strategic value, system efficiency, resource allocation, improved population health outcomes
Clinicians and public health teams	Service providers, quality improvement teams, outbreak monitoring teams	Granular, near-real-time data	User-friendly interfaces with minimal technical barriers, rapid turnaround, actionable outputs	Limited	Service evaluation, quality improvement, outbreak detection, direct clinical practice and public health intervention support

Table 2: HDRS Intended Service Users

Table 3: HDRS Expected Use Cases

USE CASE	EXAMPLES
Interventional studies	Clinical trials for drugs, devices or behavioural interventions
Observational studies	Epidemiological studies, patient safety and quality of care analyses
Genomic analyses	Precision medicine, including drug design

These use cases and user groups will have different service requirements, ranging from standard data reports to self-service analysis and more advanced data modelling. The service must be designed to support all these use cases and must be able to clearly articulate how. All of this must be mapped to specific use-case archetypes before meaningful service architecture can be developed. Beyond the user definition, the HDRS requires robust technical infrastructure and adequately trained staff capable of building, maintaining, and supporting the system over extended timeframes. There is a risk of developing an infrastructure that either fails to meet actual user needs or cannot be sustained operationally. Whilst successful models such as OFH offer potential templates, uncertainty persists regarding which elements merit adoption. As it stands in England, each existing data service has negotiated distinct trade-offs between usability, data comprehensiveness, and access controls that reflect their specific user base and governance arrangements.

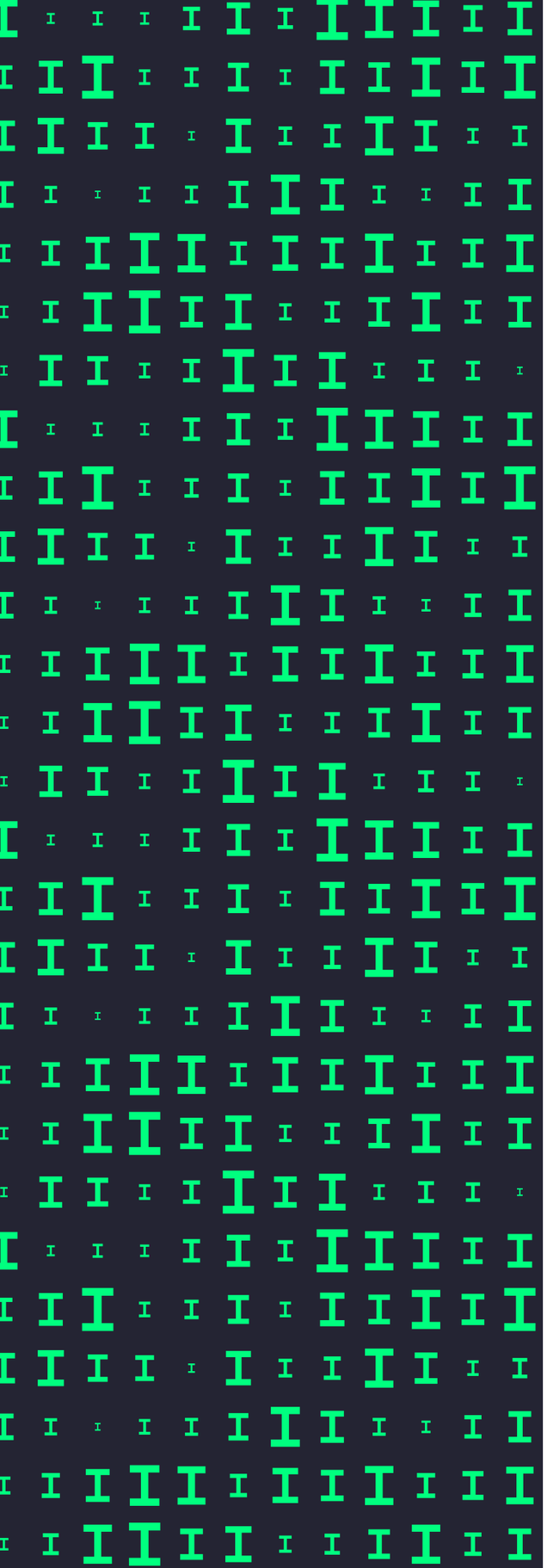
Whether HDRS forms a 'shop window' into existing services or a more unified data platform, it will also need to be considered within the context of timescales of deployment, the costing model and its interaction with existing data repositories. The core challenge for the HDRS is resolving the fundamental service architecture: whether to be a unified, consolidated platform or a federated marketplace that connects its users to distributed services. This decision is critical as it determines the governance models, pricing structures, user experience (UX), and technical infrastructure, reflecting the existing complex trade-offs between usability and control found in other data services.

Designing the technical infrastructure

The technical infrastructure must be designed for immediate and rapid adoption of AI capabilities. The system must be built for scalability and performance to handle computationally intensive AI workloads, adopting cloud-native architectures and providing access to necessary computational resources like graphics processing units (GPUs). Standardised, version-controlled data pipelines are essential for effective model training. This environment must be safeguarded by stringent access controls, and robust auditing and provenance capabilities. Rigorous, mandatory output checking procedures must be implemented for all research findings before they are released, ensuring adherence to security and ethical protocols. Before developing meaningful service architecture, these workforce and infrastructure requirements must be strategically mapped against the specific user archetypes (**Table 2**) and the service's chosen consolidation model.

Health Data Research UK launched their Gateway in 2020 with the aim to provide "a portal for researchers to search, discover and request access to health datasets in the UK". [26] It includes metadata from 800+ datasets and is being developed to become the 'single front door' for the SDE Network. The HDRS should be more ambitious in its aim. While a 'service wrapper' would largely be the most feasible to deploy within the 12-month timeframe proposed for version one of the service to go live by the end of 2026, care should be taken to not duplicate what is already there. To add value to potential users, the HDRS should have a strategic roadmap with clear milestones to evolve from a marketplace to a platform that offers users access to unified datasets. This process of data integration should be demand-driven, balancing user needs with technical feasibility.

The main advantage of a single-service layer is the provision of a single point of entry for industry partnerships. Finally, as described, the UK boasts some of the most granular cohort and data repositories globally, including UK Biobank, Genomics England and Our Future Health. A single coherent service is needed across consented cohorts. How HDRS complements and enhances the useability of the data derived from these programmes will be directly related to the nature of the service being delivered. For instance, it is unlikely that a marketplace format could enable interrogation of these datasets in their current context, which would mean users would still need to rely upon bespoke requests and linkage procedures.



