HDRUK
Health Data Research UK

UK Research and Innovation

# INDUSTRIAL STRATEGY CHALLENGE FUND
# DIGITAL INNOVATION HUB PROGRAMME
# **SUMMARY REPORT**

HDRUK
Health Data Research UK



# Contents

# Executive summary

The Digital Innovation Hub (DIH) Programme has achieved a step change in **improving the quality of research-ready data** and making it easier for researchers to discover, as well as developing **services to support its effective and trustworthy use**. The Innovate UK funding of £37.5 million with which the programme was established has **leveraged an additional £140 million of resources** to increase the scale and the impact of the programme.

Alongside the data curation and services provided by the Health Data Research Hubs, and the improved discoverability of data assets via the Innovation Gateway, the UK Health Data Research Alliance has achieved **consensus and convergence in key areas of data custodianship**.

Whilst the pandemic inevitably disrupted the programme, and in particular some of the engagement with industry users, it provided an unexpected opportunity to **demonstrate impactful use of UK health data for research and innovation**, with the emerging DIH infrastructure **contributing to both the national and international COVID-19 response**.

Looking to the future, the **components of the programme are being sustained**; providing infrastructure and services that are **building capacity to support a growing commercial research and innovation environment**. Lessons learned through the course of the programme are also **helping inform and shape new investments**.

***

The UK Government's Industrial Strategy (2017) recognised the untapped potential of the UK's health data as an opportunity to boost economic and health outcomes. As part of the Life Sciences Data to early diagnosis and precision medicine challenge (D2EDPM), Health Data Research UK (HDR UK) was funded by Innovate UK to establish the £37.5 million DIH Programme. Its aim was to enhance routine National Health Service (NHS) data and the UK's rich cohort, registry, and research data and make it available to industry, researchers, and innovators to use for impactful research.

Following an initial Design and Dialogue Phase that included consultation with over 1,200 stakeholders from across the four nations, HDR UK published a DIH Programme Prospectus in May 2019 that set out proposals to establish three new capabilities to address challenges faced by industry and academic researchers in accessing and using health data. These were:

- Seven Health Data Research Hubs (the Hubs): UK-wide centres of excellence which focus on data curation and create the expertise, tools, knowledge, and ways of working to maximise the insights and innovations developed from health data.
- The Health Data Research Innovation Gateway (the Gateway): a web-based platform to discover and request access to UK health datasets for research and innovation.
- The UK Health Data Research Alliance (the Alliance): an independent, legally non-binding alliance of health data providers, custodians and curators who develop and share standards, policies, and best practice to demonstrate trustworthiness in health data use for research to improve human health.

During the implementation phases (May 2019 to March 2023), HDR UK managed and distributed funding across its UK partners.

Less than six months after the Hubs were launched (in October 2019), the COVID-19 pandemic disrupted the research priorities and environment for life sciences and development work on the Innovation Gateway. All staff involved had to adapt to new ways of working, non-COVID-19 research was delayed or halted, clinicians were diverted away from research to deliver front-line care, and the ability of industry to set up new research programmes was severely hampered. Despite these challenges, the pandemic helped to accelerate the development of some aspects of the DIH infrastructure. For example, the Hubs delivered services and research to identify and address the direct and indirect impacts of COVID-19, including NHS DigiTrials' support to the RECOVERY trial - the world's largest clinical trial into treatments for COVID-19 - and the BREATHE Hub's research into vaccine safety and effectiveness, which shaped global policy. The Alliance and Gateway formed the basis of a major leveraged national COVID-19 research programme (the COVID-19 Data & Connectivity National Core Study (NCS)) as well as an international one (the International COVID-19 Data Alliance; ICODA).

Despite the challenges of the pandemic, and guided by the Principles for Participation, Hubs have enabled trustworthy use of data: the programme has engaged over 150 companies as partners or users, involving patients and the public in decision-making throughout.

The Gateway has provided a common portal for data discovery that far exceeds the scale of previous platforms for health data research. It has over 2,500 registered users and 10,000 or more searches per month across metadata from nearly 800 diverse data sets, as well as tools, papers, data uses and people. The Gateway provides a reference implementation of metadata specification, data use register standard and data access request process based on the Five Safes Framework (safe data, projects, people, settings, outputs) that have been developed by the Alliance (which now has a membership of over 70 data custodians). Its White Paper, *Building Trusted Research Environments - Principles and Best Practices; Towards TRE ecosystems* has been downloaded over 6,000 times and has helped shape the Secure Data Environment (SDE) Policy for the NHS in England.

The DIH Programme has created strong foundations and enduring infrastructure for future activities. The shared ecosystem components of the DIH Programme (Gateway and Alliance) are embedded as core features of HDR UK's long-term strategy, together with the necessary funding. Hubs have adapted their models to sustain their services and activities, with regionally focused Discover-NOW and PIONEER Hubs proving integral to the NHS England Data for Research and Development (R&D) investment.

The lessons learned through the DIH Programme have shaped and informed the approach to delivering subsequent significant research programmes both through HDR UK and beyond, including Research Data Scotland, Our Future Health and the aforementioned Data for R&D programme. They are shared in detail through this summary report and the associated component reports to help the work of others.

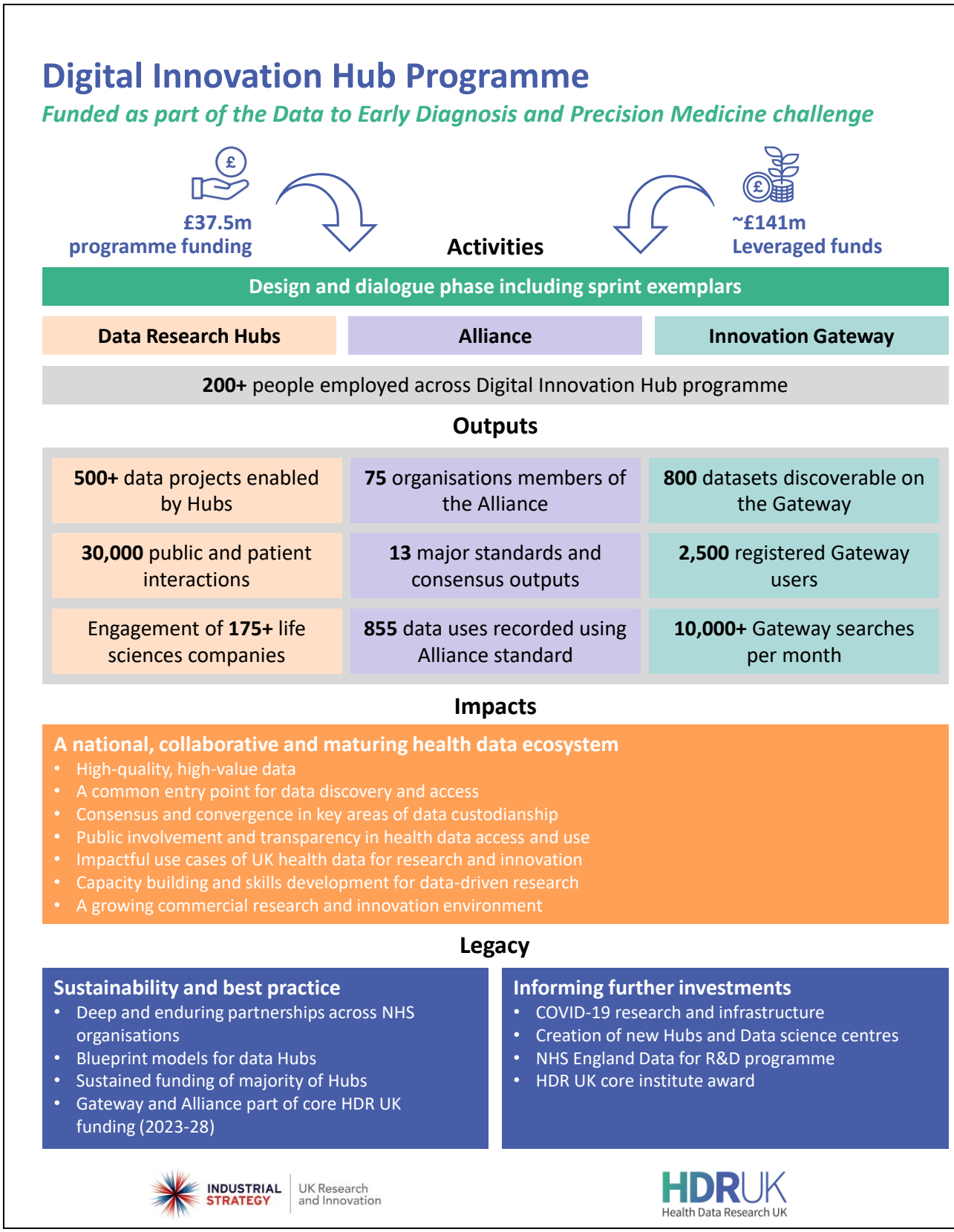



**Figure 1a**: Overview of the DIH Programme's funding, activities, outputs and impacts, and legacy.

# Overview of the DIH Programme

## Background and requirements gathering

In September 2018, UK Research and Innovation (UKRI) allocated £37.5 million from Innovate UK's (IUK) Industrial Strategy Challenge Fund (ISCF) to HDR UK to lead the health data component of the £210 million Data to early diagnosis and precision medicine challenge (D2EDPM). This was known as the Digital Innovation Hub (DIH) programme and its vision was to make the UK home to data-driven research, scientific advances, and innovation in healthcare to improve patient outcomes through enabling the development of infrastructure that supports health data research. Other programmes within the challenge focused on the application of Artificial Intelligence (AI) to diagnostic imaging and whole genome sequencing for precision medicine.

To inform development of the infrastructure and effective use of funding, HDR UK undertook a six-month Design and Dialogue Phase, starting in October 2018. This gathered feedback from over 2,700 individuals across industry, NHS, academia, government, and the public. Over 200 companies and 1,200 industry stakeholders engaged in the process to establish a broad view of the needs of users and the desired attributes of a world-leading health data research ecosystem.

The consultation confirmed existing challenges of navigating a complex and fragmented landscape particularly for industry users, and the difficulties in accessing high quality data. This highlighted the need for predictable routes for accessing data that could be completed swiftly. Ideally, users wanted a single front door to discover and request access to data, together with standardised terms on usage, and access to expertise and services to make it easier access and use the data.

Whilst broadly there was a wish for ecosystem-wide enhancement in the NHS health data and alignment in standards and processes, it was recognised that the programme would need to have focus and exemplars in specific diseases or types of data. Companies were interested in a range of different use cases, especially delivery of more efficient clinical trials and better access to, and use of, real-world data. From an initial plan to develop Hubs based exclusively on geographical populations of three to five million people, the Design and Dialogue Phase highlighted that disease-specific data curation was a priority for industry. It also highlighted that only a small sub-set of potential data that could be made available for research was discoverable. As a result, the programme focused on building a data discovery platform that covered a far wider and more diverse range of data assets than the 13 national data sets and five NIHR Health Informatics Collaborative datasets that were listed on the NIHR Health Data Finder at the time.

## Infrastructure design

The Design and Dialogue Phase provided the evidence for the DIH Programme prospectus which led to the creation of three new complementary and interlinked capabilities:

- The Health Data Research Hubs (Hubs): UK-based collaborations between NHS, industry, academia, and patient groups/medical research charities that focus on data curation and create the expertise, tools, knowledge, and ways of working to maximise the insights and innovations developed from health data

- The Health Data Research Innovation Gateway (the Gateway): a technology partnership between HDR UK and industry to develop a web-based common entry point to discover and request access to UK health datasets for research and innovation
- The UK Health Data Research Alliance (the Alliance): an independent, non-legally binding alliance of health data providers, custodians and curators who develop and share standards, policies, and best practice to demonstrate trustworthiness in health data use for research to improve human health.

These core components were informed by user need and are designed to work in synergy. The Alliance was designed to be inclusive and open to all data custodians, addressing users' needs for common practices and standards across the UK health data ecosystem. It was informed by the Global Alliance for Genomics and Health, which has driven practices for the responsible sharing of genomics data. A limited number of Hubs were selected as vanguards for health data and services in focused disease areas or clinical trials and real-world evidence.

Standards and recommendations developed through the Alliance have informed the design of technical infrastructure and tools delivered through the Gateway. The Hubs are early adopters of the Alliance outputs and Gateway technology and infrastructure and were able to provide feedback on their implementation. This relationship is set out in Figure 1.

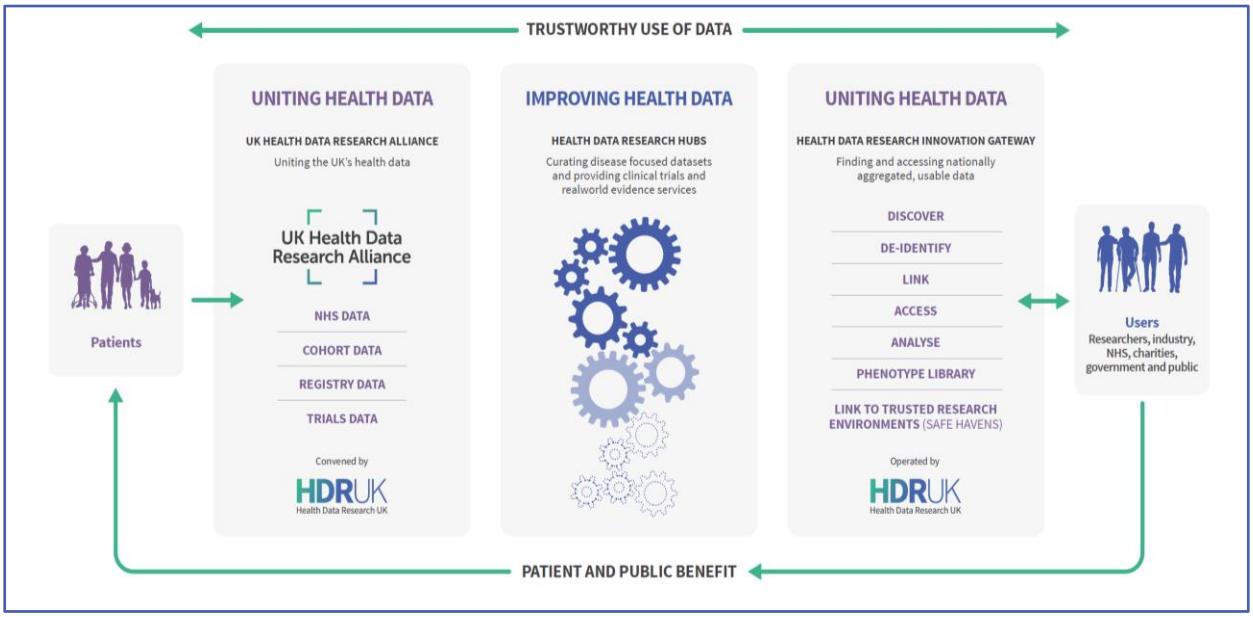



**Figure 1b**: The DIH Programme's major components and their relationships with stakeholders.

## Programme components and allocated funding

Under the guidance of the ISCF Programme Board, the funding was allocated across different components of the programme (Figure 2).

| Component | ISCF funding allocated |
|---|---|
| Health Data Research Hubs | £21.8m |
| Innovation Gateway | £8.1m |
| UK Health Data Research Alliance | £1.2m |
| Sprint exemplar projects | £2.7m |
| Central functions and programme Management | £3.7m |
| **TOTAL** | **£37.5m** |

**Figure 2**: Allocation of ISCF DIH Programme funding across constituent components (excludes all leveraged funds into the programme).

The Appendix to this report includes six impact case studies from across the programme components. Each programme component is also set out in more detail in its own report, including purpose, the approach, the outcomes, lessons learned and legacy.

## Sprint exemplars

As part of the Design and Dialogue Phase, HDR UK launched a call for sprint exemplar projects to develop proof-of-concepts for technology, methodology and research services that could inform the implementation of the DIH Programme infrastructure. The 10 projects that were funded provided early use cases for the programme's focus on research services and infrastructure across the NHS, academia and industry to enable high value linked datasets to be used for UK-scale research. Partnerships formed through the sprint exemplar programme formed the basis of two successfully funded Hubs: Gut Reaction and NHS DigiTrials. Insight sharing days also helped to build a community and keep stakeholders informed of developments.

Report 2 provides more detail on the purpose, the approach, the outcomes, lessons learned and legacy of the Design and Dialogue Phase, including the Sprint Exemplars.

