The Digital Innovation Hub Programme

Discussion Paper of the Emerging Model

A first step towards

A UK-Wide Infrastructure for Health Data Research and Innovation

Purpose: This paper aims to stimulate further discussion on the emerging model for the Digital Innovation Hub Programme in advance of the publication of the final prospectus in May 2019. The programme, as announced in the Life Sciences Industrial Strategy, is an important first step in our quest to create a UK-wide infrastructure for health data research and innovation. We welcome organisations who would like to be involved in helping to shape this exciting and emerging model to contact us at enquiries@hdruk.ac.uk

11 March 2019

1. Overview

The UK has some of the richest health care and research data assets world-wide. Some assets are well organised, such as the Clinical Practice Research Database (CPRD) and UK Biobank, but only a fraction of NHS and research data is accessible and being used for research and innovation. Achieving this across the UK will create game-changing capability in identifying causes of disease; the analysis of genetic, lifestyle and social circumstances to prevent disease; using data to spot disease trends and make care safer and more personal, as well as supporting the discovery of new treatments and medicines.

To address this need, the UK Government has committed £37.5m to develop interoperable **Digital Innovation Hubs** (DIHs). This is as an important first step towards a national approach to enable the safe and responsible use of health-related data at scale for research and innovation. The DIH programme, part of UK Research and Innovation's Industrial Strategy Challenge Fund (ISCF) Data to Early Diagnosis and Precision Medicine Challenge is being led by Health Data Research UK (HDR UK), the national institute for health data science. This programme will act as an accelerator to enable the UK to be at the leading-edge of the growth of the global big data analytics in the healthcare market and will ensure that benefits which are developed through increased use of data are returned to the NHS.

2. Vision

Our vision is to support improvements for patients and the public through an exponential increase in the use of UK health data for research and innovation.

Our aim is to increase the access and use of health data in a trustworthy and ethical way in order to develop improvements in the UK's health technology. This will be achieved through embedding a robust governance framework across the three layers of the DIH programme: (i) the **UK Health Data Research Alliance** which will facilitate partnership working across NHS organisations and a consistent approach to governance, data provision and public engagement; (ii) an **infrastructure layer** which will provide access, de-identification, policy, standards, tools and best practices for using data; and, (iii) the **Digital Innovation Hubs** which will provide specialised research services based on a particular area of expertise (e.g. disease area, therapeutic area, research activity, region).