3. Legal and Ethical Considerations

Data Protection Laws
Legal challenges will need to be navigated to facilitate HDRS's development. In particular, the legal basis of the data-sharing structures between current data holders, for example, GP practices and HDRS will need to be clarified alongside related contractual arrangements and risk allocation (see section 5).

The law is already beginning to evolve to facilitate more easily the sharing of data for secondary research purposes. *The Data Act* [10] materially changes how data is used and accessed across various sectors. With respect to healthcare data, it seeks to establish a framework for a sector-specific smart data scheme via secondary legislation. The new amendments of UK GDPR[27] came into force in June 2025, clarifying that where data is collected for one purpose and intended to be used for a secondary purpose, such purpose shall be considered compatible (and therefore permitted) if (among other things) it is for scientific research purposes. The Act also clarifies the rules around 'broad consent' for scientific research purposes and provides an exception to transparency requirements where the controller intends to process the data for scientific research purposes and providing the required information to data subjects would involve a disproportionate effort. 'Scientific research' is defined to include any research reasonably described as scientific and conducted in the public interest, regardless of the funding source or commercial status. This change enables large-scale health datasets collected for healthcare purposes by care providers to be shared with the HDRS for scientific research, statistical and other purposes.

The government should assist current data controllers with these assessments by providing relevant key information. Likewise, guidance from the ICO on appropriate data sharing arrangements with DHSC via the Secretary of State for Health and Social Care would help to support current data controllers. Ideally, to provide ultimate clarity, the government would introduce a statutory amendment to UK General Data Protection Regulation (GDPR) expressly providing for the sharing of health data with HDRS and/or setting out a statutory defence to liability for data controllers who comply with government requirements for the sharing of data with HDRS.

The HDRS governance structures should be designed to assess whether proposed research meets the 'reasonably described as scientific and conducted in the public interest' standard, providing a clear decision-making framework aligned with the legal basis. The HDRS should also include a mechanism for transparency information to be made publicly available to rely on the "disproportionate effort" exemption. If data can be fully anonymised, the legal pathway becomes considerably simpler and reduces regulatory burden. However, complete anonymisation may limit the value of the insights derivable from data, potentially undermining the service's utility for precision medicine, longitudinal research, and other high-value applications.

Patient Confidentiality Laws
In addition to ensuring compliance with data protection laws, the proposed sharing of patient medical data will need to comply with patient confidentiality laws. The common-law duty of medical confidentiality requires that identifiable patient information held by NHS organisations can only be shared outside the care team under the following scenarios:

- with the patient's express consent
- with the Confidentiality Advisory Group's support (which allows the common-law duty of confidentiality to be lifted for a period of time so that confidential patient information can be used without breaching the duty of confidentiality pursuant to section 251 of the NHS Act 2006)
- if there is a legal requirement to do so (e.g. a court order)
- if there is overriding public interest to do so.

Because HDRS activities such as research, population-health analytics and service planning are not direct care, implied consent cannot be relied upon.

The HDRS must either use de-identified data or ensure that identifiable data is shared only under a lawful gateway described above, such as under a Section 251 Control of Patient Information (COPI) approval (as was done during the Covid-19 pandemic)[1].

To facilitate this (among other things), the government could introduce a new long-standing COPI approval for HDRS purposes and/or clarify that pseudonymised medical records may be shared in the same way as de-identified or anonymised data, provided a reasonable threshold of pseudonymisation is undertaken. Strong governance, transparency, and effective patient opt-out mechanisms are essential to maintain public trust and the 'social licence' for HDRS to operate responsibly.

Ethical Foundations
One of the great value propositions of a consolidated data resource is the capacity to use this in the development of data-driven technologies, including in the training of AI. Whilst the ethics and regulation of AI in health is beyond the scope of this paper, certain areas must be considered in relation to the HDRS. A consolidated data resource may give the impression that algorithmic biases are less likely due to the numbers of data points represented. However, given that data-collection activities in health are largely influenced by human activity, the development of AI systems that reinforce existing health inequalities are likely. With HDRS there may also be an assumption that integration necessarily equates to complete representation. Whilst databases such as Clinical Practice Research Datalink (CPRD) encompass 18 million patients[28] currently registered with a GP, they are only broadly reflective of the population. Understanding the limitations of the representativeness of the data will help avoid selection biases, especially as the data from groups including those who have chosen to opt-out will be missing.

Many of these issues can be mitigated by adopting a 'responsible-by-design' approach that embeds the principles on the safe development and deployment of health AI.[29, 30] In addition to upholding core values regarding transparency, explainability and equity, these frameworks also call for a model of collective responsibility for everyone involved in AI development. The governance of HDRS should include a clear code of conduct for AI development and use this as an opportunity to reinforce the principles of responsible AI to all users, both commercial and research-based.

Whilst the prospect for HDRS to support commercial activity is welcomed, this must be transparent and with appropriate oversight. The potential for a consolidated health-data resource to be the foundation of hostile algorithm development, or to be used in biodefence is entirely possible. In addition, there are scenarios in which HDRS makes it easier for algorithms to be developed/trained using national health data, only for them to be subsequently deployed in other more lucrative territories, and where NHS patients fail to receive the benefits of their data use. Governance and communication need to be proactively considered, with measures put in place to mitigate these risks, but not to hinder innovation.

4. Integration and Technical Requirements

Mapping the Data Landscape
England's health-data landscape comprises numerous overlapping initiatives with varying effectiveness. The HDRS will need to align with these existing structures to prevent over-duplicating services. Multiple data structures can be considered, given the existing ecosystem of assets such as the SAIL Databank and NHS DigiTrials. The SDE network includes eleven subnational/regional and one national SDE enabling NHS organisations, researchers, and analysts to access health data without transferring sensitive patient information.[31] The nationwide rollout of SDEs has occurred simultaneously across multiple NHS organisations without sufficient time for phased learning and adaptation. Implementation has been inconsistent; whilst some SDEs function effectively with potential for national scaling, others struggle to deliver value. The eleven regional SDEs vary considerably in operational maturity, technical capabilities, governance frameworks and indeed success. This has led to duplicated effort, incompatible standards, and uneven regional access. It complicates multi-regional studies, as researchers must navigate different approval processes, technical specifications, and data formats. Establishing and maintaining a SDE requires sustained investment in infrastructure, multi-year funding, personnel, and training.