## Health Data Research Hubs

A competitive call for Health Data Research Hubs was launched in May 2019, and the outcome was announced in October 2019. Of the 25 applications received, seven were selected for an award. Five Hubs (INSIGHT, Discover-NOW, Gut Reaction, DATA-CAN and BREATHE) were funded through the ISCF funding, with a further two (NHS DigiTrials and PIONEER) funded at a lower level from available HDR UK discretionary core funds.

Each Hub were and remain to be a consortium of organisations from multiple sectors (e.g. industry, academia, the NHS, charities), with a specific area of deep domain expertise, such as a disease or research theme.

Hubs provide services and access to data to support research and improve or curate the data to increase its utility and impact. They have patient and public involvement and engagement (PPIE) embedded within their structures, which provides robust governance whilst meeting industry needs.

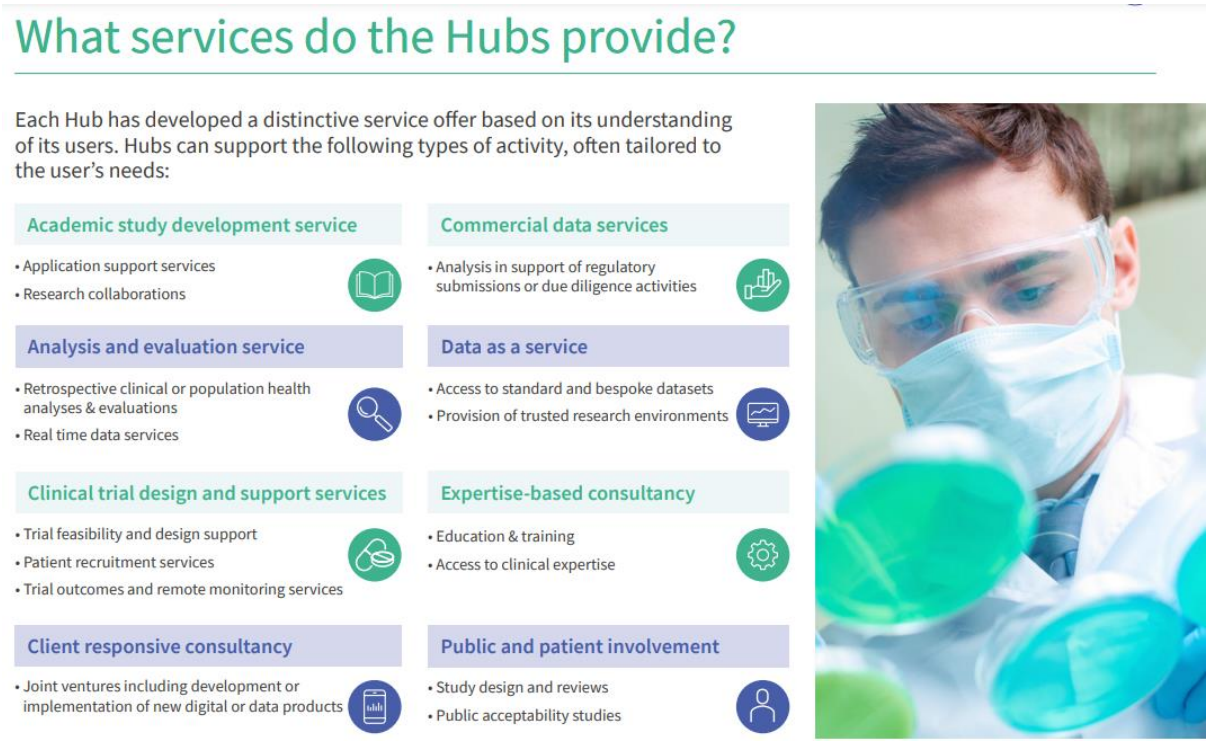



**Figure 3**: The range of services provided by Hubs.

Funding a tranche of Hubs at the same time created a cohort effect. Despite being hampered by the onset of the COVID-19 pandemic within weeks of them being established, the Hubs formed a collective and cohesive network which was greater than the sum of its parts. Facilitated by the central HDR UK programme office, the Hubs have led multi-sector, cross-cutting workstreams that involve development and sharing of best practice across PPIE, commercial sustainability and data standards and quality. The programme office also guided the Hubs through an innovative process of assessing progress against three predefined major milestones, which was undertaken by an independent panel.

The Hubs have benefited from having a focus on sustainability from the outset. All Hubs have continued to operate beyond the ISCF funding period with a variety of models for long-term legacy. NHS DigiTrials and Discover-NOW (London SDE) are directly involved in the NHS England Data for R&D programme, whilst PIONEER is collaborating closely with the West Midlands SDE in particular. DATA-CAN, INSIGHT and Gut Reaction have all changed host organisations and are pursuing mixed sources of funding including grants, research collaborations and service contracts with industry and other users. BREATHE, DATA-CAN and PIONEER are also supporting HDR UK's Research Driver Programmes from April 2023.

In 2021, the UKRI Medical Research Council (MRC) supported the award of two more Hubs, Alleviate (pain) and DATAMIND (mental health), as part of a growing Hub programme.

Further information about the Hub component of the programme can be found in the case studies included in the Appendix and Report 3, which also includes the Hub programme delivery approach.

## Health Data Research Innovation Gateway (The Gateway)

The Gateway was launched as a minimum viable product (MVP) in February 2020 through a partnership with IBM. Its aim is to provide a web-based user interface for researchers and innovators to discover and request access to UK health-related datasets. The MVP replaced, and significantly enhanced, the NIHR Health Data Finder, which, at the time of the switch-over provided a metadata and discovery service for 18 datasets and approximately 200 users per month. By contrast, the MVP included metadata from over 400 datasets across more than 20 custodians and saw over 2,000 users per month.

Development of the Gateway was expanded through a technology partnership with PA Consulting, following a competitive process that included three shortlisted potential partners undertaking an extended evaluation. As part of plans for sustainability, HDR UK brought the Gateway software development capability in-house at the end of 2021.

Through the course of the programme, the Gateway has developed tools, services, and products to serve the research community and promote FAIR (Findable, Accessible, Interoperable, Reusable) data science practices (Figure 5). In addition to a search functionality connected to an extensive metadata catalogue which contains both contextual and technical metadata elements, the Gateway provides a data use register, cohort discovery user interface, and Data Access Request application form and functionality based on the Five Safes Framework (safe data, projects, people, settings, outputs). It also connects to the HDR UK Phenotype library that contains over 1,000 definitions for hundreds of diseases from structured and unstructured data sources.

The level of interest and engagement in the Gateway has also helped to identify challenges that still need to be addressed. Navigating access to relevant data assets for specific health data research use cases is a complex process, particularly with regard to data that has been collected for a different purpose. Contextual knowledge about the data collection, intermediate processing and curation is required, alongside data profiles and counts, to understand its suitability and reliability. Developing this level of metadata can be resource intensive and time consuming and so often this can only be understood by interacting with data custodians directly. Therefore, data custodians' partnership with the Gateway is essential to deliver a data discovery functionality that meets data users' needs.

Integrating with custodian and Trusted Research Environment (TRE) providers' existing systems is also necessary to avoid duplication of effort, both for the sharing and updating of metadata as well as the processing of data access requests. Agreed standards and specifications to enable automated exchange of information via Application Programming Interfaces (APIs) is key and whilst some progress has been made, the Gateway is just at the start of this journey.

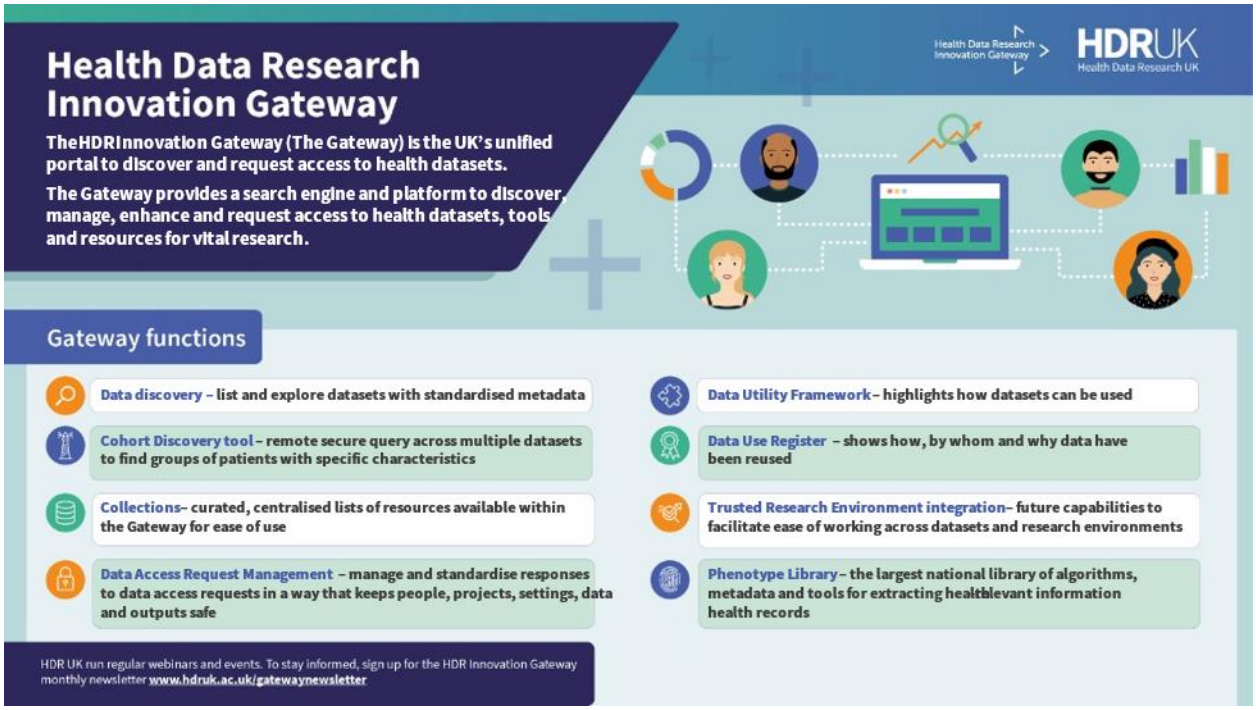



**Figure 4**: DIH publicity materials highlighting Gateway functionalities.

Whilst data discovery will remain a core function of the Gateway, the future vision is for the Gateway to act as a portal to all health research data that is shareable, without the need for the privacy-enhancing features of an SDE or TRE. This will be achieved through establishing a **vibrant open-source technology ecosystem community**, including researchers and data custodians, to ensure that its development and adoption is led by real world use cases. In the short term, following the rapid growth and development of the platform during the pandemic, some re-engineering of the Gateway is needed for scalability, and to meet emerging researcher needs. Baseline funding has been secured through HDR UK core funding, with the expectation of leveraged funding to support enhanced capabilities.

Further information about the Gateway component of the programme can be found in the case studies included as an Appendix, and Report 4, which also includes details on the different phases of developments, approach to development and major releases.

## UK Health Data Research Alliance (The Alliance)

The Alliance is a non-for-profit association of leading health data providers, custodians and curators. Alliance members are dedicated to improving human health by maximising the potential of multiple forms of data to accelerate progress in biomedicine, health and care through development and adoption of standards, policies, and best practice. Convened by HDR UK, the Alliance has 75 data custodian members (as at February 2023) from across the four nations and including all the main recipients of ISCF D2EDPM funding: health data research Hubs; five digital pathology, radiology, diagnostics and AI research centres; and UK BioBank and Genomics England. Members are committed to improving access to data for research that benefits the public across research cohorts, biobanks, medical research charities, NHS trusts, programmatic investments, and national agencies (see Figure 5, below).

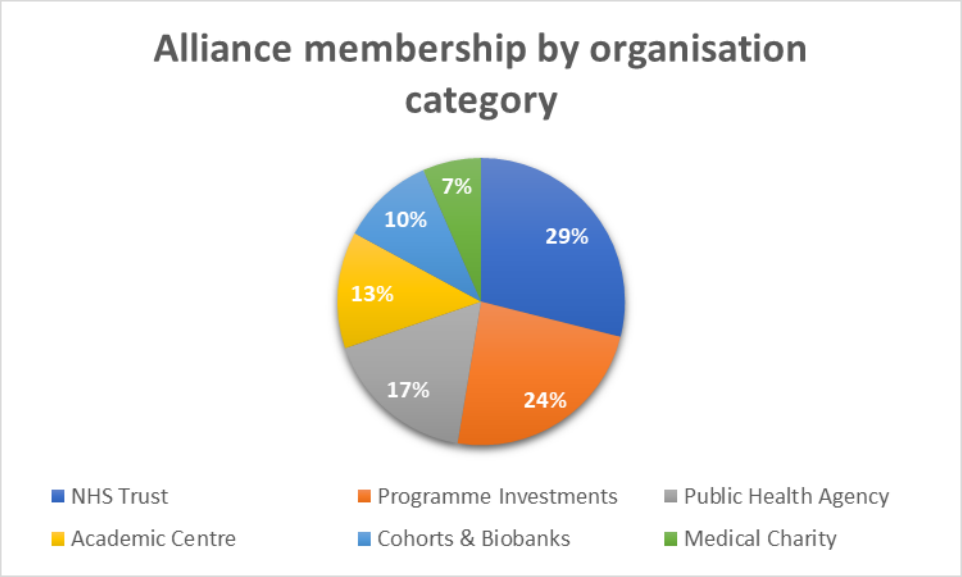



**Figure 5**: Alliance membership by organisation.

The Alliance is governed by a set of Principles for Participation developed in consultation with stakeholders, to which every DIH partner organisation has signed up. These principles provide a framework of collaborative and open working practices across the DIH Programme. They draw on national and international best practice frameworks and recommendations, including the FAIR Guiding Principles for scientific data management, and the Five Safes Framework. The principles are referenced in the Alliance Letter of Intent which all members sign prior to joining.

Alliance members collaborate to develop common standards, formats, and tools for stakeholders in the health data research community and then make them available through a reference implementation on the Gateway. The Alliance has published a series of papers modelled on initial open consultation (Green Papers) and subsequent White Papers on key work areas. The aim is to shape the development of a responsible and ethical infrastructure for UK health research and innovation through developing, sharing, and adopting best practice, taking a common approach and avoiding duplication of effort.

Bringing these parties together facilitates, for the first time, a federated and coordinated approach to health data research infrastructure. The Alliance addresses industry feedback, provided through the Design and Dialogue Phase, for more standardised approaches to data, governance, and access terms across the NHS. Achievements have included development of a consensus around embedding the Five Safes Framework in custodian application processes and a new data use register standard. Work on the Trust and Transparency elements of data use are being taken forward by the Alliance's Pan-UK Data Governance Steering Group.

Alliance baseline funding has been secured through HDR UK core funding, enabling it to continue to maximise trustworthy use of health data by working across the entire UK 'ecosystem' and connecting with relevant international developments. Priority areas include: improving the availability of, and access to, priority datasets for research and innovation; and improving the quality, completeness and diversity of data through development and implementation of data standards.

The Alliance will be strengthened further by bringing together national decision makers and regulatory bodies, industry associations and research funders who have not yet had a direct role in the Alliance, but who are critical to ensuring development of an efficient health data infrastructure ecosystem. Details around key changes proposed are publicly available via the Alliance website.


**Figure 6**: Promotional flyer about the UK Health Data Research Alliance.

Further information about the Alliance component of the programme can be found in the case studies included as an Appendix to this report, and Report 5, which also includes details on the purpose, approach, outcomes, lessons learned and legacy of the Alliance.

## Public and Patient Involvement and Engagement (PPIE)

From the programme's outset, HDR UK recognised that public trust would be integral to successful, high impact delivery – but that this trust would need to be earned continually. The first Principle for Participation, developed after a workshop on commercial principles in January 2020, is to *'Demonstrate active and ongoing engagement with patients and the public in the design, development and governance of their activities involving health data, to provide assurance that these activities are in the public interest'*. The DIH Programme has advanced innovative models of involving the public and patients in data-intensive research.

