



**IMPROVING UK HEALTH DATA:
IMPACTS FROM HEALTH DATA RESEARCH HUBS**

April 2021

Pioneering data research centres enable cutting-edge research and innovation and improving lives

Health Data Research Hubs – the story so far

In October 2019, seven [Health Data Research Hubs](#) were launched as UK centres of excellence for expertise, data, tools, scientific knowledge and open innovation to maximise insights and innovations developed from UK health data. This network of Hubs is pioneering and informing the development of health data research in the UK, and globally, and is demonstrating how insights from data are improving people's lives.

The Hubs were established by Health Data Research UK (HDR UK) as unique models of partnership and collaboration that bring together datasets, safely and securely, and make them available to academic and industry groups wishing to conduct health research in a trustworthy way.

Since launching, the Hubs have made 157 datasets - from genomic data to clinical and public health data – discoverable on the [Health Data Research Innovation Gateway](#) (the 'Gateway'). This is an important first step in producing a transparent single front door to health data for research in the UK.

The Hubs have delivered over 300 projects involving hundreds of partners, over 20,000 meaningful patient and public interactions, and 2,300 training activities. The impact of these projects has informed UK policy decisions on the effectiveness of COVID-19 vaccines, created tools to improve clinical decision-making in the management of patients with vascular disease, and driven research in cancer, heart disease and hospital care pathways by linking routinely-collected data. Other examples include:

- **The world's first evidence of the effectiveness of the single dose of the AstraZeneca and Pfizer vaccines** on reducing COVID hospitalisation and mortality at 28 days through data research on 5.4m people (BREATHE Hub)
- **Accelerated recruitment to the PRINCIPLE trial**, linking routine datasets to NHS Test and Trace and integrating patient engagement, resulting in the world's largest Phase 3 trial of community-based treatments for COVID-19, and finding that inhaled budesonide shortens recovery time (NHS DigiTrials Hub)
- **Rapid development of new guidance for venous thromboembolism** (blood clots in the veins) that has been adopted by 60 hospitals across the world (PIONEER Hub).

Throughout the pandemic, all Hubs utilised their rich data sources to support areas of analysis that included infection rates, vaccine roll-out, and to understand the impact of COVID-19 on specific disease areas.

Collectively the Hubs have driven significant improvements in data, guided by HDR UK's innovative [data utility framework](#), particularly in the richness of metadata (descriptions of the datasets), pathway coverage, and data management processes. Each of the Hubs are developing their offers based on industry requirements and are becoming sustainable through industry and academic partnerships.

To ensure data is accessed and used in a trustworthy way, the Hubs are guided by patients and the public, with representatives contributing to decision making, projects and processes. Plans for meaningful patient and

public involvement and engagement (PPIE) were embedded from day one, resulting in better research and a change in the culture of health data research by making it more accessible for everyone.

The Hubs are part of a four-year £37million investment from the UK Government Industrial Strategy Challenge Fund (ISCF) led by UK Research and Innovation, to create a UK-wide system for the safe and responsible use of health-related data at scale and to stimulate economic growth through research activity.

HDR UK and its partners are now scaling this programme to further develop a trustworthy health data ecosystem in the UK that will transparently, reliably, and competently harness the power of health data. This has the potential to position the UK as the place to do health data research and, through working with global partners, provides an opportunity to benefit many more millions of people across the world.

“The Hubs are already demonstrating the huge potential value of health data for people across the UK and for society as a whole. It has been fantastic to see the progress being made to improve health data, something that has always presented a major challenge given the fragmentation of this data worldwide.

This is a real boost for society, the life sciences industry and academia.”

John Jeans, Chair, Digital Health and Social Care Innovation Centre

“The Hubs have demonstrated real impact from their work to involve patients and the public, including in decision making; raising awareness and increasing participation; improving transparency, understanding and accessibility for members of the public.

Our hope is that they will build on this early work and that patient and public involvement and engagement will remain embedded in everything they do.”

Sarah Brooke, patient and public representative

“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.”

Ross Stone, Global Programme Director, Healthcare, AstraZeneca

Contents

The Hubs.....	5
Hub Impact and Sustainability	6
Improvements to Health Data.....	9
Engaging Patients and the Public.....	11
Appendices.....	12
• Appendix one – Data Improvement Case Studies.....	13
• Appendix two – Data Utility Breakdowns.....	56
• Appendix three – PPIE Case Studies.....	60
• Appendix four – PPIE plans.....	73
• Appendix five – Hub Projects.....	92

The Hubs

Health Data Research Hubs are centres of excellence with expertise, tools, knowledge and ways of working to maximise the insights and innovations developed from the health data.

[BREATHE](#) is enabling the use of respiratory health data in cutting-edge research and innovation, to address conditions such as asthma and chronic obstructive pulmonary disease (COPD).

[DATA-CAN](#) is a UK-wide partnership that aims to unlock the power of health data to improve cancer care.

[Discover-NOW](#) works with data from the NHS in London and Manchester, with appropriate controls and consent, to support research to find new ways to diagnose and treat disease and make care for patients better and safer.

[Gut Reaction](#) harnesses the power of health data to help identify the best treatment for each person with Crohn's Disease or Ulcerative Colitis. Working with patients, academia and industry, it is uniting high-quality health datasets from trusted partners and supports their use for research.

[INSIGHT](#) is focused on eye disease and its application to wider health, including diabetes and dementia. It uses anonymised large-scale data and advanced analytics, including artificial intelligence, to develop new insights in disease detection, diagnosis, treatments and personalised healthcare.