Other major initiatives operate in parallel with limited coordination. Consented cohorts, national data assets and public charity data remain uncoordinated and siloed. Primary care data remains under-utilised for research despite earlier digitisation, as noted in the *Independent Investigation of the NHS in England* by Lord Darzi in 2024.[32] Ongoing challenges exist for the Federated Data Platform (FDP),[33] which seeks to integrate data across NHS trusts and Integrated Care Boards whilst maintaining local governance. While the FDP is designed to support real-time data sharing and integration across the NHS, it remains under-utilised. A lack of tangible real-world examples demonstrating the benefits of FDP has made it difficult to secure wider engagement across the system. Without a unified strategic vision, there is a significant risk that the FDP becomes a technically sophisticated but under-used tool, rather than a catalyst for meaningful healthcare improvement. Genomics England has created one of the world's largest national genomic datasets, linking genomic information with longitudinal health records. UK Biobank provides access to data from half a million participants. Each of these initiatives operates under different governance structures, consent frameworks, and data access mechanisms. Critical interoperability gaps persist.

Integration Strategy and Governance Model
Duplication across different data programmes represents poor value for taxpayers and undermines the potential of health data to improve patient outcomes. However, instead of combining all existing initiatives, the priorities must be to build trust across the health-data community and to deliver a coherent national service through careful sequencing. A phased, 'best-of-breed' approach offers the optimal path forward. This involves identifying priority services aligned with key use cases, extracting the most effective elements from existing initiatives, and assembling these into a national service guided by a clear roadmap. This recognises a critical technical reality: a single monolithic data stack should not be pursued. Genomic and imaging data are simply too large and costly to centralise, making distributed approaches inevitable. Building a comprehensive longitudinal patient record is the technical foundation to pursue for the longer-term vision. This requires aligning with and extending central initiatives such as the SDEs and single-patient record programmes. However, this addresses only part of the requirement. The strategy must also solve for service delivery models and breadth of data access, particularly for imaging and genomics that will necessarily remain in situ.

A virtualised, federated architecture may provide one solution, delivering a 'network of nodes' that builds upon existing infrastructure. This architecture comprises an enhanced service layer providing consistent access across distributed data sources, coupled with a shared analytics toolkit enabling standardised methods and reproducible research. It enables flexible, organisation-specific, SDEs with 'Bring Your Own Data' [34] capabilities. For example, researchers can develop code offline and execute within controlled confidential environments with output review. Such distributed approaches deliver national-scale infrastructure without forced consolidation, balances accessibility with data protection, and future-proofs the system for increasingly complex analytical demands.

The HDRS must address growing concerns around cybersecurity, data privacy, and the responsible use of AI. The number of cyber-attacks in healthcare has increased significantly over the past five years with healthcare becoming the most targeted critical sector.[35] Cyber-attacks and data breaches threaten public trust which is a core component of data sharing. As a result of the increasing number of cyberattacks on healthcare globally, people may be less willing to share data for care or research. In ransomware attacks, hackers may encrypt or corrupt records and disrupt data integrity. Even when systems are restored, data accuracy may still be in question and have a significant impact on the utility of the impacted data. The HDRS need to prioritise cybersecurity and data privacy by design to ensure threats can be mitigated where possible.

Avoiding Planning Blight
The short-term alignment strategy must be to establish shared standards and application-programming interfaces for future compatibility. This requires proactive engagement with regional initiatives to prevent paralysis in strategic planning. Long-term planning must provide clarity on development timelines, phased capability delivery, and integration pathways. The integration strategy must prioritise interoperability, modularity, and minimal disruption to functioning systems. Rather than wholesale replacement, the approach must build upon existing infrastructure and borrow from proven initiatives for potential templates. This ensures data remains under the governance of original custodians whilst facilitating insights that drive efficiency and improved outcomes. SDEs should be examined selectively and effective ones scaled nationally. A clear understanding of which initiatives should be included within HDRS must be defined, given that there have been less successful data programmes, as well as ones that fulfil a niche which may not be met by HDRS.

Regional concerns about maintaining momentum must be acknowledged and ensure that local data capabilities are not lost. Fundamental uncertainty about the HDRS development timeline compounds these concerns. Without a clear roadmap showing which capabilities will be available when, organisations cannot determine whether to continue development, pause to avoid duplication, or begin planning for integration. The strategy must therefore include transparent timelines, clear communication of migration paths, and commitment to preserving functioning capabilities throughout the transition. This will ensure continuous service delivery while building toward an integrated, interoperable national health data infrastructure.

Technical Requirements: Architecture and Components
The 'federated' versus centralised architecture decision has implications for the technical requirements needed. The federated approach would minimise unnecessary data movement from source systems whilst leveraging existing proven capabilities. Open, system-wide standards, including common-data models such as the Observational Medical Outcomes Partnership (OMOP) Common Data Model[36] and standardised vocabularies, must underpin any chosen architecture. This will enable consistency, scalability, and effective integration across services. Data interoperability requires both structural standardisation through common-data models and semantic standardisation through shared terminologies.

Establishing clarity around metadata (what data exists, where it is stored, its standardised format, and how it can be accessed) is an immediate priority. Secure data-access environments building upon the SDE model must incorporate strict role-based access permissions, audit trails, and computational access principles. Metadata registries and data catalogues connecting distributed providers enable researchers to identify datasets and leverage standardised analytics tools without compromising security.

Skills and Sustainability

The HDRS necessitates a multidisciplinary workforce that transcends traditional data management, focusing on three key areas. Staff must be proficient in AI/ML model development, deployment, and governance, including data engineers, machine learning specialists, and data scientists, to support next-generation research. Professionals with strong commercial skills are required for managing industry interaction, negotiating complex pricing structures, and ensuring the service delivers demonstrable value and return on investment. The team must possess an intimate knowledge of health data and its clinical context to ensure data integrity, ethical compliance, and that research outputs meet actual clinical and public health needs, mitigating the risk of operational failure or user misalignment.

Critical operational questions about staffing and sustainability require policy attention to ensure any fundamental gaps can by addressed. Key considerations include defining workforce requirements, identifying necessary specialist skills in service design, commercial service and product design and delivery, data engineering, information governance, and developing competitive recruitment strategies against private-sector salaries. Protected funding streams independent of political cycles need to be available to ensure long term planning. It is vital for the success of the HDRS to attract and retain these skills. This will create clear organisational ownership, accountability frameworks, and develop sustainable financial models that extend beyond the initial £600 million mandate. Setting out these operational needs during the planning phase, rather than reactively, are essential for avoiding the trajectory of previous NHS IT initiatives that experienced operational difficulties.