The Hubs have spearheaded this approach. Each Hub works directly with patients, relatives, carers and the public to bring their distinct voices to complex conversations about data and participation in research. PPIE members review, approve and can veto proposals or contracts. Hubs have established training programmes for patients and the public with an interest in patient data which builds their confidence to speak up and encourages better, more effective partnerships in making decisions, especially when it comes to commercial activities. The approach of DATA-CAN for example has been featured in the *British Medical Journal* (BMJ)'s Partnerships in Practice blog. Other examples include:

- The Open Data Institute (ODI) worked with eye health charity Action Against Age-related Macular Degeneration (AAAMD) to support the INSIGHT Health Data Research Hub to put patients and the public at the heart of data sharing.
- PIONEER's public-facing website has been praised by the HDR UK Public Advisory Board for its clarity and transparency.
- With the support of Crohn's & Colitis UK, Gut Reaction's Patient Advisory Committee was formed at inception of the Hub to facilitate patient involvement in planning and decision-making. Gut Reaction has also done a significant amount of work on developing training materials to support patient involvement in health data research.

Public representatives have been involved in each funding decision and every milestone review of the programme. The HDR UK Public Advisory Board has also provided a continuing source of input to both the Alliance and Gateway developments. Further information on PPIE is included in the programme outcomes and benefits section below and in a number of case studies included in the Appendix.

## Industry involvement and engagement

Central to the design of the DIH Programme, and to the delivery of its infrastructure, was close partnership and collaboration across multiple industry sectors. The predominant route of engagement for industry as users has been through the Hubs, but all elements of the programme have sought to engage and involve industry in different ways (see Figure 7).

| Component | Industry involvement | Main outcomes |
|---|---|---|
| Design and Dialogue | Ensure industry needs were captured in the programme design | 200 companies engaged (1,200 individuals) over six sectors. In-depth interviews with 32 industry leaders |
| Sprint exemplars | Industry partners included as part of project proposals | 13 industry partners involved across 10 selected sprint programmes |
| Hubs | Industry partners as core part of Hub consortia, provided resources and expertise | 22 different industry partners (25 partnerships) selected in successful consortium pledging cash and in-kind resources valued at £19 million |
| Hubs | Industry as users of Hub data and services | 175 projects contracted with commercial organisations (Milestone 3) |
| Gateway | Industry providers as core part of a Technology Partnership, providing resource and expertise | • IBM as technology partner for Gateway Minimum Viable Product<br>• PA Consulting as technology partner for main build<br>• MetaDataWorks and Parity as metadata onboarding partners<br>• BC Platforms as cohort discovery partner |

| Component | Industry involvement | Main outcomes |
|---|---|---|
| Gateway | Industry as users of Gateway to find and access datasets | 201 registered industry users (source: Quinquennial Review) |
| Gateway | Industry to list metadata on Gateway | Two industry created datasets listed on Gateway (in partnership with Hubs) |
| Alliance | Industry to have opportunities to inform and influence standards and Alliance outputs | Commercial TRE providers provided comments on TRE Green Paper |

**Figure 7**: Strategies of industry engagement in the DIH Programme and outcomes.

## Programme management

Integral to the impact of the DIH Programme has been the coordination and project management provided by the programme team hosted by HDR UK. Its purposes included (a) realising the complementary benefits of the Alliance, Gateway, and Hubs, (b) accelerating knowledge transfer and sharing of practice, (c) ensuring that delivery of the programme is guided by its Principles for Participation, (d) maximising beneficial stakeholder engagements, and (e) providing reassurance and monitoring of the investment of public funding, in particular using strategic milestones and independent review.

Programme management activities included:

- Coordination of the Design and Dialogue Phase, and the design, delivery, and management of the sprint exemplars programme (see Report 2 for more details)
- Design and delivery of the research data Hub competition
- Innovative milestone management of the Hubs programme, and support to individual Hubs (see Report 3), alongside regular performance management, metrics, and funder reporting
- Management of the cross-cutting Hub workstreams, (i) participation and sustainability (ii) data utility (iii) communications and PPIE, to drive impact and support the Hubs' work toward the three milestones of the programme
- Market research and user testing of products and services
- Marketing and communications of the DIH services and infrastructure and communicating the impact and public benefit of the programme.

**Building a legacy for the DIH Programme**

This flagship event took place in Birmingham on 28 June 2022. It celebrated the DIH Programme's successes and is helping to develop its legacy. It was a popular event both in-person and online, with over 700 people registering to attend from a range of sectors, including industry, the NHS, academia, and charities both nationally and internationally.

Keynote addresses were given by former BBC Technology correspondent Rory Cellan-Jones; Janet Valentine, who at the time was Challenge Director; Emily Jefferson, who at the time was Director of Health Informatics Centre at the University of Dundee; Sir Mark Walport, who had also launched the DIH prospectus in May 2019; and Claire Bloomfield, the SRO for NHS England's Data for R&D programme. The event showcased each component of the programme, both on the main stage and in the 'market place', with networking opportunities built into the programme. A recording of all the plenary sessions is available on YouTube.

Rory Cellen-Jones


Emily Jefferson


Sir Mark Walport

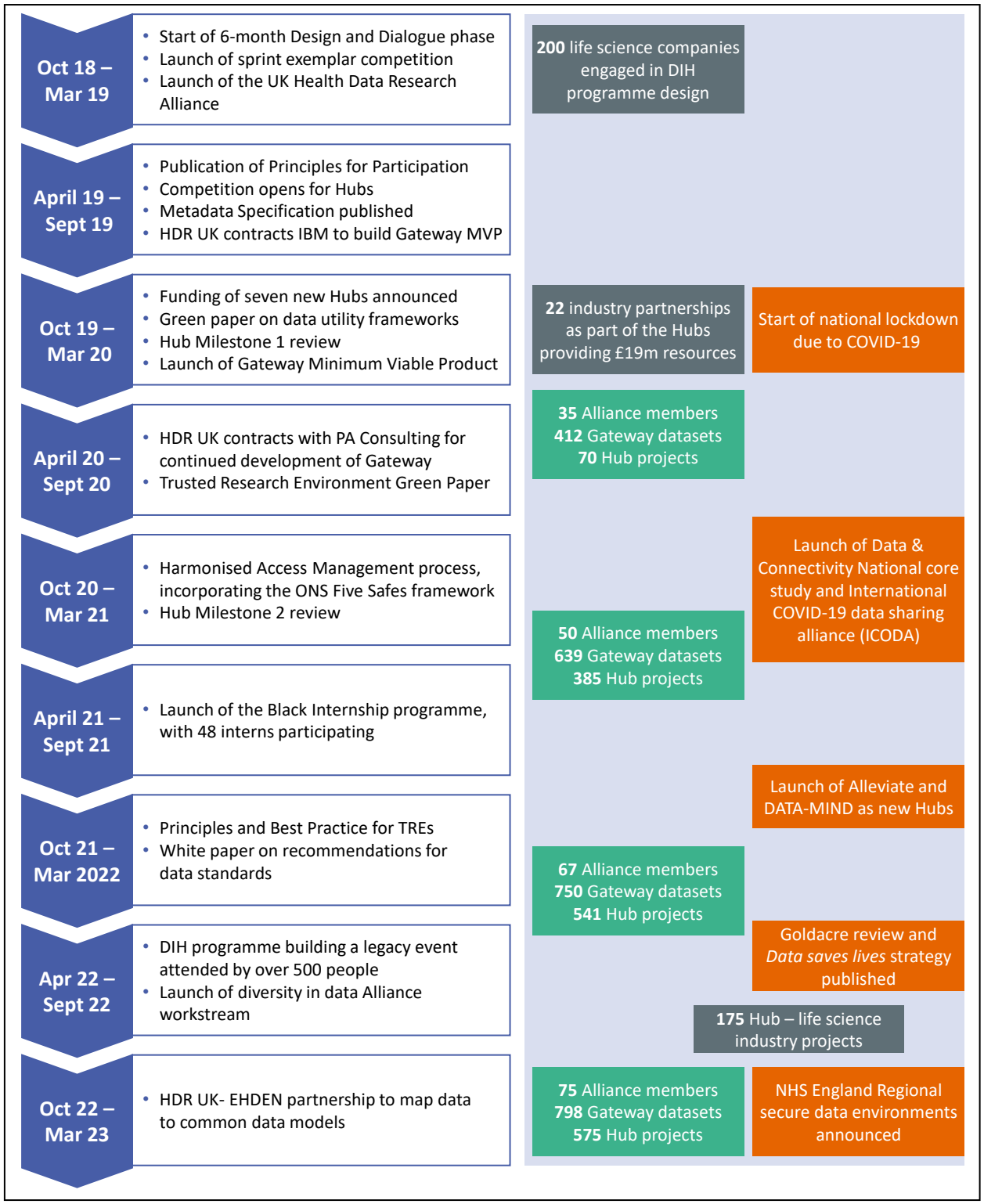



**Figure 8**: A summary timeline of how DIH Programme activities unfolded.

# Programme outcomes and benefits

The DIH Programme has addressed the needs of industry and academic users by increasing the volume of curated health-relevant data, which is now readily discoverable, and which has established, transparent governance so that it can be efficiently accessed for approved research purposes. The programme has supported progress towards a more collaborative and maturing 'many-to-many' health data research ecosystem. Through some early exemplars, the DIH Programme has explored: federated approaches, allowing users to work between datasets held by different custodians; technologies that support secure and safe use of data; and increasing and automated levels of transparency to the public.

The benefits of the DIH Programme have been synthesised under seven themes, which are aligned to the original purpose of the DIH Programme and overarching ambitions of the industrial strategy:

- Curation of high-quality, research-ready data
- Making data discoverable and accessible at scale
- Consensus and convergence in key areas of data custodianship
- Enhancing public involvement and transparency in health data access and use
- Impactful use of UK health data for research and innovation, including contribution to the national and international COVID-19 response
- Building capacity and developing skills for data-driven research
- Enabling a supportive commercial research and innovation environment
- Leveraging £140 million additional funding.

The programme has supported many exemplars of advances in each of these areas. Some of these are discussed in the rest of this section, with supporting case studies in the Appendix.

## Curation of high-quality, research-ready data

At the start of the DIH Programme, few datasets met the user needs articulated during the Design and Dialogue Phase, which were for datasets that are longitudinal, event-based, linked (such as primary and secondary care, genome, and clinical measures), and large-scale (e.g., UK wide) to enable identification of disease pathways or sub-groups.

The seven data Hubs curated over 200 datasets, making them discoverable via the Gateway and available to users for research purposes in a safe and secure way. Data made accessible by the Hubs spans real-world NHS data, including: health records, test results and high-quality imaging; patient registries and clinical cohorts; and research studies including observational cohort studies and clinical trial data (see Report 3 for further information, including outputs). The Data Standards and Quality workstream of the Alliance led to the development of a first-of-its-kind data utility framework that enables objective measurements of improvement in data quality and utility and can help demonstrate a return on investment for curation activities. It was trialled with the Hubs between the three-month and 18-month milestones, during which time it demonstrated an improvement in 70% (51 out of 73) of datasets that had been made discoverable over that period.

The Hubs have undertaken many activities to make data more 'research-ready', i.e. (i) accessible; (ii) broad – so as to be applicable to multiple uses; (iii) curated; (iv) documented; and (v) enhanced for research purposes (*McGrath-Lone, 2022*). The Hubs have also offered services that support users to

navigate processes to access, analyse and derive insights from data. Figure 9 summarises how the activities and infrastructures that Hubs have collectively undertaken have developed research-ready data.

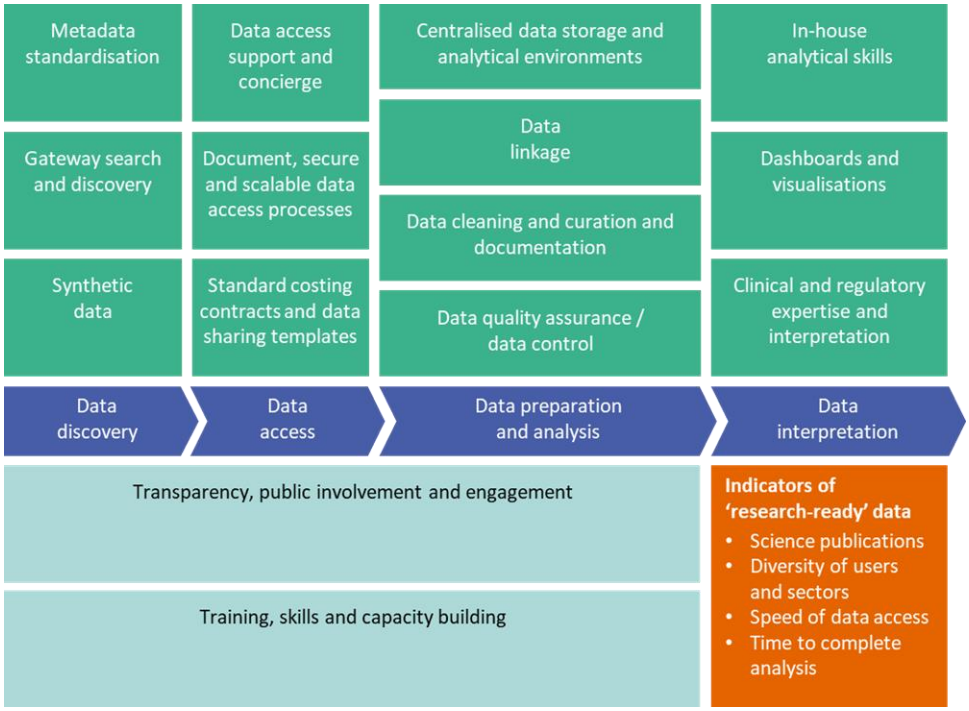



**Figure 9**: Hub activities (green) that have contributed to the development of research-ready data. Some indicators of research-ready data are shown in the orange box.

## Making data discoverable and accessible at scale

The Gateway is the data discovery and access platform for the DIH Programme, and the Hubs and most other Alliance members have used the tools and services it provides. The Gateway has grown to hold metadata across 780 datasets (as of November 2022) from 68 different organisations, which allows a user to search and understand requirements for accessing a dataset. The content of the Gateway was accelerated through requirements built into the Hub development and the delivery partners and researchers of the COVID-19 Data & Connectivity NCS.

Working with Alliance members and researchers, HDR UK categorised the data access request questions that were common across data custodians using the Five Safes framework. This resulted in a common framework to which data custodians align their data access request applications, which has been operationalised through the Gateway in a Five Safes Data Access Request (DAR) form. When fully implemented, the potential benefits are that users will be able to request data from multiple studies via a single application, and there will be greater standardisation, transparency and clarity over the information required to assess and review a data access request. This is in line with one of the recommendations of the Goldacre Review that was published during the course of the programme

The Gateway has also provided the platform for the development of more innovative features, such as enhanced models of discovery and integration with existing data custodian or TRE provider systems. These projects, which are still in development and are being tested with pilot users, include:

- Data profiling, which enables the user static counts of the datasets to understand the profile of characteristics such as age, gender, and ethnicity
- Cohort discovery, which supports dynamic querying of datasets to support cohort building through identifying participants
- Federated metadata catalogues, which allow custodians to integrate their existing catalogue with the Gateway through APIs
- Custodian dashboards, which provide a visual overview of all user touchpoints and interactions with datasets onboarded by a custodian.

By default, the DIH Programme uses, and extensively contributes to, open-source software, and has helped inform HDR UK's wider Development Principles. Source code for the Gateway, and APIs for Gateway integration are available on GitHub, as part of a broader collection of >150 repositories of open standards, data and source code, creating an accessible resource for health data researchers.

## Consensus and convergence in key areas of trustworthy data use

The Alliance has supported the development and adoption of standards that are essential for a collaborative, harmonised and trustworthy approach to health data research across the UK (see Figure 10).

| Date | Output |
|---|---|
| May 2019 | Principles for Participation |
| June 2020 | Principles for Data standards |
| July 2020 | Trusted Research Environments (TRE) Green Paper |
| November 2019 | Metadata Specification and GitHub Schema |
| December 2020 | Harmonised Access Management process |
| March 2021 | Green paper on data utility frameworks Gordon et al, 2021; Gordon et al, 2022 |
| July 2021 | Green paper on data use registers |
| November 2021 | White Paper - Recommendations for Data Standards in Health Data Research |
| December 2021 | Building Trusted Research Environments - Principles and Best Practices (TREs) |
| January 2022 | White Paper on data use registers (Karrar et al 2021) |
| February 2022 | Five Safe Data Access Request application |

**Figure 10**: Major consensus outputs arising from the Alliance.

Data access policy and practice in the UK and beyond have been shaped by a critical policy paper which has set out principles and guidance for building TREs (known also as Secure Data Environments; SDEs). Co-authored by HDR UK, the paper has been downloaded over 6,100 times since it was published (as of February 2023). Case Study 9 in the Appendix explains how it is shaping UK and European policy on safe data use.