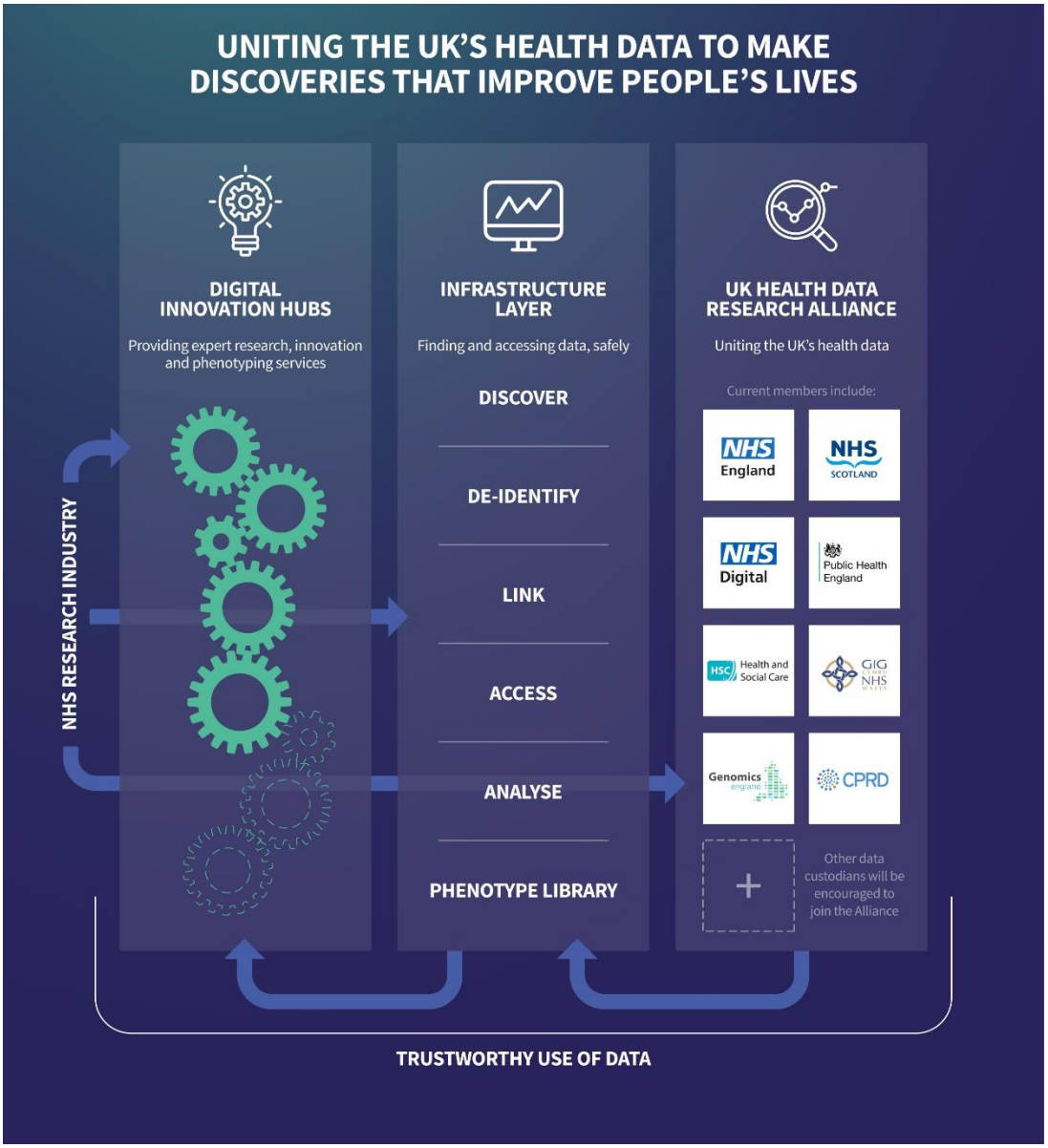
The DIH programme will therefore bring the NHS, academia and the public together with industry to establish a robust health data research infrastructure, with the aim to:

- Improve the clinical care of individual patients and health of the population
- Become a world leader in health data-led fields including genomics, 'real world trials', diagnostics and related areas
- Bring together pioneering infrastructure work and existing trusted research environments across the UK including National Institute of Health Research (NIHR) infrastructure (e.g. Health Informatics Collaborative, Clinical Research Networks), Secure Anonymised Information Linkage (SAIL) Databank in Wales, National Services Scotland (NSS) and NHS Digital, and engage with other initiatives such as the Digital Pathology, Imaging and AI centres of excellence and Advanced Therapy Treatment Centres
- Catalyse the development of high-energy areas of data-driven innovation and excellence on different data types including electronic patient records, imaging, "omics" and wearables between NHS, academia and industry across the UK.

Our aim is to act as the catalyst for data driven innovation in healthcare, not only making a significant contribution to the health of the nation, but to increase the UK's GDP by sparking and growing UK companies, and encouraging inward investment in diagnostics, imaging, artificial intelligence (AI), technology companies and pharmaceuticals

3. Model

The DIH programme will seed a thriving health data research ecosystem in the UK. The three layers of the DIHs are illustrated in the figure. The design and delivery of each will be co-created with public and patient involvement.



3.1 The UK Health Data Research Alliance

Established in December 2018, and announced in the Sector Deal 2, NHS England, NHS Scotland, NHS Wales, Health and Social Care Northern Ireland along with NHS Digital, Genomics England, Health Data Research UK, Public Health England and Clinical Practice Research Datalink (CPRD) are the founding members of the **UK Health Data Research Alliance** of health data controllers. This is a non-profit association to accelerate progress in medicine and health and is the first time these key bodies have come together with a focus on health data research.

The purpose of the Alliance is to unite expertise to establish best practice in the stewardship of the UK's health data – including patient data from the NHS, genomic data and other molecular data – **to enable faster, more efficient access for research at scale**. This will speed up progress in science and medicine at a pace never achieved before in the UK.