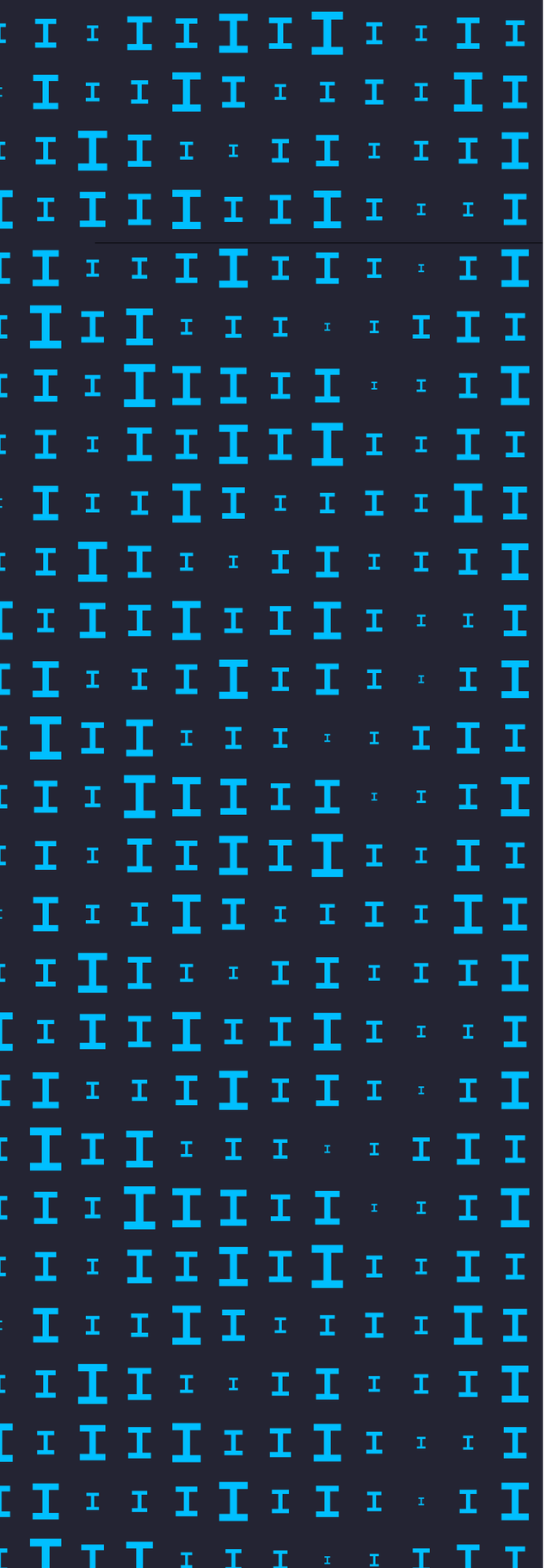
Governance Structure

The HDRS must be positioned clearly within the health system to operate effectively and independently. Three main options exist:

- **Within DHSC:** A Government Company (GovCo) which aligns with policy and enables administrative efficiency, but risks limited independence and prioritisation conflicts.
- **Arm's-Length Body (ALB):** Offers operational autonomy and credibility across sectors, though funding and accountability may be more complex.
- **Bespoke Statutory Body:** Provides the strongest independent governance and stable funding but requires legislation and longer setup time.

This choice determines HDRS's credibility, funding stability, workforce appeal, and ability to partner across NHS, academia and industry. Early clarity on structure is essential to guide governance, finance and recruitment, and to prevent 'planning blight' for ongoing initiatives. Alongside this, governance groups must be considered including funders, charities, PPIE and NHS representation. A GovCo is an entity fully or majority owned by the government to oversee designated commercial operations/assets while operating with greater autonomy than a conventional government department. The GovCo approach enables administrative efficiency and will mitigate the risk of a protracted set-up time. It will be important to outline the Wellcome Trust's role in HDRS's governance and accountability, and how this partnership can be enhanced using the Wellcome Trust's capabilities and expertise.

The HDRS must address growing concerns around cybersecurity, data privacy, and the responsible use of AI. The number of cyber-attacks in healthcare has increased significantly over the past five years with healthcare becoming the most targeted critical sector.



5. Primary Care Data

Primary care data represents the most comprehensive longitudinal health record for the UK. It is also one of the most under-utilised data sets for national-scale change for healthcare. Incorporating primary care data into the HDRS will be critical for the success of the service and to help address population health and prevention at scale. This will only happen if GPs and patients see clear, tangible benefits that outweigh any perceived risks or burdens. Patient trust and engagement is paramount. People need to see the justification for this data usage with concrete examples of how care can be improved, learning directly from the GPDPR failure. This endeavour will be a complex, multi-year process requiring technical alignment, trust, and collaboration across the system. Success will be dependent on genuine buy-in from GPs, vendors, and patients, with clear value, transparency, and shared benefit driving sustained participation.

The technical foundation must prioritise automated, passive data extraction using GP system APIs that operate with minimal practice burden, and standardised feeds across all major primary care electronic healthcare record (EHR) systems (principally EMIS, SystemOne). Practices must also see compelling returns on their participation through tools that directly support performance and patient care.

Primary care practices must also receive compensation for contributing data to the HDRS. Fair financial compensation should cover up-front infrastructure costs. Reimbursement can also include per-record contributions, quality bonuses, or annual infrastructure grants. In-kind benefits such as analytics tools, population-health insights, and funded data-management support can also reward participation.

Legal Liability

One of the biggest issues preventing this integration is the perception of legal risk for GPs if they share their data with a third party (i.e. HDRS or the end user of HDRS). The case for clarifying the legal basis on which data would be shared and the corresponding liabilities of each party is compelling, as further set out in Section 3 above. Likewise, patient confidentiality laws present a material legal hurdle for the operation of the HDRS (as set out in Section 3) which will need to be resolved.