Establishing common standards for healthcare data and its associated metadata is a fundamental requirement to enable health data research at scale to improve people's lives. HDR UK's Principles for Data Standards (June 2020) set out an approach to adopting data standards for health data research concerning structured electronic health records data. A data officers' group, bringing together those responsible for data quality across the ecosystem, has convened regularly. This has enabled the Alliance to build on these principles, propose recommendations for data standards and promote their adoption.

## Enhancing PPIE in health data access and increasing transparency of use

PPIE was layered throughout the DIH Programme and was included in all Hub milestones. A regular PPIE sub-group of the Alliance shared experiences and best practice, and PPIE was built into all key consultations, such as the creation of the data user register standard and the positioning taken on TREs.

The major achievement from the DIH Programme has been the development of processes to enable greater access to health data, including for commercial use, which has been scrutinised and co-developed with public contributors. Wide ranging activities, involving smaller groups from 8 to 12 people to larger public deliberations involving over 100, have supported each Hub to discuss, advise, review, debate and help inform positions on arising topics.

A major step forward in transparency has been the development and delivery of a common approach to data use registers through both community development of a standard and via a reference implementation on the Gateway, which has worked across the Alliance, Gateway, and Hubs (see Appendix, Case Study 10).

## Impactful use of UK health data for industry-enabled innovation

The enabling infrastructure established by the DIH Programme has supported hundreds of users to engage with UK health data to undertake research. The Hubs have contracted 175 projects to support industry-enabled innovation in drug-discovery, precision medicine, diagnostics, AI and care delivery.

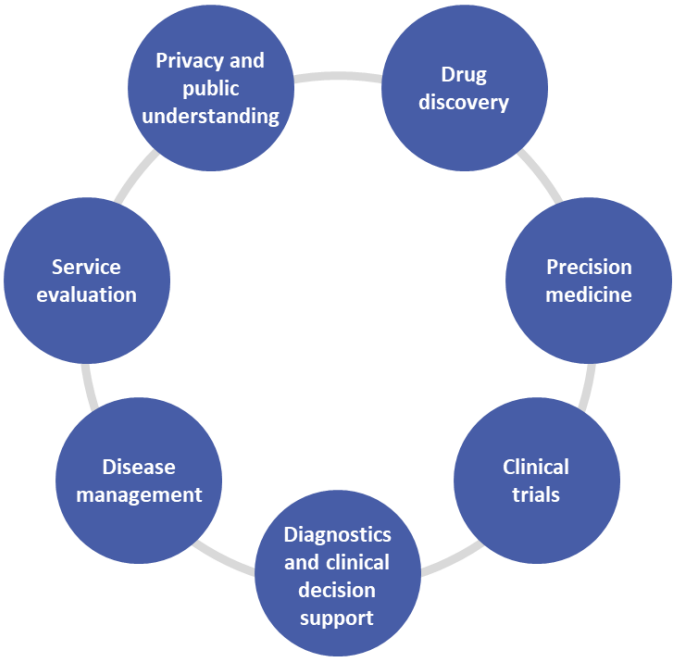
Clients include global pharmaceutical firms and large health tech companies, through to start-ups and SMEs. Some companies have developed a portfolio of projects with the Hubs, with many examples of repeat customers. The type of services is varied. Not all contracts result in a data-sharing agreement, as Hubs also provide feasibility and scoping services, PPIE activities, and expertise through consultancy services.

The full range of projects is captured in the data registers, publications and project summaries of the individual Hubs, as well as in the Milestone 2 report. Occasionally, details of projects have been redacted from publication due to commercial sensitivity and this was a particularly contentious area whilst developing the data use register standard.

The range of partnerships and projects are summarised in Figure 11, whilst the AstraZeneca partnership with Discover-NOW to transform care for diabetes, heart and kidney disease is included as Case study 1 in the Appendix.

**Gut Reaction** and **Privitar** collaborated to bring together and integrate different sources of data collected from IBD patients whilst maintaining highest levels in security and data privacy.

**DATACAN** were commissioned by **FlatIron Health** to work with patients to advise on how the company could build public understanding and trust into its business activity.




**INSIGHT** and **Roche** collaborated in developing a data-driven understanding of the impact of COVID-19 pandemic on treatment for wet AMD.

**Discover-NOW** and **AstraZeneca** implemented a risk stratification model to identify patients at highest risk of type 2 diabetes complications and designed a new integrated pathway.

**BREATHE** and **PrecisionLife** collaborated to identify multiple promising drug targets for the development of personalised therapies for non-allergic asthma patients.

**Gut Reaction** and a **major pharma** collaborated to investigate genetic variations in patients with Inflammatory Bowel Disease to look for differences in symptoms and responses to treatment.

**DATACAN** and **Roche Products Ltd** worked to understand early triple negative breast cancer, treatment options and health outcomes for patients, to better support drug development and access to treatments.

**DigiTrials** supported recruitment to the **Grail LLC** funded Galleri trial recruiting 140,000 participations in 10 months.

**PIONEER** and **Microsoft** collaborated on the development of AI to support clinicians interpreting chest radiographs as a way to diagnose COVID-19 and predict the likely outcomes for patients.

**Discover-NOW** and **OWise** evaluated a mobile support tool for patients with breast cancer and to understand if users interacted with the health system differently.

**Figure 11**: Examples of Hub and industry partnerships.

## Contribution to the national COVID-19 response

The DIH Programme made a significant contribution to health data research in response to the COVID-19 pandemic. The pandemic both pivoted and accelerated development of the capabilities of the Gateway, Alliance, and Hubs so that they provided infrastructure to quickly answer the most pressing research questions and fast-track them into policy in a way that would have otherwise taken far longer. HDR UK, in partnership with Office for National Statistics (ONS), led the Data & Connectivity National Core Study (NCS), one of the six NCSs commissioned by UK Chief Scientific Advisor Sir Patrick Vallance to lead the UK research response to the pandemic.

The Gateway was critical to the study because it provided the UK research community with a single platform to access information about pandemic-relevant datasets and a way to request access to them, as demonstrated through the partnership between the Gateway and BREATHE Hub. This enabled data being collected by the Kings College London/ ZOE COVID-19 Symptom Study Application from its 4.5 million public contributors to be onboarded to the SAIL Databank (Secure Anonymised Information Linkage Databank) and made discoverable via the Gateway for research and analysis within days of the App's launch. This provided a single route for researchers working at a national, regional and local level to analyse the data securely and use it to influence government policy, particularly within Wales and Scotland, with minimal delay.

Through the Data & Connectivity NCS, 110 datasets are available via the Gateway together with transparent information on the 271 research teams using these datasets for research via the Gateway Data Use Register. In many cases this includes the outputs of their research, therefore providing an exemplar of how to complete the loop from requests for data through to outputs and impact.

Further information can be found in the Appendix, Case Study 9.

## Building capacity and developing skills for data-driven research

Whilst the programme was not a training award, the DIH Programme has built a community of health data professionals and demonstrated innovative data science to operationalise FAIR data principles, with the potential to scale in the future.

The DIH Programme nurtured national capacity in the different domains of health data science, and in building research infrastructure, data governance and management. The core DIH award was used to pay all, or part of, the salaries of over 200 members of staff, either based at HDR UK (25 FTE supporting the Gateway team, Alliance, and central programme functions) and a headcount of approximately 180 across the five ISCF-funded Hubs.

Analysis of Hub finance returns showed each Hub claimed for between 30 and 50 posts at steady state, many part-funded roles, reflecting the critical mass required to deliver a data infrastructure. Analysis of job titles provides some insight to the multidisciplinary teams that are required to deliver Hub-like entities. Across job families there was a high proportion of technical, enabling and service-orientated roles created through the Hub programme, as opposed to more traditional research roles.

## Enabling a supportive commercial research and innovation environment

The data and services requirements of different industries are not the same. In the experience of the Hubs, small-to-medium-sized enterprises (SMEs) often have deep knowledge in their technology domain area, or in general business acumen. However, they often have limited understanding of the clinical research domain, the governance of data (particularly NHS patient data) and contracting. Hubs have tailored their offer to better meet the needs of SMEs. The expertise brought by the Hub runs throughout the entire pathway of working with data, from first enquiry and designing a study, through to supporting the governance and data access request process, conducting analyses, and interpreting results within the clinical context.

The Hubs have supported SMEs to work with UK health data for the first time, growing industry capacity and capabilities. The business development and service to SMEs includes:

- Helping to refine the companies' research questions
- Extending understanding of NHS Data controller legal duties and responsibilities
- Providing clinical research advice and expertise
- Visualising deployment in the real world
- Assessing data suitability and funding opportunities
- Being responsive to emerging trends.

INSIGHT has encouraged applications from multi-sector research users and has a particular interest in enabling the success of SMEs by meeting their need for efficient and safe access to high quality eye health data (see Appendix, Case Study 6).

BREATHE has worked with multiple commercial companies to better understand customer needs and to build a collaboration committed to its vision of driving the use of health data for research and innovation to transform respiratory health - see Appendix, Case Study 7.

## Leveraging additional funding

It is estimated that the original £37.5 million (£36.9 million actual spend) from the ISCF to deliver the DIH Programme has so far leveraged approximately £140 million of further investment.

The ISCF award supported five Hubs, but due to funding from other sources a further four Hubs have been created. HDR UK allocated Director's discretionary funding to set up the PIONEER and DigiTrials Hubs, as well as supporting an additional sprint exemplar. Further support was also provided through the UKRI World Class Laboratories Fund. UKRI Medical Research Council funded the creation of the DATAMIND and Alleviate Hubs (with Versus Arthritis), building on the Hub model developed through the DIH Programme. These awards totalled £7.4 million.

As part of each Hub competition, bidders were requested to match (or better) requested ISCF funding with in-kind or actual matched funding from host organisations and partners. Cash and in-kind commitments across all Hubs totalled £41.1 million, which included £19.1 million of industry funding committed from 25 companies. The reported spend was £32.4 million.

The Alliance and Gateway also provided a scaffold upon which further infrastructure and research programmes could be leveraged. These included the UKRI World Class Laboratories Fund programme (excluding Hubs), the COVID-19 Data & Connectivity NCS, the International COVID-19

Data Alliance (ICODA) and CO-CONNECT (COVID - Curated and open analysis and research platform). The proportion of these awards attributed to supporting and expanding the activities of the DIH infrastructure was £10.2 million.

The Hubs have also generated funding that have supported and expanded their activities. Contract and grant income from delivery of projects and data access commissioned from the Hubs is estimated to be £15 million. Whilst the value of individual contracts was not disclosed, an aggregate analysis of 100 projects (out of the 575 completed by the Hubs) delivered an income of £5.9 million. Therefore, a reasonable estimate would be in the range of £10 million to £30 million, with £15 million being a conservative assumption.

The Hubs have also been awarded research and funding bids that will either wholly or partially support the Hubs' activities. Analysis of ResearchFish returns and financial statements prepared by the Hubs for Milestones 2 and 3, shows that there has been a further £66.6 million of leveraged income from the Hubs.

Follow-on and leveraged funding from the sprint programme amounted to £9.1 million (not including the Hub funding which resulted from some sprint programmes). This is summarised in Figure 12, below:

<table>
<tr><th></th><th>Hubs</th><th>Sprints</th><th>Alliance and Gateway</th><th>Programme Support</th><th>Total</th></tr>
<tr><td>ISCF funds</td><td>£21.25m</td><td>£2.6m</td><td>£9.3m</td><td>£3.75m</td><td>£36.9m</td></tr>
<tr><td rowspan="3">All co-investment</td><td>£121.4m</td><td>£9.4m</td><td colspan="2" rowspan="3">£10.2m<br>(Additional programme investments)</td><td rowspan="3">£141m</td></tr>
<tr><td colspan="2">Made up of:</td></tr>
<tr><td>Additional Hubs £7.4m<br>Matched or in-kind £32.4m<br>Contracts £15m<br>Awards and funding £66.6m</td><td>Additional sprint £300,000<br>Awards and funding £9.1m</td></tr>
</table>

**Figure 12**: DIH Programme summary expenditure and totals of co-investment.

# Programme reflections and learning

Through its delivery, management, and coordination, the DIH Programme has evolved in response to its partner and user feedback, as well as a global pandemic.

Feedback has been gathered through formal milestone reviews with independent panels, stakeholder interviews and questionnaires, Alliance Board meetings and consultations, and informal feedback gathered during interactions across the programme. The rich insights gained through this process will guide HDR UK in its future activities, as well as providing guidance for funders of, and organisations which deliver, data infrastructure and services projects. This section summarises the programme-wide reflections and learnings, whilst the associated reports provide richer, component-specific insights.

## Ingredients of effective data and services infrastructure

The purpose of a Health Data Research Hub is to support users, particularly industry users, in accessing high-value health data and working with it in a secure way which is transparent and trustworthy to the public. The Hubs have been empowered through the programme to be innovative in delivering this goal, and the early reflections of the Hubs' leadership on what is needed for success was captured during the Summer of 2020 (and updated during Winter 2022) into a Hub playbook that documents Hub structure, staffing, activities and advice for the set-up of further Hubs. Some common core 'ingredients' of the Hub model are discussed in Report 3:

- Close engagement with data controllers of high-value data assets that enable the Hub to process and make data available
- Data curation, information governance and technical expertise to generate 'research-ready' data assets and secure research environments for conducting analyses
- Public, patients and patient organisations to ensure that all activities of the Hub are aligned to delivered public and patient benefit, and that models of data sharing are transparent and trustworthy
- Clinical expertise to provide context and interpretation of datasets and associated analysis, understanding of clinical and patient need, and adoption of products and tools into NHS practice
- Supportive (industry) 'customers' who are willing to invest, pilot and learn with the Hub to inform early iterations of data and services offer
- Business development and acumen, legal and commercial services, and project management to provide potential clients with a reliable and high-quality service offering
- Communications and marketing to disseminate service offer and impact, and build a community of stakeholders aligned to the mission of each Hub.

However, Hubs cannot deliver this in isolation and have benefited from organisational connections and collaborations across the 'ecosystem'. These links have been built at a local level (to build awareness within the host organisation), regional level (for example regional business and entrepreneurship networks, Innovate UK Knowledge Transfer Networks (KTN)) and national level (for advocacy, funding opportunities and elevation of Hub activities to national exemplars). The community of practice, through the network of Hubs, has provided the informal sharing of

knowledge, insight, and resources, in addition to providing each Hub with assurance in the approach taken to common challenges.

## Building communities

The DIH Programme has depended on building, nurturing, and convening multiple groups of stakeholders and communities, and judgement has been required throughout to balance their priorities. In particular, the relative activity and resources invested by the programme in understanding and responding to the requirements of data custodians to users of data, especially in the development of Gateway technology. Industry, academic, and NHS users also have different expectations and motivations for using data infrastructure. Generally, industry users want a higher level of supported service, and through the programme there has been a developing understanding of how and where to tailor approaches. A further dimension was ensuring diversity in the stakeholders that HDR UK worked with in shaping the programme.

The programme has shown the benefits of building the Alliance data custodian community. Its members value being part of something larger than their own organisation, and for individuals it provides a way to recognise and promote what is seen as good practice. The UK data ecosystem is fast moving. Those working in it are challenged to keep on top of developments in policy, legislation, and research initiatives. Ways of working and guidance from the Alliance is valued. It is robustly developed, reliable and provides confidence to senior stakeholders that governance processes are legal, trustworthy, and best practice. Closer working of the Alliance with regulatory agencies or funders could lead to more formal adoption of Alliance outputs. Whilst convergence and adoption is crucial for an efficient 'ecosystem', some organisations will always want to do things their own way, which remains important for innovation.

## Programme management and ways of working with the funder

Through the DIH Programme, HDR UK has played a dual role of representing the funder through delivering programme management and the milestone process across the Hubs, as well as collaborating with Hubs to support delivery of their objectives. These roles sometimes conflict, for example in openly sharing a Hub's challenges or difficulties. Overall, this was managed through maintaining clear governance and communication in the milestone process, whereby recommendations of the diverse and independent panel were ratified by the ISCF Programme Board, representing the funders. This enabled HDR UK to play the role of supporter and collaborator, and this learning has been carried forward to consider how HDR UK works with its parent funders (often UKRI) in the delivery and management of research programmes.