The function of the Alliance will be to develop and co-ordinate the adoption of tools, techniques, conventions, technologies and designs that enable the use of health data in trustworthy and ethical ways for research and innovation. Its founding members, along with public involvement, will formulate best practice and standards in areas such as privacy, transparency, public engagement, inclusivity and governance to ensure that health data is accessed and used responsibly by researchers and innovators. As the Alliance grows and develops, additional health data custodians will join, bringing further datasets to contribute regional and local depth to support further understanding. NHS trusts and other high-profile data controllers are now requesting to join the Alliance.

3.2 DIH Infrastructure

For those wishing to access health data across multiple datasets, it can be confusing, fractured and disjointed.

The purpose of the common DIH infrastructure layer is to provide a single, secure, accessible data infrastructure, underpinned by a consistent governance framework and advanced data discoverability tools (indexing, metadata etc.) to provide information on the available data. This will allow users to understand the array of data that has been made available through the Alliance, and the range of possibilities for linking this data with others to develop even greater insight. It will be delivered to robust industrial standards, cloud-enabled and available to access either directly or through a hub. The infrastructure layer itself will not hold any data but will compile and build on the availability through the individual Alliance members.

This will ensure that the programme delivers interoperability and secure data provision across the DIHs at scale. The design will be informed by dialogue with a number of national data custodians (e.g. NHS England, NHS Digital and equivalents in the devolved nations), existing large infrastructures in the NHS (e.g. Genomics England), research (e.g. UK Biobank) and industry.

This common technology layer will build on infrastructure and best practices that already exist in the UK. Wherever possible, we are working in partnership to evolve existing assets and commercial agreements. Where new solutions are needed, we are going to market for existing 'best of breed' solutions and will only custom develop where a suitable product or application does not already exist.

The infrastructure layer will be developed in stages. The initial delivery phase (from May 2019) will focus on proof of concept projects to define the critical functions (Minimum Viable Product).

The functions of the infrastructure will be to:

- **Discover data:** provide a single point of access providing a view of available datasets, enabled by federated metadata management and search tools to provide improved visibility, comparability and navigation of existing datasets.
- **Access data** that retains data controllers' requirements but helps harmonise processes, reduce transaction costs and improve access to more isolated, important datasets. This will also include a single point of access across multiple data sets and workspaces based on common authentication software and assigned access permissions.
- Use **privacy** and interoperability through **de-identification** where appropriate, in line with information governance requirements. Deploying best in class de-identification and encryption software across the Alliance will enable secure novel data linkages across a wide range of datasets to meet researcher needs while ensuring privacy of sensitive health data.
- Harness **interoperability** to support access to linked datasets across the data custodians in the Alliance. This will enable a single point of access across multiple datasets and workspaces based on common authentication. The interoperability support will also extend to support mapping between datasets. This will focus on enabling re-use of mapping carried out as part of planned research projects, as well as codifying interoperability issues between different datasets, for example through the calculation of an interoperability score which would be available to other researchers considering working with the same datasets.
- Provide an **open source repository** for common components such as a phenomics library, analytics code snippets and scripts.

3.3 Digital Innovation Hubs

The purpose of DIHs will be to create regionally-developed centres of excellence that provide expert research data services to enable NHS, academia and industry to create real world insights, innovations and improvements from structured and unstructured health data. Each Hub will create an interdisciplinary focal point for research and innovation that harnesses massive data, fast computation, and new approaches to data curation and analysis. Each will be a collaborative effort between NHS, industry and academia
The DIHs will develop common core approaches to key data services, facilitating co-ordination towards national coverage.

The Hubs will be formal collaborations between NHS, academic organisations and industry. They could be developed as new capability within existing organisations, joint ventures or new entities. They will initially be funded through the Industrial Strategy Challenge Fund (ISCF) to accelerate activity, with the expectation of local matched funding to increase impact.

They will operate in line with the National Data Guardian's recommendations on patient data[1] and will be capable of using and integrating diverse health data (e.g. from Local Health and Care Records), adapted to different baseline capabilities, frequently changing data requirements and diverse user needs.

[1] National Data Guardian, 2016. Review of data security, consent and opt-outs. Available at: https://www.gov.uk/government/publications/review-of-data-security-consent-and-opt-outs

Additional Hubs will be accredited and encouraged to use the DIH infrastructure and provide services for specific medical technologies, disease areas or data innovation, funded through means other than the initial ISCF investment. Such an example is the proposed UK Cardiovascular Data Science Knowledge Hub being developed in partnership by the British Heart Foundation (BHF) and HDR UK.