At present, primary care practices operate as independent data controllers under UK GDPR, bearing legal responsibility for the data they hold. There is significant concern among these practices that they will be held liable for breaches resulting from or connected to downstream uses of data by HDRS or other third parties if they share their data. The DHSC via the Secretary of State for Health and Social Care or HDRS can potentially take on this responsibility depending on how HDRS as a corporate entity is structured. For example, clarifying that GP practices would share healthcare data with HDRS on a controller-to-controller basis (being separate, not joint, controllers), for a complementary purpose, on the basis of legitimate interests and scientific research, seeking to ensure that original data controllers will remain liable only for their own acts and omissions with respect to the shared data; if this is the case, then appropriate data notices will need to be provided to data subjects, legitimate interest assessments and data protection impact assessments should be performed, and robust data sharing contracts entered into, to ensure that such sharing is lawful.

However, designating DHSC via the Secretary of State for Health and Social Care or HDRS as a separate entity as data controller will require careful policy justification. It must take a considered approach with active stakeholder participation. Concerns about diminished local accountability and concentration of power in a single entity must be addressed proactively. International examples include Finland's Findata model [37,38] which operates as a centralised data permit authority for the secondary use of health data when combining data from multiple sources. Under this model, data sources including primary care operate under clear legal frameworks, with Findata granting permits for secondary use when data originates from multiple public data controllers, private service providers, or the Kanta Services.[38] Practices retain responsibility only for data quality, accuracy, and respecting patient opt-outs at source. This preserves appropriate clinical accountability without imposing systemic infrastructure risks beyond their control. Denmark's system similarly centralises responsibility for health data through Statistics Denmark and the Danish Health Data Authority,[39] which manage national health registers and provide access to researchers through secure environments. These models have achieved high participation rates and enabled extensive research as they removed the legal concerns that might otherwise obstruct engagement.[40]

Maintaining public and professional confidence is essential and will rely on robust safeguards. These may include independent oversight mechanisms, and meaningful patient and professional representation. The Finnish model shows that such frameworks must balance comprehensive data-protection requirements with practical research needs. Overly rigid processing requirements can paradoxically hinder the research the system aims to enable.

Data Quality and Integrity

The overall quality of primary care data remains uneven and will require significant curation to extract maximal value. As generative AI and advanced analytics improve our ability to interpret unstructured data, the foundational principle remains: analytical outputs are only as reliable as the data that underpins them. While the UK benefits from a relatively homogeneous primary care EHR landscape, with a roughly 60/40 split between the two dominant vendors (EMIS and SystemOne), significant variation in data completeness, consistency, and accuracy persists across practices. Recent trends have further complicated data quality. Multiple third-party systems, including remote monitoring tools, digital scribes, and triage platforms, now write directly into primary care records. This has increased data volume but not necessarily data integrity or standardisation. Variation in data-capture processes and inconsistent adherence to coding standards continue to limit data usability.

Sensitive Non-Health Information

The primary care record is distinct in its inclusion of highly sensitive, non-clinical information such as safeguarding notes and disclosures of domestic abuse. Established practices exist for restricting access to such data within primary care, but new data-sharing frameworks must include clear, auditable processes to prevent inappropriate disclosure. Risk-mitigation protocols, including selective data exclusion and robust consent management, should be core features of HDRS data governance.

Building and Maintaining Staff Trust

Addressing legal liability, whilst necessary, is insufficient to secure professional participation in the HDRS. Professional engagement is crucial for the success of integrating primary care data to a greater extent. Professional engagement cannot be tokenistic and requires protected budgets from the outset to train and educate professionals and GPs in governance structures. The system must clearly demonstrate reciprocal value. Whilst the use case for clinical care integration is immediately apparent, research benefits require careful translation into tangible improvements in service delivery and patient outcomes. Success depends on establishing genuine dialogue about data commercialisation, transparent benefit-sharing mechanisms, and demonstrable evidence that primary care data-sharing advances both individual patient care and public good.

Neighbourhood Health Services

The Health Data Research Service (HDRS) directly enables the government's strategic shift from hospital-based to community-based care through the neighbourhood health model. Launched in January 2025, Neighbourhood Health Services[6] operationalise

one of three key shifts in Lord Darzi's 2024 NHS Review by enabling a move from hospital to community care.[32] The guidelines mandate that integrated care boards ensure person-level, longitudinal, linked datasets encompassing general practice, community and public-health data. The HDRS provides the foundational data infrastructure for this transition by enabling secure, lawful access to linked health datasets across organisational boundaries. Two of the Neighbourhood Health Service's objectives can be addressed with the HDRS. Firstly, there can be consistent population health management methods to segment and risk stratify populations. Secondly, there can be a priority on patients with complex needs (approximately 7% of the population associated with 46% of hospital costs[6]) to prevent unnecessary hospital admissions and residential care placements whilst improving independence. The guidelines specify clinical data systems must support care navigation, case management and risk-based prioritisation. The HDRS can thus deliver tangible value by operationalising the data ecosystem essential for neighbourhood multidisciplinary teams.

Primary care data represents the most comprehensive longitudinal health record for the UK. It is also one of the most under-utilised data sets for national-scale change for healthcare. Incorporating primary care data into the HDRS will be critical for the success of the service and to help address population health and prevention at scale.

6. Value Generation and Measurement

Dimensions of Value
Determining appropriate value for health data access requires moving beyond simplistic transactional models to recognise multifaceted value creation across clinical, financial, and societal domains as well as recognition of the UK's standing in a globally competitive marketplace. The NHS should receive fair value for data use, whether financial or in-kind, particularly when used on a large scale for the development of AI models. However, value consideration should balance the opportunity to deliver to patients better clinical outcomes, improved medications and services, and enhanced public health provision with the opportunity to drive investment into the UK. This could enable jobs, growth and wider prosperity as well as the long-term sustainability of HDRS as a standalone organisation.

Our first report in 2020[41] described three types of value: health and social, economic and societal, and financial:

1. **Health and social value:** provide benefits to patients and to the public by using data to improve preventative care, have access to new medicines and devices and enable more accessible and cost-effective provision of care
2. **Economic and societal value:** create jobs and economic growth by enabling the life sciences and technology industries to develop data driven solutions, technologies and therapeutic interventions that directly benefit people in the UK
3. **Financial value:** provide income streams for the NHS through appropriate licensing and value-sharing agreements with the right partners.