## Industry engagement

A challenge through the delivery of the DIH Programme has been a need for a clear and simple call to action for industry and for-profit organisations to engage. There was a model of engagement for industry partners with the Hubs, either as founding partners providing funding, expertise, or resources to the Hub's development, or as clients of Hubs contracting for data access and services. Individual Hubs built strong one-to-one relationships with companies, although some industry users expressed the desire to have a way to engage at a cross-programme national level.

Despite the best endeavours of the Hubs, many UK health data sets remain unavailable to commercial companies, or custodians are reluctant to engage with commercial users. Looking

forward, strategies will be needed for (a) advising funders and collectors of health data on collecting consent and governance for future commercial use; (b) easily findable metadata on commercial permissions; and (c) more guidance, and adoption of standards and common approaches for industry use.

Future investments could consider the value for introductory services that can match data users with potential data providers that are likely to be able to service data needed, like the capabilities provided through Innovate UK's KTNs. The planned Data Navigation Service as part of NHS England's Data for R&D programme may provide a similar capability.

## Areas for development

The DIH Programme has shown that creating research-ready data assets that include the use of data linkages has a high administrative and governance burden. The process can take years, with similar or duplicative processes and approvals required at different points in the system. As exemplified by the experience of the Gut Reaction Hub (NIHR BioResource), linkages to routine NHS records (primary, secondary care, high-cost drugs) for consented studies remains a significant operational challenge, and a burden for study teams. In their experience, they were unable to 'link once (to a nationally held NHS dataset) and use many', instead requiring permissions for linkage on a project-by-project basis. Furthermore, gathering data from individual NHS Trusts involved navigating the complexity of different data collection methods and systems, and the need to agree data sharing arrangements with each Trust's governance teams (who often had different interpretations of what could be agreed).

Other reported challenges faced by Hubs which may benefit from general, scalable solutions were the routine extraction of digital images for linkage, and the combining of datasets across national boundaries where standardised, low-resource approaches to harmonisation would be beneficial. Licences and permissions for data are commonly granted at an institutional or organisational level, which can be challenging to negotiate for multi-sector, multi-university working. These represent future data science challenges which, if addressed, could further unlock the potential of UK health data.

The DIH Programme and national pandemic response has created a wealth of new, linked data assets which include valuable data such as primary care, and prescribing. These data often rest under the custodian or banner of a disease-specific initiative, but include whole-population data and could be applied to many different use cases and diseases. There are opportunities to make these data available to new communities of researchers, whilst recognising the need for disease-specific expertise to maximise the insights and interpretation that can be drawn from the data.

The uptake and use of the Gateway was accelerated when there was a scientifically-driven use case in the pandemic, namely, to enable the rapid discovery, sharing and access of COVID-19 data. This mirrors wider learning from other HDR UK delivered programmes, that 'build it and they will come' for infrastructure is not enough. The strategy of developing infrastructure with a high impact and specific first use identified can guide decision making and demonstrate early success. In addition, having funding calls to the community to use specific infrastructure, as was the case for the COVID-19 Data & Connectivity NCS, can support user-centred development and promote awareness and adoption of the infrastructure. Funders should consider specific calls aligned to the use of new infrastructure, to enable providers to 'test and learn' and also overcome barriers to researchers

adopting new ways of working and new relationships. The requirement in the Hub competition to partner with industry was a key ingredient of their success, since industry partners brought real-world, early use applications that guided and accelerated design and delivery decisions.

The funding profile of the DIH Programme, in which all components were required to establish other funding sources beyond the first four years, has been a challenge for some elements of the programme. Whilst many components of the programme are expected to be sustained through different models in the immediate future, funding 'cliff edges' and uncertainty inevitably leads to a loss of staff and their expertise.

It remains easier to secure funding for new developments than it is to obtain longer term funding to sustain existing platforms and data as infrastructure. This is particularly true for infrastructure that enables the re-use of data and/or aggregation of smaller scale or more diverse data assets, whereas large scale data collections such as UK BioBank, Our Future Health, and Genomics England are viewed as long term commitments. Much of this infrastructure is, or will become, funded by multiple organisations, and there is a strong case for funders to pool their resources and commission these infrastructure and services together. For UK-wide infrastructure and services, a four-nations approach is essential, so devolved administration health research funders need to be core partners.

## Programme legacy and conclusions

One of the Programme's major legacies has been to exemplify at scale the models by which industry users can find, access and use UK health data in a way that is deemed secure and appropriate under public scrutiny. Multiple case studies are provided from the programme, which highlight these successes that have been achieved from the ISCF funding.

The DIH Programme represents a defined, four-year piece of funding with specific goals, but was delivered against the backdrop of a broader, longer-term, multi-decade strategic ambition for wider, safer and better use of UK (NHS) health data for public benefit. The programme has supported advancement of this agenda.

As the DIH Programme funding ends, HDR UK and NHS England have announced a joint workplan to continue working together on many of the areas started in the DIH Programme which are being continued through the NHS England Data for R&D programme. This includes activities on information governance, commercial principles for data access industry partners, and models of public, patient, and professional involvement, and engagement. The scope of this partnership includes continuing to work on a joined-up data ecosystem in which the Hubs, including those that hold research datasets, rather than those derived from real-world patient interactions, are 'docked in'. This will create a more streamlined and easy-to-navigate data access offer, across all four nations of the UK.

Hubs have set out their own approaches to sustainability following the Milestone 3 assessments and some are forming an important part of the HDR UK research Driver Programmes. The Hubs also have opportunities to align to the SDEs for R&D investment across England, which from 2022 to 2025, will aim to create access to curated NHS data, giving regional coverage of three to five million of the population. Discover-NOW and PIONEER, given their regional footprint, are already informing development of the London and West Midlands SDEs in this new data ecosystem. For NHS DigiTrials,

the 18 months of HDR UK Hub funding has served as a springboard for further investment. Through demonstrating its value as part of the DIH Programme, DigiTrials has received multi-year funding from NHS England through the Data for R&D investment to grow and develop its services on a more ambitious scale and support delivery of industry-backed trials to help transform large-scale clinical trials in England (see the Appendix, Case Study 3).

The foundations of PIONEER, BREATHE and DATA-CAN will be furthered as part of HDR UK Core Funds with involvement across three of its six core-funded Driver Programmes. The Gut Reaction Hub model could potentially be extended across NIHR BioResource, which is now its host organisation. The DATA-CAN Hub has changed host organisation at the end of their ISCF funding to ensure that they are positioned to move forward whilst continuing to deliver on their current data contracts and commitments. The INSIGHT Hub transitioned in September 2022 to a self-sustaining NHS partnership between Moorfields Eye Hospital and University Hospitals Birmingham, with Moorfields becoming the lead organisation leveraging the Trusts' reputation as a world-leading eye health service provider, and research and education centre.

# Appendix: DIH Programme Impact case studies

## Case Study 1: Discover-NOW (Improving Care) - AstraZeneca partnership transforms care for diabetes, heart and kidney disease

**Summary**

**AstraZeneca's partnership with the Discover-NOW Hub for real-world evidence has enabled research across healthcare systems and industry. The work has provided deeper insights into how care is delivered, which key interventions work and how to improve personalised care at scale. Through Discover-NOW, the industry researchers have accessed analytics and expertise and validated new treatments and interventions by securely accessing 2.5 million de-identified patient records in north-west London. With ongoing commitment from AstraZeneca to the collaboration, tangible achievements to date include: better remote care for type 2 diabetes patients; remote monitoring support for NHSX during the COVID-19 pandemic; redesign of an NHS Trust's heart failure pathway; and earlier identification and management of chronic kidney disease.**

**Case study**

Discover-NOW is a real-world evidence Hub that provides secure access to patient data to test health care innovations and treatments and prevent disease. Through access to 2.5 million de-identified patient records in north-west London, the Hub combines robust data infrastructure and access through its Trusted Research Environment, with the analytical capability and technical expertise to offer valuable services to clinicians, researchers and scientists.

From the outset, the Hub has worked closely with AstraZeneca, with the aim of working across healthcare systems and industry to discover how delivery of care can be improved, to target key interventions and to improve value-based and personalised care at scale.

Through the partnership, health data is collected in real world settings, such as routine appointments or through mobile phones, and is used by the industry researchers to target and model changes as a result of specific clinical interventions.

AstraZeneca and the Discover-NOW Hub have collaborated on several projects since 2020, including:

**1.Remote self-care for high-risk type 2 diabetes patients**

Diabetes care in North-West London costs around £600 million, with patients seen twice a year or less, making it difficult for doctors to know if care plans are being followed, or are effective.

In 2020, Discover-NOW and AstraZeneca worked with healthcare technology company HUMA to assess use of remote monitoring to identify high-risk patients and improve their care. An initial pilot through local primary care networks helped the partners to navigate working at scale, improving patient experience and outcomes, integrating clinical pathways, commissioning care across several providers and helping to support self-care for patients from diverse backgrounds. This culminated in the design and roll out of 'Fresh Start', a remote care service that includes an app-based remote monitoring programme, virtual consultations and an education campaign – all delivered to patients in their own homes.

Fresh Start is assisting healthcare providers to pick up high-risk patients earlier, to help them to self-manage their care, and it maximises clinical efficiency. It is currently being launched across north-

west London and an economic and outcomes-based evaluation is underway, using Discover-NOW's patient identification and recruitment service.

**2. Supporting NHSX through remote monitoring during COVID-19**

The Hub's existing partnership with AstraZeneca, as well as the infrastructure and digital platforms associated with the diabetes project, allowed Discover-NOW to respond rapidly to a request from NHSX. The request was to support national roll-out of COVID-19 remote monitoring and development of 'Hot Hubs' across north-west London, to diagnose and advise COVID-19 patients that were not in hospital.

Within two weeks of its launch, the service had over 150 patients on a virtual ward, 96% of whom were using the service as prescribed. Between April and July 2020, every patient using the HUMA remote monitoring app in primary care across the pilot sites recovered, with an average three minutes saved per Hot Hub patient per day, according to a 2020 NHSX evaluation.

**3. Redesigning an NHS Trust care pathway for heart failure**

Most cases of heart failure (HF) are diagnosed too late, with 80% of patients in the UK receiving a diagnosis in secondary care and emergency services, even though more than 40% will have experienced symptoms at an earlier date. Earlier diagnosis would allow for better treatment and quality of life and fewer hospital admissions.

A partnership between the Hub, patients, Eko Health, GPs and clinicians set out to redesign Imperial College Healthcare NHS Trust's HF care pathway to diagnose the condition earlier. The project created new GP referral templates within GP systems, using existing information. This in turn supported the development of improved, integrated primary and secondary care plans. To monitor outpatients and patients in primary care from their homes, the Eko Duo digital stethoscope and the Luscii remote patient monitoring platform were employed, with the aim of improving 30-day hospital re-admission rates.

Evaluation of over 100 patients showed a reduction in unplanned hospital admissions compared with standard care. The service was more efficient, allowing more time for specialists to spend on other clinical priorities. There were also more referrals to specialist HF services and better system coding of HF patients – leading to better care and ultimately better patient experience.

**4. Better care for north-west London's chronic kidney disease patients**

Chronic Kidney Disease (CKD) is often diagnosed when patients reach end-stage disease, require dialysis or transplant surgery, or have follow-on heart complications. Given that there are now effective treatments for CKD, the earlier doctors intervene, the greater the benefits to patients and the economy.

An integrated pathway that optimises early identification and management of CKD across north-west London has been developed with £200,000 funding from the NHS Insights Prioritisation Programme (NIPP). This was led by a partnership between AstraZeneca, Discover-NOW, the north-west London Applied Research Collaborative and London Kidney Network (LKN), with dedicated patient and public involvement and engagement and health inequalities workstreams.

LKN is commissioning a health economic evaluation from Discover-NOW to support wider work. Moreover, the partnership is looking at how to make the pathway improvements sustainable, so that it is embedded in clinical practice and can be scaled nationally at a later date.

**Overall influence and impact**

HDR UK investment and in-kind support from AstraZeneca has positioned the Hub at the forefront of research and clinical interventions for treating disease. Longer-term support has allowed the partners to build up expertise and collaborative ways of working over three years – capitalising on the multi-disciplinary approach across the organisations. The work has improved clinical effectiveness and begun to address the gap between national policies and clinical practice.

AstraZeneca has committed to extending the Discover-NOW consortium agreement, which will enable sustainable implementation of the integrated CKD pathway across this region of London.

The Hub has also delivered a series of feasibility, retrospective and health economic evaluation studies for AstraZeneca using the TRE, amounting to £300,000 additional income for Discover-NOW.

Ross Stone, Global Programme Director, Healthcare, at AstraZeneca, adds: “The Hub’s role as a facilitator between the NHS, academia and industry partners to provide a level playing field and a safe environment to experiment is critical to the success of innovative and transformational work.”

## Case Study 2: PIONEER (Improving care) - Errors in emergency overdose treatment cut to zero by digital prescribing tool

**Summary**

**Human errors in prescribing an antidote to paracetamol overdose have been cut from 40% to zero following the introduction of an electronic prescribing tool led by the PIONEER Hub. People coming to hospital after taking an overdose can now be confident that they will be treated swiftly and accurately, and the digital prescribing tool has taken the headaches out of the complex paracetamol overdose prescribing regimen for clinical staff. Such is the success of the tool that its roll-out is being trialled more widely in other hospitals and the Birmingham team has attracted NIHR funding to embed similar tools in both secondary care and on-the-go ambulance care.**

**Case study**

Paracetamol overdose accounts for 48% of all poisoning cases seen in UK hospitals and is linked with an estimated 100 to 200 deaths each year. People who come into hospital after taking a paracetamol overdose are usually treated with N-acetylcysteine (NAC), an antidote which breaks down the toxic metabolite of paracetamol and prevents it from damaging the liver. If it is given within eight hours of the overdose, NAC is highly effective in preventing death from paracetamol-induced liver damage, reducing deaths from 5% to 0.7%.

Unfortunately though, the treatment regimen for paracetamol overdose is complicated, which makes prescription errors common. Professor Elizabeth Sapey, Director of the HDR UK PIONEER Hub for acute care and Honorary Respiratory Consultant at the University Hospitals Birmingham, explains:

“You need different bags of fluid, of different volumes, with different amounts of drugs added to them, running over different time periods. It relies on you knowing an accurate patient weight and what their paracetamol level was, and you have to get the bags in the right order, which is why errors are common. Often people with paracetamol overdoses tend to come in late at night because there’s a bigger influx of patients with overdoses in the evenings, when there are fewer senior staff around than during the day.”

With PIONEER funding, Professor Sapey collaborated with clinicians, pharmacists and the programming team behind the Birmingham Systems Prescribing Information and Communications System (PICS) – a rules-based prescription-support system that provides real-time drug prescribing checks and recommendations. Together they developed and trialled an electronic prescribing tool which proposed the complete NAC regimen, with the aim of reducing prescription errors and providing patients with faster treatment within the critical eight-hour time window.

The project assessed error rates and delays in getting NAC treatment to patients after paracetamol overdose both before and after the tool had been introduced at the Queen Elizabeth Hospital Birmingham, in July 2018. The tool proposes the three weight-dosed NAC infusions after clinicians have entered the patient’s weight. NAC prescriptions were reviewed during a three-month period before and after the tool was introduced. Error rates were divided into dose, infusion volume or infusion rate categories. Emergency Medicine guidelines were used to spot any delays in prescribing NAC to patients.

Between July to September 2017 (before the tool's introduction) and August to October 2018 (afterwards), all 108 adult NAC prescriptions from patients admitted to the Accident and Emergency department of the Queen Elizabeth Hospital were analysed.

Findings showed that prescribing errors fell from 25% to 0% before and after the tool was brought in, and NAC was provided within the recommended timeframe in 47.4% of cases, compared with 11.1% of cases previously. Other research has shown that the prescribing error rate nationally is even higher – 40% – highlighting that this was a widespread and urgent problem.

**Influence and impact**

Following its success in Birmingham, the tool is now being implemented more widely in other secondary care hospitals, making an impact on NHS care delivery. Trials are underway to further assess its benefits and to discover if the electronic tool's improvement in prescribing and efficiency has any effects on patients' length of stay.

Professor Sapey says: "The clinical decision support tool has directly led to the award of an NIHR Patient Safety Research Collaboration that will be looking at how clinical decision support tools can be embedded into not just secondary care but also into ambulance conveyance of patients to reduce harms associated with medicines. That's a theme within the West Midlands Patient Safety Research Collaboration with the idea of testing these algorithms more widely across different populations and seeing what their true utility is in different care settings."