The function of the DIHs is to provide a range of research data services. They will be designed as advanced data handling and analytics environments, where academic, NHS and industry researchers will "breathe the same air" and be supported by people who understand the data, and a secure data processing area and flexible analytical workspace. Hubs will be able to act as a concierge service, using expert knowledge and skills to provide data curation, governance, analytics, computing power, physical location and safe havens for restricted data access within a trusted research environment.

Hubs will provide expert services needed by academic, industry, healthcare professional and NHS users, and will build on areas of specialty. Hubs should leverage distinctive areas of expertise that meet a specific need, and have an identified market and users for the services.. This expertise could be:

- **High value data curation services**: Providing expertise in specific areas of interest, e.g. disease specific datasets (rheumatoid, cancer, diabetes etc) and co-morbidities based on patient need and industry developments
- **Clinical trial services**: Providing expert services in large-scale data to dramatically increase the speed and quality of clinical trials by automating feasibility, eligibility, recruitment and follow-up; and providing innovation in trial design ideas that have emerged in 21st century trials embedded in health systems
- **Precision medicine services***:* Providing drug and medical treatment companies with expert services on linked clinical and genetic data for target validation, identifying early disease states and validating biomarkers, discovering and validating diagnostic and prognostic markers, using artificial intelligence (AI) and machine learning (ML) to develop more targeted and effective drugs and treatments
- **Regional insight:** Combining in-depth knowledge of a local area and context with rich regional datasets, this could provide specific insight into a local area and detailed public health information to support improvements in patient health and appropriate commissioning and evaluation of services
- **Analytical Research Platform:** Provision of data import from diverse sources into a secure research data warehouse; with variable quantities of data and ability to scale the warehousing capacity flexibly; ability to aggregate and link data items that match from different sources (e.g. EPR data with legacy data for the same patient in a Vendor Neutral Archive data, or EPR data with genomic/imaging data); and, application of common terminology.

Hubs will make the UK attractive for international investments and actively stimulating the start-up market, for example, by making fully anonymised data sets openly available as 'simulated data'. Hubs will be selected for initial funding where there is credible ongoing demand for these services from industry, with the expectation that each Hub will leverage the ISCF investment through matched-funding and will be self-sustaining after this initial funding.

Through the ISCF, Health Data Research UK will fund hubs to demonstrate the UK's value as a world leading research centre. We encourage other funders to develop their own 'digital innovation hubs' that work with and alongside the other hubs, but subscribe to the common ways of working, ethics frameworks,

governance principles, interfaces, protocols and standards of the DIH Infrastructure. Increasing the use and availability of data will bring greater benefits to patients and the public.

Example case studies:

James is a researcher working in a small company in Glasgow, which is developing new treatments for rare cancers. Following appropriate data governance processes, the Digital Innovation Hub has enabled him to discover and access a dataset of 1 million cancer patients across the UK combining NHS and genomic data. By working directly with the Hub, James has also accessed expert analytics services and world leading experts in cancer in the NHS and academia. The Hub services have ensured the data was usable, proportionate and directly applicable to his research. He is reassured that the data is accessed using trusted and secure networks and complies with the national governance standards, and this is made clear to the original owners of the data.

Alison is an NHS commissioner working to develop preventative services to support children and young people's mental health, one of the NHS Long Term Plan priorities. She is interested in which alternative to admission services would be most effective for her area, West Yorkshire. She approaches a Digital Innovation Hub to help understand the problem. The Hub helps to formulate the question and lets Alison know what information is available. Having identified the right datasets to support the research question, the Hub coordinates the data request and information governance processes. Having received the secure, anonymised information, members of the Hub team analyse the data in the trusted research environment, providing Alison with the outputs of this analysis. Informed by real world evidence, Alison is able to commission the most effective services for the children and young people in her area.

4. Relationship to other initiatives

This ISCF DIH programme is being managed and delivered as a ring-fenced programme within Health Data Research UK, but will add value and act as an accelerator to other parts of the health data research landscape:

- The Local Health and Care Record Exemplar (LHCRE) programme has been closely involved in the development of the DIH programme. LHCREs will participate through the Alliance bringing regional data and will be eligible to participate as part of a Hub.
- Health Data Research UK operates across six substantive sites that collectively comprise 21 university and research institutes across the UK. These sites are engaged in nationwide research programmes using health data. These sites will be encouraged to use data through the DIH programme and will also be eligible to participate in the DIH programme by developing Digital Innovation Hubs themselves. However, Hubs do not need to be linked to a Health Data Research UK site.
- Health Data Research UK separately funds the development of new career pathways and training for health data scientists who will be encouraged to work in the different parts of the DIH programme
- There are a number of ongoing digital health initiatives across the UK with which the DIH programme can link in order to demonstrate the potential impact on health outcomes, such as the Digital Pathology, Imaging and AI centres of excellence.