Mechanisms for Value Generation
The HDRS has received significant pledged government funding of £500 million with a further £100 million from the Wellcome Trust.[3] To represent value for taxpayers, the HDRS must ultimately achieve self-sustainability through a hybrid funding model that balances commercial revenue generation with public value delivery.[43] Charges for access need to include a decision fee, the costs involved in a data controller extracting data, the costs of HDRS cleaning or combining data and the cost for using any secure-data platforms. Key revenue streams should include tiered licensing fees for commercial data access, with academic users receiving subsidised access. International precedents demonstrate the viability of such models, with Finland's Findata[44] and Denmark's health data[45] authorities successfully operating cost-recovery mechanisms whilst maintaining public benefit priorities. Research services, analytics consulting, and revenue-sharing agreements providing royalties from products developed using HDRS data offer additional income potential. The opportunity to develop products, either within HDRS or to enable a UK-based SME ecosystem, will be critical to delivering value. Policy frameworks need to retain these commercial activities within the UK as they could be a significant growth driver. An example would be differentiated data access to trusted UK research environments.

Supplementary value creation could emerge from HDRS or the NHS developing and licensing clinical decision-support tools, providing data quality and interoperability services to healthcare organisations, and offering training and certification programmes. To maintain public trust, commercial revenues must be transparently reinvested into health-system infrastructure, public-benefit research, and patient-engagement activities. This approach leverages network effects where increasing participation and longitudinal data accumulation enhance value over time, demonstrating measurable returns through improved health outcomes and system efficiencies benefiting the broader economy. This can be demonstrated by publishing audited financial reports, tracking measurable impact metrics, establishing independent oversight, demonstrating research outcomes, and communicating reinvestment benefits accessibly to build public trust.

Measuring Value
The HDRS must establish baseline measurements at inception, set phased implementation plans, and track progress systematically to ensure sufficient value exchange. Comprehensive value measurement requires the integration of quantitative and qualitative metrics within robust evaluation frameworks.

Economic impact metrics should capture healthcare cost savings, return on investment, commercial revenue generation, regional economic development, and workforce productivity gains.[46] These measures align with broader national objectives for life sciences competitiveness and economic growth through health-data innovation. Research excellence indicators must track peer-reviewed publications, the number of clinical trials supported, and novel discoveries successfully translated into clinical practice.[47] As *The O'Shaughnessy Review* into clinical and commercial trials emphasised, reducing the time and cost barriers to conducting commercial clinical trials represents a critical measure of infrastructure effectiveness, with explicit targets for trial set-up timelines providing concrete benchmarks for success.[48]

System performance requires monitoring data quality, reliability, interoperability achievements, utilisation rates, and user satisfaction metrics. Trust and legitimacy measures should assess public confidence through regular engagement surveys, compliance audits, and governance effectiveness reviews. Health outcomes, whilst requiring extended observation periods, constitute the ultimate validation of infrastructure value and should include mortality and morbidity reductions, quality adjusted life years gained, health inequalities narrowed, and preventable disease burden avoided. Evaluation frameworks must emphasise continuous monitoring with periodic independent review, measuring progress incrementally over time rather than attempting comprehensive transformation assessment at single time points. This approach reflects lessons from previous health data initiatives that promised revolutionary change but struggled to demonstrate appropriate progress against defined milestones, ultimately undermining stakeholder confidence and political support.[49]

Funding Models and Industry Collaboration
The HDRS's ability to attract investment and its long-term sustainability beyond initial government funding must be seen as a key driver of its development. This will require a clear articulation of the key industry use cases and value to them, in the context of global competition, to drive foreign direct investment into the UK. Recouping costs through a pay-to-play model is a positive start. However, it potentially undervalues the utility of the HDRS as a life sciences asset. A potential of measure of its success will be based upon institutional investment. Sustainable operation requires diversified funding streams that extend beyond initial government investment. Philanthropic and industry collaboration represents a crucial pathway toward financial sustainability. This will require carefully constructed frameworks that ensure mutual benefit whilst protecting public interests and maintaining trust. *The O'Shaughnessy Review* highlighted the imperative of creating an environment where commercial research can thrive whilst delivering public benefit. It also noted that regulatory efficiency, data accessibility, and that clear value propositions for industry investment constitute essential prerequisites for attracting commercial research activity.[48]

The HDRS should function as a single service, providing an interface and set of user experiences for engagement, eliminating the current fragmented landscape where companies navigate multiple NHS organisations with inconsistent requirements and conflicting technical specifications. This involves publishing clear service-level agreements and transparent access frameworks that define permissible use cases, standardised application processes with defined timelines and clear pricing structures.

Effective prioritisation frameworks require substantive patient co-design moving beyond tokenistic engagement to create mutually beneficial access policies and pricing structures. Researchers benefit through enhanced analytical capabilities, and industry partners through accelerated development timelines and robust evidence generation. Only through establishing and demonstrating this mutual benefit can the HDRS achieve the broad-based support necessary for long-term viability. Transparent reinvestment of commercial revenues into health-system infrastructure, public-benefit research, and patient-engagement activities remains essential for maintaining public legitimacy.[50] This approach recognises health data as a public asset requiring stewardship models that balance accessibility for public-benefit research with appropriate value capture from commercial applications. This ensures sustainability without compromising the foundational principle of health data serving the public good.

Implementation Roadmap

We propose the HDRS is implemented through a phased approach, establishing governance and infrastructure in 2026 before scaling to population-level integration by 2031, underpinned by a transparent opt-out framework that balances citizen control with essential NHS planning needs.

Phase 1: 2026

Implementation Timeline

KEY MILESTONES:

Q1: CEO appointed, Hiring of executive team/secondments

Q2: Board established

Q3: MVP launch - 25 research teams, 2-3 SDEs integrated

Opt-out system launched with public awareness campaign

Q4: Scale to 4-6 SDEs; datasets deployed covering primary care, Hospital Episodes Statistics (HES), prescribing

Clinical trials offering building on a Common Data Platform

Unite consented cohorts of existing data

Phase 2: 2027-28

KEY MILESTONES:

Full integration with Genomics England, trial matching capabilities

Streamlined governance reform, devolved nations partnering

AI models for discovery projects; MHRA/NICE-grade evidence generation

Phase 3: 2029-31

KEY MILESTONES:

Full population-scale coverage: GP, HES, prescribing, mortality data

AI platform with 50+ projects; MHRA/ NICE-grade evidence generation

60-70% cost recovery achieved whilst prioritising public benefit

In conclusion, this proposed implementation roadmap for the HDRS represents a pragmatic yet ambitious pathway toward a sustainable, ethically grounded, and innovation-enabling national-data infrastructure. The HDRS must build trust and demonstrate value iteratively. Embedding transparency, patient co-design, and equitable reinvestment throughout each stage will be critical. There is an urgent need to maintain public legitimacy and ensure that commercial collaboration aligns with public benefit. Success will depend not only on technical excellence but on demonstrating that data-driven health innovation can co-exist with the values of stewardship, accountability and inclusivity. Through this phased, ethically anchored, and partnership-oriented approach, the HDRS can establish the foundations of a world-leading health-data ecosystem.