Importantly, in the longer term, the work with the PIONEER has demonstrated that it is possible to bring data together not just at scale, and of quality, but also at pace when there is a problem that needs solving – whether it's an emerging pandemic or the high rate of error in prescribing a complex drug regimen.

"That combination of patient partners, healthcare providers, data scientists coming together and working together has meant that we've found really important solutions—yes for the paracetamol prescribing problem, but it also gives us now a platform to look at well, what other decisions can we improve by bringing those clinical decision support tools in there? How can we reduce error rates for other areas where we know errors are commonly made?" adds Professor Sapey.

## Case Study 3: NHS DigiTrials (Data access) - Safe access to national data for faster, more representative trials

**Summary**

**HDR UK Hub funding has served as a springboard for the NHS DigiTrials service, giving it the foundation to help transform large-scale clinical trials in England. NHS DigiTrials' services make trials faster and more efficient, better represent patients and make trial outcomes data more complete. Furthermore, NHS DigiTrials has provided the data behind some landmark studies. These include RECOVERY, which rapidly identified dexamethasone as an effective treatment for COVID-19, and the NHS GALLERI cancer diagnostics study, which has recruited 140,000 healthy volunteers in just 10 months.**

**Case study**

NHS DigiTrials is developing a range of data services to support clinical trials. In a nutshell, its focus is on securely using routinely collected health care data from across England to help 'turbo charge' clinical trials to work at scale, progress faster and more efficiently, and enable more complete data for outcomes and follow-up.

The Hub started life as the NHS DigiTrials Digital Innovation Hub, with 18 months of HDR UK funding to develop a clinical trials feasibility service. This was as part of a consortium with the University of Oxford, IBM and Microsoft.

However, the legacy of the original Hub funding has been much bigger, and NHS DigiTrials has since secured additional funding to grow and develop its services on a far more ambitious scale. Now funded by the Data for R&D Programme, NHS DigiTrials also now has a patient recruitment service in pilot testing, a communication service to keep participants updated on trial progress and results, and an outcomes service to use routine healthcare data to study the long-term impacts of treatments.

Along the way, NHS DigiTrials has provided the data services muscle behind some truly ground-breaking clinical trials, while continuing to build the services it offers as part of the long-term health data infrastructure for the UK.

**Influence and impact**

NHS DigiTrials has recently launched its clinical trials feasibility service, building on the service developed using the original Hub funding, that will provide direct access to national data for trialists to run their own feasibility queries. Typically, a principal investigator has to quickly gather information from several different sources and put it together to determine if there are enough eligible patients to run a trial, or to decide on the best location for it. In contrast, NHS DigiTrials provides access to a wide range of data from across England, making the process more streamlined, especially for multi-site trials, and more representative of the population.

Heather Pinches, Head of the Clinical Trials Service at NHS Digital and DigiTrials, explains:

"All too often, feasibility and trial planning is done based on where the leading expert is for the disease being studied. So the research tends to go to where they are, over and over, rather than going to where the burden of disease is. This will also enable the opening up new trial sites. There's a direct correlation between quality of care and the degree to which that setting is engaged in research. So there's a real win-win there in terms of breaking that cycle."

Beyond its original remit, NHS DigiTrials came into its own during the COVID-19 pandemic. It pivoted from working on feasibility to the more urgent pandemic priority of clinical trials outcomes. Notably, NHS DigiTrials provided outcomes data to support to the RECOVERY trial which discovered that dexamethasone cut the risk of death from COVID-19 by one-third.

Rather than the trial team having to assemble data from 200 individual trial sites, NHS DigiTrials was able to provide them with an efficient and easier-to-manage single flow of data for England. This covered many of the trial sites and brought together data from primary care, hospitals, covid testing and vaccinations into one study. As a result, the dexamethasone finding was reported at unprecedented speed.

Heather adds: “We also removed some of the burden from the frontline NHS when they were very busy delivering care to people, because they didn’t have to capture the trial data outcomes, which are often collected locally. And you don’t lose people during follow-up because you’re not relying on knowing where they are - as long as they’re still in England, even if they’ve moved, we can provide the data for them.”

More recently, NHS DigiTrials has provided a unique direct-to-patient recruitment service for the NHS GALLERI trial to determine if a new blood test can help detect cancer early when used alongside existing NHS cancer screening. This is the fastest recruiting large-scale clinical trial in the UK – no mean feat, given that healthy volunteers are harder to reach through traditional recruitment methods.

Heather explains: “They needed 140,000 people in ten months. It would have been hard to do that scale of recruitment if they’d had to engage with individual GP practices or hospitals. But we were able to do it centrally and nationally. Also, that allowed them to keep an eye on who they were recruiting and we could dynamically adjust the invitation to make sure the trial was representative, for example over-sampling women if there weren’t enough joining.”

Reflecting on the overall impact of HDR UK Hub funding, Heather says: “While it has been only one part of the picture for NHS DigiTrials, our origins as a Hub have been really important for where we’ve got to now. It’s been a real springboard for us to broaden our ambitions and develop our full range of services. Through being a Hub and providing a valuable service during the COVID pandemic we were able to show the value of that service, develop and build confidence in our ambition and attract further funding to develop all four services.”

## Case Study 4: Gut Reaction (Public involvement) - Giving patients an active voice in how their data is accessed

**Summary**

**Working with people with inflammatory bowel disease (IBD) and members of the public has been a central part of the Gut Reaction Hub's quest to help researchers to discover better treatments for IBD and to improve patient outcomes through safe, transparent and responsible use of patient data. The Hub's Patient Advisory Committee (PAC) has helped shape Gut Reaction's strategy and processes from the outset, guiding patient data sharing decisions, engaging over 1,600 people with health data research and, along the way, co-producing the first ever training programme for Patient and Public Involvement and Engagement (PPIE) in health data research. While HDR UK's funding term for Gut Reaction ended in August 2022, NIHR BioResource is continuing to support its activities, including the national Patient and Public Review Group (PPRG) which will take Gut Reaction's PPIE work forward.**

**Case study**

The Gut Reaction Hub is a secure data resource for research in inflammatory bowel disease (IBD). It is a partnership between the NIHR BioResource, Cambridge University Hospitals NHS Foundation Trust, Crohn's & Colitis UK, Wellcome Sanger Institute, AIMES and Privitar. The initial 3-year HDR UK funding period was led by Eastern Academic Health Science Network (EAHSN) as the project management office.

One of Gut Reaction's unique strengths is that it brings together genomic, phenotypic and clinical data routinely collected from more than 38,000 IBD BioResource participants and combines it with real-world data from participating NHS Hospitals and other national datasets.

Gut Reaction's mission is to be a world-class, multi-dimensional integrated data resource for research and innovation, which uses a model based on open tools and data standards that could ultimately be scaled up and used in a similar way for other disease areas.

Specifically, the Hub includes information from BioResource participants with Crohn's disease and ulcerative colitis who have given consent for their records to be used for research, and who have had their genomes profiled, along with digital pathology, images, Hospital Episodes Statistics and genomic data.

PPIE has been 'baked in' from the start, forming one of the Hub's six workstreams. The Hub has an active PPIE group, formerly the PAC, during the HDR UK funding period and recently renamed and remodeled to a PPRG, which meets quarterly to discuss topics including data access, artificial intelligence and commercial models with other relevant workstreams. Over 1,627 patients and members of the public have attended meetings and other events. An animation describing the programme has been viewed over 1,500 times, five webinars involving PAC members have taken place, and an array of blogs and news articles have been co-developed by PPIE representatives. The team has worked with the charity Crohn's & Colitis UK to present at two PPIE day events, engaging a wider audience. What is more, listening to the needs of the PPIE group, Gut Reaction recruited seven new members to ensure that diverse views and experiences are heard.

Another notable achievement is that Gut Reaction has co-produced a PPIE in health data research training programme – the first of its kind. To date, 236 people have signed up and feedback has been positive, with one PAC member saying: "First rate online training course about PPIE – I frequently refer back to my notes from this." PAC members are also offered yearly one-to-one meetings to support their learning and development and give them the chance to feed back.

**Influence and impact**

Gut Reaction's PAC has had a meaningful impact on decision-making to achieve the delicate balance between ensuring data safety and supporting important research. For example, they have helped to define when it might be appropriate for Gut Reaction's data to be downloaded, and which safeguards would be required if this were to happen. Following in-depth discussions, Gut Reaction now only shares data as an exception, directing researchers to use a neutral TRE as the next best option, but if this isn't possible then a list of safeguards is put in place before data are shared. An application from industry has served as a test case, supporting further discussions on the practicalities of how to implement this approach.

Surveys show that 75% of the Gut Reaction team feel that PPIE influenced decisions some of the time, and 25% all of the time. Encouragingly, 63% of PAC members either agree or strongly agree that their views are heard. One PAC member says: "As a group, we raise questions about the use of health data, which wouldn't happen without us."

Rosanna Fennessy, another PPRG member, says: "Individuals with experience of IBD also bring important insights into what types of research are most needed, and our views are not always the same as study investigators. For example, I would like to see research into how you can best live with IBD, including dietary and mental health considerations, alongside studies into new drugs."

Over 88% of PAC members surveyed felt better informed about health data and 63% of PAC members strongly agreed that the initiative has been a good use of their time. Another member adds: "It feels really good to take a project all the way through to completion. It's been great to interact with professionals in a number of different fields, and to be treated with respect and parity."

The PAC and now PPRG have been essential for developing and defining the Hub's approach for accessing patient data, which goes a long way towards improving and maintaining public trust in responsible use of health data. NIHR BioResource has implemented PAC proposals for how PPIE could be improved and the PPRG review. All IBD research applications received by the NIHR BioResource will be central to the data access process in future. While the HDR UK funding term for Gut Reaction came to an end in August 2022, NIHR funding for the BioResource is confirmed until November 2024 and it has committed to sustaining the Hub's PPIE activities and remains an active member of the HDR UK Alliance. Crohn's & Colitis UK will provide support to deliver and review the PPRG.

Sarah Sleet, CEO of Crohn's & Colitis UK and PPIE lead at Gut Reaction explains: "Although Gut Reaction have always involved patients in the data approval process, this new approach confirms patient input as one of the first steps, with every IBD data request reviewed by patient representatives. The PPRG will help to ensure that Gut Reaction projects are centred around the views and needs of people with Crohn's and colitis and are ultimately driven by delivering benefits to this community."

## Case Study 5: Gut Reaction (Industry involvement) - How Tech companies help balance patient privacy with industry Inflammatory Bowel Disease research

**Summary**

**The Gut Reaction Hub is a secure data resource for research in inflammatory bowel disease (IBD) which brings together genomic and phenotypic data from patients in the IBD BioResource with real-world data from participating NHS Hospitals and other national datasets. The Hub works with two technology partners to achieve the difficult balance between keeping patient data safe and secure while allowing health data research to progress unhampered. Over its first three years, the resource has brought benefits for many users. Among them is a major pharmaceutical company project that has been a test case for the real-world complexity and analytical power demands of an industry study, helping to shape health data research policy so that others can follow in future.**

**Case study**

As anyone who works in the field will know, the perennial difficulty for any health data research is striking a balance between maintaining the privacy of those to whom the data belongs and sharing information with trusted researchers to allow vital research to progress.

The Gut Reaction Hub has enlisted the help of technology partners AIMES Management Services and Privitar Limited to provide resources and infrastructure to find a solution that is the best of both worlds.

AIMES provides the Trusted Research Environment (TRE) which safely stores the information, whilst also allowing approved researchers to securely access and view it. The AIMES TRE also contains specific tools for analysing data to support research, alongside control and monitoring technologies that stop anyone from copying, downloading or getting unauthorised access to secured information.

Another technology partner, Privitar, provides the Gut Reaction Hub with a privacy-enhancing technology to ensure that anyone who donates their information for health data research cannot easily be identified. At a basic level this 'de-identifying' process means removing unique, sensitive details, such as name or hospital number, and replacing it with a consistent key to link records from the same person together. But when it comes to linking different datasets, it is also important to consider details which could be used to reveal someone's identity or health condition when different information is combined – so the technology also de-identifies data fields such as dates and times of hospital appointments or types of treatment to make sure they are not too exact.

Watch a short video to hear from the Hub team on the role of Privitar in Gut Reaction.

Through working with its technology partners, Gut Reaction has been able to provide sustainable data services that have brought research benefits to a wide range of users – including those from industry. Lessons learned along the way are defining the regulatory and legal framework for secure data sharing and helping to develop tools and processes for working with industry, paving the way for future projects with industry.

One recent example is the Hub's support of a major pharmaceutical company study. The company wanted to investigate whether genetic variation in patients with IBD is associated with different clinical features of the disease, and how genetics might affect individual patients' varying response to treatments.

The company applied for access to Gut Reaction's dataset to investigate genes and clinical classifications of interest. After careful review by the NIHR BioResource Data Access Committee (DAC) as well as its national steering committee, the application was approved in February 2021. However, it soon became clear that the Gut Reaction TRE at AIMES was not suitable for this work, given the high-performance computing needed for genomic data analysis, and the complexity of the company's own algorithms and tools.

Gut Reaction has the clear principle that the TRE should be the default research environment, with data sharing only in exceptional circumstances. However, in this case, the Hub's Patient Advisory Committee (PAC) supported the movement of Gut Reaction data onto the company's own secure research environment, with provisos. These included: ensuring patient consent and NHS Research Ethics Committee approval and assurance that linked data could be transferred to the company's secure research environment without patients being identified. The company also demonstrated that the research environment was secure and explained who would have access, how long the data would be kept, and how it would be disposed of at the end of the project.

**Influence and impact**

The industry-led genetics of IBD research project went ahead in March 2022 and has brought in commercial income to the Hub, which has since been reinvested to further develop Hub infrastructure. But, more importantly, it has had longer-term impact for the sustainability of Gut Reaction's research resource offering. That's because it has allowed new processes and tools – the i2b2 cohort discovery tool and clinical data catalogues – to be tried and tested for working with industry in order to benefit IBD patients.

This industry-led case study and interactions with the Hub has helped to define the regulatory and legal framework for secure data-sharing that permits data to move from the Gut Reaction TRE to a company's own secure research environment. Once it had become clear that Gut Reaction's TRE was not suitable for the research task in hand, the Hub used the Privitar Data Privacy Platform to enable policy-driven de-identification of data. It has worked up an information security agreement for the company's restricted-use secure research environment. This extends the standard Hub expectations for such a research environment, with proportionate checks to verify that it is being used as intended.

At the same time, the company is developing and applying multi-modal, representation learning approaches to create patient digital 'fingerprints' that encode different views of the patient's histology, genetics, and laboratory test results data. Such fingerprints could be used for other tasks at a later stage of the research. For example, this could include sorting patients into different sub-

groups according to their genetics, making data anonymous, and generating synthetic data – artificial data used for training or testing algorithms.

Such is the success of the project that the company has made a second application to use its data. While analytical results are currently confidential, they are expected to better define the different sub-types of IBD, which should ultimately lead to more accurate and speedy diagnosis of patients. The findings could also be used to predict the course of the disease, enabling IBD patients to be better monitored. Ultimately, the bigger goal is to unpick the process underlying how the disease unfolds in different genetic sub-groups of patients, allowing for more personalised treatments for people with IBD rather than a 'one size fits all' approach.

## Case Study 6: INSIGHT - Support to catalyse innovation from small and medium-sized enterprises (SMEs)

**Summary**

**INSIGHT has encouraged applications from multi-sector research users and has a particular interest in enabling the success of SMEs by meeting their need for efficient and safe access to high quality eye health data. The Hub extended advice and expertise that is bespoke to SMEs. This counsel incorporates an assessment of data requirements, providing access and help to support Data Use Applications, and working with SMEs to translate understanding and ensure compliance with the 'Five Safes' criteria for data usage. INSIGHT's work with SMEs has highlighted a range of areas and opportunities that differ from working with global industry, NHS, charity sector and academic research users, and the Hub continues to invest resource to refine their offering for the SME sector.**

**Case study**

As the world's largest ophthalmic bioresource, INSIGHT provides SMEs in the United Kingdom with an exceptional opportunity to develop and validate their product lines in preparedness for health settings and other commercial and/or regulatory applications. A key element of their engagement activity has been to attract SMEs from outside the eye health space, but with clear relevance to advancing machine learning and Artificial Intelligence (AI) development using eye data, including for systemic conditions.