5. Governance

The Hubs, infrastructure and Alliance need to assure the public that they operate within trusted data standards and requirements for security, privacy, ethical approval, and that the status of national "opt-outs" are respected, as per the recommendations of the National Data Guardian (NDG).

Members of the public have been engaged throughout the programme, through events and webinars. The HDR UK Public Advisory Board will provide strategic advice on the DIH programme, with particular emphasis on ensuring that health data is used responsibly for research and innovation.

The DIH programme is inclusive and encourages a diversity of users. This means that the Alliance and infrastructure layer will remain open to direct access from researchers and users from the NHS, academia, industry and the third sector.

5.1 Progress to date

Since September 2018 Health Data Research UK has:

- Recruited dedicated DIH leadership and delivery team
- Established a Public Advisory Board to advise on a range of activities, including DIH Programme development
- Agreed governance with ISCF Steering Group
- Contributed to 27 industry, NHS, patient and academic events to understand requirements for the DIH Programme
- Run a competition and launched 10 Sprint Exemplar Innovation projects to showcase DIH capability
- Produced an infrastructure roadmap of the DIH Programme, in partnership with the NHS
- Launched the UK Health Data Research Alliance, in partnership with NHS Digital, featuring eight high profile NHS data custodians,

5.2 Next steps

In May 2019, Health Data Research UK will launch a call for applications for Digital Innovation Hubs. This will be supported by a detailed prospectus and roadmap. Key milestones for 2019/20 include:

- The operating model for the DIH programme will be refined through further consultation, including face-to-face interviews with a large number of industry partners, widely disseminated industry surveys (in partnership with ABPI, the Pistoia Alliance, BIA and the Medicines Discovery Catapult), eight NHS and academic roadshows across the UK and public/charity engagement events. This will feed into the development of the DIH prospectus
- A survey to ascertain which datasets are most in demand, and which do we want to prioritise making available and to what level of curation
- Infrastructure developmental projects will be requested from March 2019
- The inaugural UK Health Data Research Alliance meeting will be held in April 2019
- The DIH programme and competition for initial hubs will be launched formally in May 2019

- The vendors to develop specific components of the infrastructure layer (the Minimum Viable Product) to be selected in September 2019
- Sprint Exemplars will deliver by November 2019
- Successful hubs will be selected in Autumn 2019 and supported to go live in early 2020

Between February and May 2019, we will work closely with Alliance partners, UK Research and Innovation, the Office for Life Sciences, and key NHS partners to ensure the proposed model fully meets the needs of industry; earns public support and trust; has a sustainable business model; and is underpinned by robust information governance. This will be overseen by the HDR UK board and the Industrial Strategy Challenge Fund programme board.

6. Funding

While the initial £37.5m ISCF funds over the four years will allow the Alliance, infrastructure and Hubs to be developed, the programme will be made increasingly sustainable through attracting further funding from industry and academia, and through a commercial model that encourages data use and feeds benefits back to the NHS, in line with the Code of conduct for data-driven health and care technology. The commercial model is still subject to further development, but the below represents a possible approach:

- Fees for accessing datasets within the Alliance will be set by each custodian, as they currently are. These fees and terms will be published as part of the description of the infrastructure layer
- Use of the infrastructure layer will be on a subscription basis and there will be a small processing fee for each request based on the number of datasets requested and processing required. The subscription arrangements will provide a fair share of value to the Alliance members and cover the costs of the infrastructure service. The processed datasets may be returned to the Infrastructure layer for further use by other researchers. To stimulate use in the initial period, the fees for use of the infrastructure will be low, or waived
- The hubs will be encouraged to develop their own, sustainable, commercial models based on the services they deliver and the enhanced and processed data available directly from the hub
- Standard terms and conditions will be established to provide a transparent and consistent way of working across the DIH programme.