References

1. Ghafur S, O'Brien N, Howitt P, Painter A, O'Shaughnessy J, Darzi A. NHS data: Maximising its impact for all. Imperial College London [2023]

2. Department of Health and Social Care, 2025. The government's 2025 mandate to NHS England and NHS Operational Planning Guidance 2025-26. Statement made on 30 January 2025. Statement UIN HCWS400. https://questions-statements.parliament.uk/written-statements/detail/2025-01-30/hcws400

3. UK Government [2025] Prime Minister turbocharges medical research. 7 April. Available at: https://www.gov.uk/government/news/prime-minister-turbocharges-medical-research [Accessed: 8th May 2025].

4. Department of Health and Social Care, Prime Minister's Office, 10 Downing Street, Sir Keir Starmer and Wes Streeting [2025] Fit for the future: The 10 Year Health Plan for England. Department of Health and Social Care. Published 3 July 2025. Available at: https://www.gov.uk/government/publications/10-year-health-plan-for-england-fit-for-the-future [Accessed: 6th August 2025].

5. Department for Business and Trade; Department of Health and Social Care; Department for Science, Innovation and Technology; Office for Life Sciences [2025] Life Sciences Sector Plan. Published 16 July 2025. Available at: https://www.gov.uk/government/publications/life-sciences-sector-plan [Accessed: 6th August 2025].

6. NHS England (2025) Neighbourhood Health Guidelines 2025/26. Available at: https://www.england.nhs.uk/long-read/neighbourhood-health-guidelines-2025-26/ (Accessed: 5th November 2025).

7. Department for Science, Innovation and Technology [2025] AI Opportunities Action Plan. London: HM Government. Available at: https://www.gov.uk/government/publications/ai-opportunities-action-plan/ai-opportunities-action-plan [Accessed: 9 October 2025].

8. Health Data Research UK [2025] The Sudlow Review. Available at: https://www.hdruk.ac.uk/helping-with-health-data/the-sudlow-review/ [Accessed: 3 April 2025].

9. NHS Digital [2025] Research powered by data. Available at: https://digital.nhs.uk/data-and-information/research-powered-by-data [Accessed: 3 April 2025].

10. UK Parliament [2025] Data Protection and Digital Information Bill. Available at: https://bills.parliament.uk/bills/3825 [Accessed: 3 April 2025].

11. Department of Health and Social Care [2025] Independent Investigation of the NHS in England. Available at: https://www.gov.uk/government/publications/independent-investigation-of-the-nhs-in-england [Accessed: 3 April 2025].

12. OpenSAFELY [2025] OpenSAFELY: secure analytics platform for electronic health records research. Available at: https://www.opensafely.org/ [Accessed: 3 April 2025].

13. NHS Digital [2025] General Practice Data for Planning and Research. Available at: https://digital.nhs.uk/services/general-practice-data-for-planning-and-research [Accessed: 3 April 2025].

14. NHS, 2025. The NHS Research Secure Data Environment Network. [online] NHS Digital. Available at: https://digital.nhs.uk/data-and-information/research-powered-by-data/sde-network [Accessed 8 May 2025].

15. NHS Digital [2021] Public attitudes to data in the NHS and social care. Available at: https://digital.nhs.uk/data-and-information/keeping-data-safe-and-benefitting-the-public/public-attitudes-to-data-in-the-nhs-and-social-care [Accessed: 3 April 2025].

16. NHS Digital [2025] Powerful Moments. Available at: https://digital.nhs.uk/data-and-information/keeping-data-safe-and-benefitting-the-public/powerful-moments [Accessed: 3 April 2025].

17. RECOVERY Trial [2025] Randomised Evaluation of COVID-19 Therapy. Available at: https://www.recoverytrial.net/ [Accessed: 3 April 2025].

18. NHS England (2025) National engagement on data: cohort 3 report. Available at: https://transform.england.nhs.uk/key-tools-and-info/data-saves-lives/national-public-engagement-on-the-use-of-health-data/national-engagement-on-data-cohort-3-report/ (Accessed: 5 November 2025).

19. Department of Health and Social Care [2022] Data Saves Lives: reshaping health and social care with data. Available at: https://www.gov.uk/government/publications/data-saves-lives-reshaping-health-and-social-care-with-data/data-saves-lives-reshaping-health-and-social-care-with-data [Accessed: 3 April 2025].

20. OneLondon [2025] Available at: https://www.onelondon.online/ [Accessed: 3 April 2025].

21. Community First Yorkshire [2025] Join the Yorkshire & Humber Secure Data Environment Citizens' Jury. Available at: https://www.communityfirstyorkshire.org.uk/join-the-yorkshire-humber-secure-data-environment-citizens-jury/ [Accessed: 3 April 2025].

22. Environment Agency [2025] Citizens' Jury for Thames Valley South Chilterns. Available at: https://consult.environment-agency.gov.uk/thames/citizens-jury-for-thames-valley-south-chilterns/ [Accessed: 3 April 2025].

23. Our Future Health [2025] Our Future Health: helping people live healthier lives for longer through research. Available at: https://ourfuturehealth.org.uk/ [Accessed: 3 April 2025].

24. Our Future Health [2025] The UK shows its support for Our Future Health. Available at: https://ourfuturehealth.org.uk/news/the-uk-shows-its-support-for-our-future-health/ [Accessed: 17 October 2025].

25. Department of Health and Social Care (2020) Creating the right framework to realise the benefits for patients and the NHS where data underpins innovation. Available at: https://www.gov.uk/government/publications/creating-the-right-framework-to-realise-the-benefits-of-health-data/creating-the-right-framework-to-realise-the-benefits-for-patients-and-the-nhs-where-data-underpins-innovation (Accessed: 1st December 2025).

26. Health Data Research UK (2020) 'A pioneering programme to establish a national infrastructure for health data science', Health Data Research UK, 21 May. Available at: https://www.hdruk.ac.uk/news/a-pioneering-programme-to-establish-a-national-infrastructure-for-health-data-science/ (Accessed: 1 December 2025).