INSIGHT illustrates how eye health – with its high-volume diseases, standardised care pathways and reliance on protocolised imaging – is an attractive market for technology SMEs that are either established within health or could bring their AI development pipelines into the health sector. This is particularly salient to sector opportunity through 'Oculomics' where the eye is used as a window into systemic health, enabling technology SMEs to build innovative screening and diagnostic tools across a range of systemic health conditions, such as dementia and cardiovascular disease.

INSIGHT's SME engagement strategy has included leveraging well-established networks (university business engagement, academic health science networks), while also reaching into new networks (through HDR UK and Innovate UK). They have contributed to virtual and in-person events ranging from 'Catapult Network' events to national data symposiums. Most recently, the Hub hosted a deep-dive event – the 'INSIGHT Sandbox' – inviting SMEs and other multi-sector research users to explore the potential of discovery data and learn about key themes such as health data poverty and AI standards in health technology.

Through their engagement, the Hub has learned that working effectively with SMEs requires a differentiated approach to realising benefits and impact. Each SME faces unique operational circumstances and INSIGHT have tailored their engagement to provide services and expert support to address these challenges. In INSIGHT's experience, SME applicants frequently have deep knowledge in their technology domain area or in general business acumen. However, they often have limited understanding of the clinical research domain, the governance of data (particularly NHS patient data) and contracting. The diversity of SME users has been striking, ranging from computer scientists with no prior commercial experience, to serial entrepreneurs.

INSIGHTs experience working with SMEs also revealed a relatively low conversion rate (query to data sharing agreement). From a multitude of enquiries from SMEs to INSIGHT since late 2020, three have moved beyond an expression of interest to submitting a Data Use application, and with the advice, guidance and expertise of Hub senior leaders, these applications are now under active negotiation and project delivery stage.

Recognising the limitations and pressures on SMEs, INSIGHT purposefully incorporated this learning into developing an 'SME Playbook' (created with contributions from the University of Birmingham Business Engagement Team) which was used during the Sandbox event to showcase market opportunities for the SME sector. They also provided a structured walkthrough of the Life Sciences sector in the UK alongside offering one-to-one networking sessions.

**Enabling access for SMEs**

INSIGHT plays a holistic role in supporting SMEs to understand, access and effectively use eye health data.

***Due diligence***

The Hub is committed to supporting start-ups in their critical early stages, whilst also ensuring due diligence with respect to safeguarding NHS patient data. This is a vital step: ensuring the 'safe people' within the Five Safes process. Industry players have a historical (often international) footprint providing ample evidence to conduct due diligence and provide assurance for INSIGHT and Data Controller purposes. In contrast, small companies often do not have an established operating or research track-record.

***Defining the 'use case'***

SME research users frequently approach the Hub with broad ideas, and INSIGHT has found they need guidance to define a precise 'use case'. For example, whether the proposed AI tool would be used in community screening or in a hospital clinic, or whether it could be deployed for decision support or operate fully autonomously. In contrast, research users from industry or large academic consortiums applying for grant funding usually have predefined expert internal review processes to hone research questions prior to a Data Use Application. Enquiries from SMEs have required significantly more input from INSIGHT senior staff in order to define and refine their project aims, so as to establish an admissible 'use case' and to design an appropriate data specification.

***Defining the 'ask'***

SME research users have tended to request large data sets without a clear determination of how many data points would be required to generate analytics for intended research purposes. Additionally, there has been a tendency to alter data discovery parameters at a late stage, even after the data specification has been finalised. The Hub has observed that SME research users do not always recognise the value of data, for example, if their main comparators are free publicly available datasets or the UK Biobank. In contrast, industry research users have extensive experience of working with other data repositories coupled with a clear understanding of market value for licensing used to tailor their requests accordingly. Industry research users have demonstrated a clear idea of the numbers of images and data points that need to be acquired for successful analysis.

**Commercial arrangements**

SME applicants are more often significantly constrained in how they can fund access to data. In contrast to industry research users that prefer to define and pay for access via a one-off licence fee, SMEs have sought to defer expense to support cash flow and reduce risk. INSIGHT has worked with the Data Controllers and the applicants to develop financial models tailored for each SME that support their growth, but also demonstrate a fair-value return to the NHS. Options considered in negotiations have included revenue shares and equity stakes. Determining appropriate models of financial return and shared risk associated with data licensing to SMEs is complex and is an area of continued development. This requires time and detailed discussion with the SME to work out the most appropriate commercial relationship with the Data Controller.

**Familiarity with PPIE**

SME research users often have limited awareness or recognition of the importance and value of PPIE. INSIGHT has invested time with the SMEs introducing core principles and practice of PPIE, facilitating patient involvement so as to ensure the patient voice within their Data Use Application. This reflects how patients and public are involved in the leadership and governance of INSIGHT.

**Organisational stability and navigating change**

SMEs are notably more agile with an ability to reach decisions faster than industry, but they are also more volatile. One SME underwent an unexpected change in Chief Executive during negotiations with the Data Controller regarding their Data Use Application, which resulted in several weeks of uncertainty as to whether negotiations would continue. On occasions, the engagement with nascent SMEs has uncovered creative research ideas that could result in exciting opportunities to address longstanding health and care challenges with positively disruptive solutions. For example, the Data Use Application to INSIGHT from an SME (micro) applicant uses AI analysis of routine retinal photographs to provide a high-street 'wellness check' to look for risk of heart disease, enabling early intervention. This 'Oculomics' application is a disruptive new technology that is unlike anything that is currently on the market. Similarly, the Data Use Application from another SME applicant (micro) addresses the well-recognised problem of large-scale screening for diabetic retinopathy but uses an 'explainable AI' approach to provide both the advantages of automation alongside the ability for human readers to understand and verify the findings of the AI system.

**Impact and sustainability**

INSIGHT is committed to enabling SMEs to achieve their potential through efficient, safe access to data, alongside provision of world-class expertise in eye health, imaging, and AI. The Hub's work with SMEs remains an integral part of INSIGHT's mission to grow a multi-sector user base, recognising the special importance that these entities bring to serve exciting opportunities for innovation, growth and healthcare transformation across the UK and globally.

## Case Study 7: BREATHE (Industry engagement) - Fostering a community of industry partners in respiratory research

**Summary**

**BREATHE, the Health Data Research Hub for Respiratory Health, has worked with multiple commercial companies to better understand customer needs and to build a collaboration of partners committed to its vision of driving the use of health data for research and innovation to transform respiratory health. To formalise this collaboration, BREATHE has set up an Industry Forum open to all companies with a special interest in respiratory disease. Sharing of expertise, information and collaboration and funding opportunities has contributed to the growth of an SME, Tiny Medical Apps.**

**Case study:**

The BREATHE Hub has taken the time to understand the challenges facing SMEs with a special interest in respiratory disease, forming an Industry Forum to bring companies together. Its members are drawn from fields including medical devices, connected devices, advanced laboratory equipment, innovative AI and ML software, and middleware that connects digital products to health care records. Membership has grown from six founding members, to over 20. These are:

- Albus Health
- Arete Medtech
- Association for Anaesthetic and Respiratory Device Suppliers- Barema
- Association of British Healthcare Industry (ABHI)
- Benevolent AI
- Breathe Biomedical
- Circasia
- Clement Clarke
- Diveplane
- Eupnoos
- GE Healthcare
- Intermedical
- Novartis
- Precision Life
- Respiri
- Savana
- Smart Respiratory
- Storm ID
- Teva
- Tiny Medical Apps
- Vatic Health
- Vitalograph

BREATHE's lay Board members and its PPIE Curiosity Group review all companies wanting to join BREATHE as Supporting Partners. Once approved, the BREATHE Industry Forum seeks to match the best ideas and solutions towards improving respiratory health that combine both commercial and academic prowess.

BREATHE and its industry members have jointly developed a Terms of Reference for the Industry Forum co-chaired by an elected forum member, alongside the BREATHE Partnerships Lead. The Industry Forum Co-Chair is also serves as a member of the BREATHE Board and provides added commercial challenge and know-how to the Hub's forward view and sustainability plans.

BREATHE can provide access to a team of respiratory experts who are passionate about improving respiratory care and outcomes. These include Principal Investigators (PIs) who are clinical and data experts in primary and secondary care, as well as epidemiologists and data scientists who are supported by the BREATHE Commercial Team.

Involving pharmaceutical and health tech leaders provides the BREATHE Industry Forum members with world-class opportunities for partnerships and collaborations. Member companies are not just from UK and this global reach provides a diverse portfolio of products, plans and ambitions linked to patient care and healthcare research. Furthermore, some forum members bring with them important datasets because of their work and products. BREATHE works with these member companies to import, curate and link their data with the wider BREATHE data 'ecosystem' housed in the Secure Anonymised Information Linkage (SAIL) Databank TRE.

The forum taps into not only direct commercial opportunity and income, but also the collaborative academic/SME grant and loan schemes supported by UKRI. Other forum membership benefits for founding and supporting partners include:

- Expertise (clinical and epidemiological)
- Peer-to-peer relationships with like-minded companies
- Support with, and collaboration on, funding opportunities
- Access to curated data/analysis to support bids and their innovation and research and development in general.

**Influence and impact**

One BREATHE Industry Forum member is the UK-based SME Tiny Medical Apps, which developed the Digital Health Passport mobile app that supports young people to self-manage their asthma. Being part of BREATHE has helped the company by:

- Providing expertise from an extensive network of world-renowned asthma specialists
- Supporting successful funding applications from Innovate UK and NHSX that have led to new products being rolled out across all of NHS England, and which have export potential
- Linking to PPIE groups for co-design and validation of the product
- Forming collaborations with other commercial organisations
- Education about new government policy, guidance, and grant opportunities.

Greg Burch, Tiny Medical Apps' CEO and Clinical Director (and BREATHE Founding Partner and Industry Forum member) said: "Our journey from a small R&D SME to having a product that is about to scale nationally would have been infinitely more difficult without the support of BREATHE."

## Case Study 8: (Infrastructure re-use) - How the Hubs, Gateway and Alliance contributed to the UK's COVID-19 pandemic response

**Summary**

**HDR UK swiftly changed course during the COVID-19 pandemic to join the national response to the emergency. HDR UK developed and coordinated existing data platforms with skilled people and emerging data from the pandemic. As part of this national effort, the Innovation Gateway, Alliance and Hubs provided a 'readymade' infrastructure to quickly answer the most pressing research questions and fast-track them into policy in a way that would have otherwise taken far longer. As such, HDR UK led one of the six national prioritisation programmes commissioned by UK Chief Scientific Advisor Sir Patrick Vallance to lead the UK response to the pandemic. The legacy of this work is a 'tried and tested' UK health data research infrastructure that is ready for whatever future shocks are coming our way – whether that is another pandemic, or some of the biggest health problems society faces, including cancer and winter pressures on the NHS.**

**The Digital Innovation Hubs – rapid data linkage to answer the big questions**

On the ground, the Health Data Research Hubs played a critical part in linking diverse data to support the UK's response to the COVID-19 pandemic, producing tangible results. Each Hub was able to find ways to use its data and services to respond, either directly tackling the virus and its effects or indirectly, such as looking at the consequences of the pandemic on health and disease.

HDR UK's NHS DigiTrials Hub provided outcomes data to support to the RECOVERY trial which discovered that dexamethasone cut the risk of death from COVID-19 by one-third. Rather than the trial team having to assemble data from 200 individual trial sites, NHS DigiTrials was able to provide them with an efficient and easier to manage single flow of data for England that covered many of the trial sites and brought together data from primary care, hospitals, covid testing and vaccinations into one study. As a result, the dexamethasone finding was reported at unprecedented speed.

Furthermore, the EAVEII study, led by the BREATHE Hub, used COVID data from the whole population of Scotland to provide the first real-world evidence that a single dose COVID-19 vaccine was effective.

Other notable achievements of the Hubs as part of the UK's response to the pandemic were:

- PIONEER curated and analysed almost real-time data analysis of the pandemic unfolding across West-Midlands. This led to early identification of ethnic differences in the impact of COVID-19 and data dashboards to allow monitoring of hospital capacity. PIONEER also identified an increase in venous thromboembolic events (VTE) associated with COVID-19.
- Linking COVID-19 testing data with all haematology, biochemistry, immunology, microbiology and therapeutic monitoring results for 2.5 million patient records by the Discover-NOW Hub
- The DATA-CAN Hub collected and analysed 'real-time' data from UK cancer centres, covering over 3.5 million people. This data provided a valuable insight into the effects of the COVID-19 pandemic on cancer patients and services and contributed to the decision to restore cancer services.
- Under the leadership of DATA-CAN's Scientific Director Prof Mark Lawler, the Hub brought together data and expertise from across Europe to delineate the impact of COVID-19 on cancer services and cancer patients which has underpinned a pan-European Time To Act Campaign. Now

translated into 30 languages, this campaign has been launched in 12 European countries by national ministers of health.

- HDR UK's INSIGHT Hub provided the first reliable estimates of the scale and severity of the vision loss arising from delays in treatment for newly diagnosed wet macular degeneration (wet AMD) during the COVID-19 period, informing NHS providers' strategies on how to optimise patients' care during the service's recovery. The IBD (inflammatory bowel disease) Registry, a key partner of HDR UK's Gut Reaction Hub, developed a COVID-19 IBD Risk Tool to allow patients to self-assess their risk of being infected.

**The Innovation Gateway – a 'one-stop shop' for COVID-critical research data and the most urgent questions**

As the COVID-19 pandemic unfolded, the HDR UK Innovation Gateway was focused to meet emerging health data research needs. Early on, HDR UK worked with the research community and policy makers to identify the most urgent emerging COVID-19-related research questions, which were made discoverable to researchers through the Gateway. This was reported on a weekly basis to the UK government's Strategic Advisory Group for Emergencies (SAGE). Use of the Gateway in this way saved considerable time and resource – had it not been there, researchers would have had to scramble to establish one-to-one relationships looking for relevant datasets to answer their research questions.

By the summer of 2020, the UK Government's Chief Scientific Adviser and the Chief Medical Officer, together with experts and funders, including UKRI, identified several key areas where there was a clear need for the UK to increase research scale or provide infrastructure. The Government established six National Core Studies (NCS) in immunity, trials, surveillance, transmission and environment, longitudinal health, data and connectivity.

HDR UK and the Office for National Statistics led one of these six studies - the Data Connectivity Core Study. On a scale never seen before, this study brought together the national health datasets needed for the COVID-19 response and connected them with other national core studies as well as forging links across the health data research landscape.

The Gateway was critical to the study because it provided the UK research community with a single platform to access information about pandemic-relevant datasets and a way to request access to them. Now, 110 datasets are available via the Gateway by Data & Connectivity NCS. What is more, transparent information on the 271 research teams using these datasets for research, in many cases including the outputs their research produced, is now captured in the Gateway Data Use Register.

**The Health Data Research Alliance – common rules, trust and partnership to launch a rapid, coordinated response**

The community of Alliance membership that had been built up by HDR UK prior to the pandemic provided a ready-made network of data custodians that were able to swing into joint action to put the datasets to best use. Thanks to the underpinning rules, relationships, trust and mutual understanding of ways of working that had previously been established, the Alliance members could respond to the pandemic in a coordinated way from the outset.

Together, the Alliance and Gateway enabled a prompt response by providing an infrastructure that enabled rapid sharing of questions, projects and data sets and allowed them to be discoverable. HDR

UK also put out around 20 rapid funding calls through the Gateway that allowed researchers to secure funding within a matter of weeks to access and work with datasets on the Gateway.

**COVID's legacy for health data research in the UK**

During the course of the pandemic, HDR UK has proven the enormous value of having research infrastructure in place to support broad health research and innovation, which can be rapidly repurposed to meet emerging needs. As a result, government and policymakers can be assured that if another pandemic emerged tomorrow, the UK could use exactly the same tried and tested methodology – namely, identifying the most important datasets, onboarding them onto the Gateway and working with data custodians to accelerate their processes.

Many of the Gateway's ways of working and early thinking and design were reused and repurposed for the international data sharing effort for the international COVID-19 data research alliance, ICODA. This was put in place by HDR UK to support a rapid response to combat the global effects of the COVID-19 pandemic, and to leave a lasting set of infrastructure and governance processes that enable an efficient data response to future pandemics and other global health challenges.

Furthermore, the Gateway was used to collect together many different packages of serology (antibody or antigen blood test) data for COVID-19 patients. Accelerated development of the HDR UK cohort discovery tool supported The COVID - Curated and Open aNalysis aNd rEsearCh plaTform (CO-CONNECT) project, allowing easier access and discovery of this serology data. The cohort discovery tool allows dynamic querying of datasets to find out how many people with particular desired characteristics for a specific study are available – for example COVID-19 patients who also had diabetes – to quickly see whether it will be feasible to carry out the research project.

Adoption and use of the Gateway and infrastructure was challenging. HDR UK learned an enormous amount by having a high value use case to drive the development and adoption of infrastructure. Through HDR UK reporting to SAGE every other week, close connections have been made between academics using the datasets and the public health agencies and government officials making decisions around policy and planning. Thanks to this, analysis could be done in a matter of days and decisions reached that influenced national policy.

One direct legacy of this work, for example, has been a recent funding call launched by HDR UK and the Department of Health inviting research on the unprecedented pressures facing the NHS in winter 2022, to pinpoint and rapidly answer the most urgent questions. Within a week of the Chief Medical Officer's announcement about the issue, HDR had put together a funding call asking the research community to put forward research ideas that would give academics the chance to understand the winter pressures facing UK health and social care. Within two weeks, 34 research groups had responded, indicating that the world is more research-ready than it was before.

Emerging crises – from pandemics to pressures on health services – can be met with a rapid response now that the underlying infrastructure, datasets, governance and public trust in research is in place.

## Case Study 9: Alliance (Policy) - Shaping UK and European policy on safe data sharing through Trusted Research Environments

**Summary**

**Data-sharing policy and practice in the UK and beyond have been shaped by a critical policy paper setting out principles and guidance for building Trusted Research Environments. Co-authored by HDR UK, the paper has been downloaded 5,550 times in the year since it was published. Its influence joins a growing movement in data science which includes not only health data research but wider social and government data research and a joint action by 25 European countries.**

**Case study**

Trusted Research Environments (TREs) are highly secure computing environments that approved researchers may access to carry out research that saves and improves lives. Rather than the data being shared or distributed, it stays safely in one place, a bit like a secure library, with access only given to trusted researchers. In this way, people whose data is within the TRE can be confident that their personal health data is being accessed securely and that their privacy is protected, while researchers benefit from a one-stop-shop to access valuable data and analytical tools all in one place.

Along with partners and collaborators, HDR UK has championed and led the use of TREs, co-authoring a Green Paper in 2020 which proposed the direction of travel. Working with stakeholders to reach a consensus, HDR UK co-authored and published an influential White paper, Principles and best practice for trusted research environments, in December 2021.

Setting out the evidence base for how TREs can work, the paper has been critical for building public and professional trust. It offers practical insights into how TREs would work within the NHS and was perfectly timed to feed into NHSX (now the Transformation Directorate) and Health and Social Care Executive developments happening at the same time.

**Influence and impact**

In the year since its publication in December 2021, the paper has been viewed over 7,000 times and downloaded 5,550 times. This is testament to the huge level of interest from data scientists worldwide.

Not only has its guidance and principles fed into the policy work of NHSX, but its TRE model was included in the wider ranging April 2022 Goldacre review to improve the use of health data in the NHS, commissioned by the Secretary of State for Health and Social Care. NHS England has formally announced that it will adopt the approach to TREs and a new NHS Digital Secure Data Environment Service is now available to a small group of users, which will ultimately become the default way for researchers to access all health and social care data held by NHS Digital.

Contributing to the bigger picture beyond health data, the TREs model is incorporated into the Data and Analytic Research Environments UK (DARE) programme, of which HDR UK is a partner. DARE has the ambition to design and deliver a collaborative UK data research infrastructure. This potentially widens the remit of TREs to include social science data, other individual level government data and all UKRI-funded research cohorts.

Safe and secure linkage, or federation, of diverse patient datasets to make them easy to find, is another key area set out in the White Paper, and DARE is working on exemplar federation projects, such as bridging the TREs of the NIHR Cambridge Biomedical Research Centre and Genomics England.

Beyond the UK's shores, the paper's vision of a federated model of secure data environments, similar to that being built under HDR UK's Innovation Gateway, is also envisaged as part of TEHDAS (Towards European Health Data Space). This joint action between 25 European countries is developing European principles for the secondary use of health data. Governance and proposals currently being drafted by the European Commission.

Professor Tim Hubbard is Professor of Bioinformatics at King's College London. He led development of one of the first TREs at Genomics England, leads the TRE workstream of the Health Data Alliance and was one of the White Paper's authors. He believes that the models, guidance and evidence set out in the paper provide a win-win solution for everyone involved - something that has been a long time coming for health data research:

"Everyone thinks there's benefit in terms of commercial algorithms, improvements in treatment, improvements in policy around looking at large amounts of individual level data - it's basically thought to be the new economic foundation. But there are the advocates on one side and people worrying about individual privacy on the other side and that's been a barrier to using health data, accessing it, for around 15 or 20 years.

TREs represent a technical and policy solution that sits between. While we've had, for example, data safe havens before, it's never been quite so clearly defined with this strong position that the data doesn't get distributed; the data custodians are retaining tight control over who accesses the data. For researchers, the downside of that is they have to work in a controlled environment, but in exchange, they will end up getting access to a wider amount of data than they did previously.

So it's a quid pro quo. I think this may be the model which can square the circle of those two opposing positions."

## Case Study 10: Alliance (Transparency) - Clearer windows on data research: standards for data use registers

**Summary**

**Data use registers are critical for trust and transparency because they show how and why people's data is accessed for research, and by whom. In 2021, the Health Data Research Alliance recommended a set of standards to improve the transparency of data use registers. This has heralded a change in practice, with 34 out of 67 Alliance members now having publicly discoverable data use registers. NHS Digital has furthermore renamed its register to align with the Alliance's recommendations. More than 93% of respondents to a public consultation also endorsed them, showing public support for the safe, secure use of health research data to make life-saving discoveries.**

**Case study**

The COVID-19 pandemic demonstrated the power of using data for research and raised public awareness of the good that can be achieved through use of people's health data. But it has also raised questions about the potential risks to individual privacy carried by the use of sensitive data, which can undermine public trust in the process. Transparency in how health and social care data is used by researchers and innovators is therefore crucial to building public trust.

A data use register is a public record of the data which an organisation shares with others for research, innovation and service evaluation. To date however, the information available in data use registers has often been difficult to access and has lacked standardisation across organisations. In some cases it wasn't even publicly available. Furthermore, it has been difficult to easily see the outcomes and outputs that have come from using the data.

The Alliance recognised that the public, researchers and funders would all benefit from a clear mechanism that highlights research projects using data and the outputs derived from those projects. They interviewed, surveyed and held workshops with more than 100 data custodians, researchers, funders, policy makers and members of the public from nearly 50 organisations to discuss standards for data use registers. The Alliance also carried out an analysis of data use registers published by health data custodians in the UK. This provided them with valuable insight into the practices of data controllers and custodians with established data use registers, as well as baseline to measure the impact of this work.

In July 2021 the Alliance published a Green Paper, proposing a new standard for data use registers which aims to make it clear to all concerned how and why data is accessed, which all organisations responsible for providing access to data can adopt. The paper outlines four recommendations focussing on content, format and frequency of publication. Development of a data use register on the Innovation Gateway has also been shaped by the standard, improving the transparency of data uses for datasets made discoverable by Alliance members through the portal.

**Influence and impact**

After the Green Paper was published, NHS Digital renamed its register to align with the Alliance's recommendations. Of the 46 Alliance members analysed in summer 2021, 48% (22) had data use

registers that were discoverable via public websites. By November of the same year, a change in practice was already underway, with 34 out of 67 Alliance members having data use registers that are discoverable via public websites. By November 2022, 19 organisations had also adopted the register standards and 845 data uses can now be discovered on the Gateway.

A public consultation of the recommendations was also held in July 2021. Results showed that the recommendations were endorsed by 93% of respondents, the majority of whom were patient and public representatives. The high level of interest and input from lay representatives the importance of the window on research that data use registers provide, as well as the importance of involving the public in all stages of the data-driven research cycle.

Nada Karrar, Project Manager - Data Access Registers at Health Data Research UK, explains: “We believe through the drive and leadership of the Alliance, there is great potential for transparency on data use to become a priority for all data custodians moving forward.

To better support Alliance member organisations meet the principles of transparency and align with the standard proposed, Health Data Research UK has developed a data use register for the Health Data Research Innovation Gateway (the ‘Gateway’). A development welcomed by many of the patients and members of the public we have engaged with, including members of the HDR UK Public Advisory Board who highlighted the need for a register in their recent recommendations to building trust in data access.”

## Case Study 11: DATA-CAN (Public Involvement) - Patient voice shapes project honing Roche's understanding of breast cancer

**Summary**

**Roche has deepened its understanding of the features, treatment options and patient outcomes in early-stage triple negative breast cancer thanks to working with the DATA-CAN, the HDR UK Hub for cancer. Working with the pharmaceutical company, DATA-CAN completed a combined analysis of curated data from two NHS cancer centres in England and Scotland covering 4.5 million patients. Strong patient and public representation has shaped the project and ensured that it went ahead despite the Hub leadership's initial reservations. This reflects the Hub's innovative way of working by placing patients and public front and centre of everything it does.**

**Case study**

Around 55,000 patients are diagnosed with breast cancer in the UK every year, and one fifth of these have so-called 'triple-negative' breast cancer (TNBC). These women don't have oestrogen, progesterone or human epidermal growth factor receptor 2 (HER2) on their cancer cells. Since many drugs are designed to target these receptors to kill the cancer cells, this means triple negative patients have fewer treatment options. These patients also have worse outcomes compared to those with other breast cancer sub-types.

The HDR UK DATA-CAN Hub is a UK-wide partnership that is unlocking the power of health data to improve cancer care. In 2021, the Hub worked with two UK cancer centres to collect and curate two historic datasets from patients with early-stage TNBC.

The first dataset, from Leeds Teaching Hospitals NHS Trust, was drawn from cancer-specific electronic health records on every cancer diagnosed in Leeds since 1990: 3.1 million patients. Detailed diagnosis, treatment and outcome data was captured on the Patient Pathway Manager platform developed by DATA-CAN Clinical Lead Professor Geoff Hall. These data are linked with other datasets including finance, Hospital Episode Statistics and cancer registration data, as well as the Cancer Outcomes and Services Dataset.

The other dataset, from NHS Lothian (Edinburgh) was from a clinical phenotyping, process, treatment and outcomes database that had been maintained over 40 years, including data from 1.4 million patients across four NHS Health Boards across the South East Scotland Region. These are integrated with regional electronic health systems to describe each patient's complete cancer journey.

The project to unite these datasets required detailed data curation by the DATA-CAN team in Leeds, conversion of both datasets into a common data model, and combination and analysis of the data in a trusted research environment to protect patient privacy. Completing a combined analysis of curated data from two NHS cancer centres spanning England and Scotland was a first for DATA-CAN, and an exemplar for future ambitious studies.

**Influence and impact**

Following curation and analysis of the data, DATA-CAN securely shared the anonymised results with Roche Products Ltd. This has aided the company's understanding of the features of early TNBC, its treatment options and health outcomes for patients to better support drug development and access

to treatments. The curated data and insights are also available to NHS clinicians and academic researchers, who can benefit from deeper understanding of how this disease sub-type is treated in two major NHS centres, as well as higher quality data for patient management and future research.

One of DATA-CAN's strengths is its innovative way of working which places the patient front and centre in all their data activities and decisions – a positive attribute of the Hub that has featured in the British Medical Journal (BMJ)'s Partnerships in Practice blog.

DATA-CAN works directly with patients, relatives, carers and the public to bring their distinct voices to complex conversations about data and participation in research. Patient and public involvement and engagement (PPIE) members review, approve and can veto any DATA-CAN proposal or contract. The Hub has a training programme for patients and the public with an interest in patient data which builds their confidence to speak up and encourages better, more effective partnerships in making decisions, especially when it comes to commercial activities.

For example, in the TNBC project, PPIE members challenged DATA-CAN's leadership to commit to the project after it had initially been rejected because it seemed too complex to deliver within the proposed timeframe. Patient and public members' passion and reasoned arguments on the project's positive impact for patients persuaded DATA-CAN that it was worth the risk.

Jacqui Gath, a cancer survivor and one of the DATA-CAN patient representatives reviewing the project, commented: "This is a vital piece of work, as patients know there is no very effective treatment for TNBC except older drugs. Consequently, this type of breast cancer is much feared."

"Because DATA-CAN has committed time and effort to hearing the patient voice, we do likewise with no regrets. I know we are making a difference, adding value and ultimately benefitting patients. Vetoing a project is 'no big deal'. Why? Because not only are we empowered; we feel empowered to make decisions," Jacqui adds.

Professor Mark Lawler, Scientific Director of DATA-CAN, explains: "We recognised from the outset that to be confident to voice their views and concerns, patient and public partners must be empowered. This is why there are two voting patient members on DATA-CAN's steering board, making them more influential than any other individual DATA-CAN partner. Two patient members also sit on the management group and on specific project groups, again making sure they are central in all the key decisions in the use of data. Patients are our north star – we follow their lead."

## Glossary

AAAMD – Action Against Age-related Macular Degeneration

AI – Artificial Intelligence

AIMES – AIMES Management Ltd, an intelligent data solutions design and development company

Alleviate – the Advanced Pain Discovery Platform (APDP) Data Hub

Alliance, The – the Health Data Research Alliance

APIs – Application Programming Interfaces

BREATHE – the Health Data Research Hub for respiratory disease

BCP – BC Platforms, a data science solutions company

CO-CONNECT – The COVID - Curated and Open aNalysis aNd rEsearCh plaTform project

CCWS – Cross-cutting workstreams

D2EDPM – the UK Government's Data to early diagnosis and precision medicine challenge

DAR – Data Access Request

DARE – Data and Analytic Research Environments UK

DATACAN – the Health Data Research Hub for cancer

DATAMIND – the Health Data Research Hub for Mental Health

DIH – Digital Innovation Hub

Discover-NOW – The Health Data Research Hub for real-world evidence

EAHSN – Eastern Academic Health Science Network

EHDEN – European Health Data & Evidence Network (EHDEN)

FAIR – data which meet the principles of being Findable, Accessible, Interoperable and Reusable

Five Safes/Five Safes Framework – safe data, projects, people, settings and outputs; a set of principles which enable data services to provide safe research access to data

Gateway/Innovation Gateway – the Health Data Research Innovation Gateway: a web-based platform to discover and request access to UK health datasets for research and innovation

Gut Reaction – the Health Data Research Hub for inflammatory bowel disease

HDR UK – Health Data Research UK

HF – heart failure

HRA – Health Research Authority

IBD – inflammatory bowel disease

ICODA – International COVID-19 Data Alliance

IG – Information Governance

INSIGHT – the Health Data Research Hub for Eye Health

ISCF – Industrial Strategy Challenge Fund

ISCF DIH - Industrial Strategy Challenge Fund Digital Innovation Hub

IUK – Innovate UK

LKN – London Kidney Network

ML – Machine Learning

MRC – Medical Research Council, part of UK Research and Innovation

MVP – minimum viable product

NCS D&C – National Core Study: Data and Connectivity

NHS DigiTrials – the Health Data Research Hub that supports clinical trials, providing safe, authorised access to patient data to help trials reach and benefit as many people as possible

NHSX – a UK Government unit with responsibility for setting national policy and developing best practice for NHS technology, digital and data

NIHR – National Institute for Health Research

NIPP – NHS Insights Prioritisation Programme

ODI – Open Data Institute

OMOP CDM – European health data Observational Medical Outcomes Partnership Common Data Model

ONS – Office for National Statistics

PAC – Patient Advisory Committee

PIONEER – the Health Data Research Hub for acute care

PPIE – Patient and Public Involvement and Engagement

PPRG – Patient and Public Review Group

PRINCIPLE – Platform Randomised Trial of Treatments in the Community for Epidemic and Pandemic Illnesses

ROs – Research Organisations

SAGE – the UK Government's Scientific Advisory Group for Emergencies

SAIL Databank – Secure Anonymised Information Linkage Databank

SDE – Secure Data Environment

SME – Small-to-medium-sized Enterprise

TEHDAS – Towards European Health Data Space

The Alliance – the HDR UK Alliance

TNBC – Triple-Negative Breast Cancer

TRE – Trusted Research Environment

UKRI – UK Research and Innovation

HDRUK

Health Data Research UK

Funded by Innovation UK as part of the Data to Early Diagnosis and Precision Medicine Challenge