27. Information Commissioner's Office [2025]; The UK GDPR. Available at: https://ico.org.uk/for-organisations/data-protection-and-the-eu/data-protection-and-the-eu-in-detail/the-uk-gdpr/ [Accessed: 3 April 2025].

28. NDRS (2025) Clinical Practice Research DataLink. Available at: https://digital.nhs.uk/ndrs/our-work/ncras-partnerships/cprd (Accessed: 1 December 2025)

29. Acharya, A., Dryden, S., Kohli, P., Mason, J., Papa, E., Singh, A., Soni, L., Darzi, A., & Koivuniemi, A. (2025) Harnessing AI to tackle AMR, Available at: https://www.imperial.ac.uk/Stories/harnessing-artificial-intelligence-tackle-antimicrobial-resistance/ (Accessed: 1 December 2025).

30. World Health Organization (2021) Ethics and governance of artificial intelligence for health. Geneva: World Health Organization. Available at: https://www.who.int/publications/i/item/9789240029200 (Accessed: 1st December 2025).

31. NHS Digital [2025] Secure Data Environment Service. Available at: https://digital.nhs.uk/services/secure-data-environment-service [Accessed: 3 April 2025].

32. Darzi, A. (2024) Independent Investigation of the National Health Service in England. London: Department of Health and Social Care. Available at: https://www.gov.uk/government/publications/independent-investigation-of-the-nhs-in-england (Accessed: 5th November 2025).

33. NHS England [2025] NHS Federated Data Platform. Available at: https://www.england.nhs.uk/digitaltechnology/nhs-federated-data-platform/ [Accessed: 3 April 2025].

34. Martin J, Cantero D, González M, Cabrera A, Larrañaga M, Maltezos E, Lioupis P, Kosyvas D, Karagiannidis L, Ouzounoglou E, Amditis A. Embedded Vision Intelligence for the Safety of Smart Cities. J Imaging. 2022 Dec 14;8[12]:326. doi: 10.3390/jimaging8120326. PMID: 36547491; PMCID: PMC9783517.

35. Ewoh P, Vartiainen T. Vulnerability to Cyberattacks and Sociotechnical Solutions for Health Care Systems: Systematic Review. J Med Internet Res. 2024 May 31;26:e46904. doi: 10.2196/46904. PMID: 38820579; PMCID: PMC11179043.

36. OHDSI; Standardized Data: The OMOP Common Data Model. Available at: https://www.ohdsi.org/data-standardization/ [Accessed: 23 October 2025].

37. Brück O, Sanmark E, Ponkilainen V, Bützow A, Reito A, Kauppila JH, Kuitunen I. European health regulations reduce registry-based research. Health Res Policy Syst. 2024 Sep 30;22(1):135. doi: 10.1186/s12961-024-01228-1. PMID: 39350115; PMCID: PMC11443657.

38. Findata [2024] Data. Available at: https://findata.fi/en/data/ [Accessed: 23 October 2025].

39. Statistics Denmark [2024] Data from other sources. Available at: https://www.dst.dk/en/TilSalg/data-til-forskning/generelt-om-data/data-fra-andre-kilder [Accessed: 23 October 2025].

40. Koivisto J, Tiirinki H, Liukko E. Identifying Individuals for Integrated Multidisciplinary Care: Lessons from Finland. Int J Integr Care. 2022 Aug 12;22(3):8. doi: 10.5334/ijic.6000. PMID: 36043026; PMCID: PMC9374024.

41. Imperial College London, Institute of Global Health Innovation (2020) NHS Data: Maximising its impact on the health and wealth of the United Kingdom. London: Imperial College London. Available at: https://spiral.imperial.ac.uk/handle/10044/1/76409 (Accessed: 5th November 2025).

42. Tennison, J. [2025]. 'The National Data Library and public benefit.' Bennett Institute for Public Policy, University of Cambridge, & Connected by Data.

43. NHS England [2025] Health Data Research Service - unlocking the potential of health and care data to transform lives. Available at: https://www.england.nhs.uk/blog/health-data-research-service-unlocking-the-potential-of-health-and-care-data-to-transform-lives/ [Accessed: 23 October 2025].

44. Findata [2025] Legislation - Act on the Secondary Use of Health and Social Data. Available at: https://findata.fi/en/services-and-instructions/legislation/ [Accessed: 23 October 2025].

45. Invest in Denmark [2022] 'New national Danish Research Health Data Gateway assists researchers with accessing data'. Available at: https://investindk.com/insights/new-national-danish-research-health-data-gateway-assists-researchers-with-accessing-data [Accessed: 23 October 2025].

46. Sanders, G.D., Neumann, P.J., Basu, A., Brock, D.W., Feeny, D., Krahn, M., Kuntz, K.M., Meltzer, D.O., Owens, D.K., Prosser, L.A., Salomon, J.A., Sculpher, M.J., Trikalinos, T.A., Russell, L.B., Siegel, J.E. and Ganiats, T.G. [2016] 'Recommendations for conduct, methodological practices, and reporting of cost-effectiveness analyses: Second Panel on Cost-Effectiveness in Health and Medicine', JAMA, 316[10], pp. 1093-1103. doi: 10.1001/jama.2016.12195.

47. Mittelstadt, B.D. and Floridi, L. [2016] 'The ethics of big data: current and foreseeable issues in biomedical contexts', Science and Engineering Ethics, 22[2], pp. 303-341. doi: 10.1007/s11948-015-9652-2.

48. O'Shaughnessy, J. [2023] Commercial clinical trials in the UK: The Lord O'Shaughnessy review - Final report. London: Department of Health and Social Care. Available at: https://www.gov.uk/government/publications/commercial-clinical-trials-in-the-uk-the-lord-oshaughnessy-review [Accessed: 23 October 2025].

49. Papanicolas, I., Mossialos, E., Gundersen, A., Woskie, L. and Jha, A.K. [2019] 'Performance of UK National Health Service compared with other high income countries: observational study', BMJ, 367, l6326. doi: 10.1136/bmj.l6326.

50. Mc Cord, K.A., Al-Shahi Salman, R., Treweek, S., Gardner, H., Strech, D., Whiteley, W., Ioannidis, J.P.A. and Hemkens, L.G. [2018] 'Routinely collected data for randomised trials: promises, barriers, and implications', Trials, 19[1], 29. doi: 10.1186/s13063-017-2394-5.