[NHS DigiTrials](#) focusses on the development of a foundation service to improve the assessment of clinical trial feasibility – supporting improved planning and delivery of clinical trials in the UK.

[PIONEER](#) allows innovative healthcare companies to develop, test and deliver advances in acute clinical care.

The first funding assessment for the Hubs, Milestone 1, in December 2019 saw the Hubs sign up to HDR UK's Principles for Participation, make the metadata about their datasets available and discoverable through the Gateway, outline their plans for patient and public involvement and engagement (PPIE), and set up physical locations around the UK. Milestone 2, 18 months on and covered in this report, focuses on the quality improvement of their datasets, the impact of the data, and engagement with patients and the public.

This report provides highlights and learning points from across the Hubs, as a head start for organisations looking to learn from and replicate the successes of the Health Data Research Hubs.

For more information on the Health Data Research Hubs visit www.hdruk.ac.uk/helping-with-health-data/our-hubs-across-the-uk/ or email Ben Gordon, Executive Director: Hubs and Data Improvement, at ben.gordon@hdruk.ac.uk

Hub Impact and Sustainability

The COVID-19 pandemic brought into focus the need for analysis of a broad range of rich datasets at scale. All Hubs have used their data improvement efforts to support the NHS, academia, and industry, in utilising the power of data to understand and overcome COVID-19.

They have had a direct impact on clinical activities; such as Discover-NOW's digital innovation testbed for remote monitoring, NHS DigiTrials support for a number of randomised controlled trials testing a range of potential treatments for COVID-19, and DATA-CAN's analysis of real-time data on excess cancer deaths that contributed to decisions to restart cancer services.

They have supported urgent analysis that has informed government policy, linked crucial data such as testing datasets with pathology to aid forecasting, and supported commercial activities such as pharma vaccine uptake and effectiveness, and answered small-medium size enterprise (SME) research questions that would not have been feasible previously.

Although it is still early days for the Hubs, with over 300 contracts, it is unambiguous that their services are useful and already being widely used by industry, academia, NHS and charities within the UK and internationally. These contracts reflect the range of commercial models identified in the [Academy of Medical Sciences report](#) and include academic and industry partnerships, demand-led support packages, and funding grants. They have developed flexible costing models to encourage partnerships across the health sector, such as with the SME community. Confidence in the Hubs and their strong service offers – particularly in response to COVID-19 – has already led to long-term commitments from industry and academia. Many Hubs are already bringing in income to cover their running costs, in some cases approaching a sustainable financial position. This will help to establish the Hubs' position in supporting research for years to come.

Impact and Sustainability case study highlights

A demonstration of how **NHS DigiTrials** was able to pivot from a singular focus, is the partnership established with the RECOVERY (Randomised Evaluation of COVID-19 thERapY) trial.^{1,6} The RECOVERY trial was established in March 2020, as a randomised controlled trial to test a range of potential treatments for patients hospitalised with COVID-19. Over 35,000 patients have been enrolled so far from 177 NHS hospitals in the UK. A novel, collaborative approach based on making key datasets available also led to significantly enhanced recruitment to PRINCIPLE – a platform clinical trial of community-based treatments for people over 50 that have tested positive for COVID-19. Mapping the data flow and timing from the NHS Test and Trace App, they were able to contact eligible candidates earlier, and registrations for the PRINCIPLE trial increased from an average of 87 people per week to 325 people per week and, by 31 March 2021, 4,671 people had been registered to the trial. This increase enabled the inclusion of inhaled budesonide as a treatment, which was found to shorten recovery times, with the trial findings having the potential to change clinical practice globally.

The success of NHS DigiTrials in supporting the RECOVERY and PRINCIPLE trials has generated significant interest from the clinical trials community. The Hub now has a sizeable pipeline of work that covers a range of market segments. As of April 2021, NHS DigiTrials has 17 trials that have an active data sharing agreement, 14 trials that they are taking through NHS Digital's Data Access Request Service (DARS), and 24 trials that are at a pre-application/enquiry stage.

Discover-NOW has built a strong service offer with a firm track record and is creating a unique testbed approach for innovative data and digital services.^{1,3} This is enabling more innovative work and significantly increasing Hub income. The Hub is investing in a Trusted Research Environment (TRE) with partner, IBM, that

will underpin Hub sustainability plans. During COVID-19, NHSX reached out for support with the national response for COVID-19 remote monitoring. This was as a direct result of their collaboration with AstraZeneca in building digital remote monitoring solutions for patients with Type 2 Diabetes and service offer expertise. Within two weeks from launch, at the peak of COVID-19 in London, the service had over 150 patients on the virtual ward and 96% of patients using the service as prescribed.

During the first lockdown, **DATA-CAN** and UCL's Institute for Health Informatics worked with UK cancer centres to collect and analyse real-time hospital cancer service data.^{1,2} Curation and analysis of the data found that, compared to pre-COVID-19 levels, there were significant drops in urgent referrals (70% decrease) and chemotherapy treatment attendances (40% decrease). Their analysis suggested that the COVID-19 pandemic could result in 7,165 more deaths in people with newly-diagnosed cancer and it could rise to 17,910 additional deaths if all people currently living with cancer were considered, with 78% of excess deaths occurring in patients with >1 comorbidity. Their rapidly-assembled MedRxiv pre-print publication was shared with the UK's four chief medical officers, the National Clinical Director for Cancer for England and SAGE, contributing to decisions to restart cancer services.

The highlight of the first 18 months of the **Gut Reaction** programme has been the commitment of commercial companies to long-term sustainability.^{1,4} Two major pharmaceutical companies agreed to provide funding over five-years to support activity – demonstrating confidence in the work of the Hub and the long-term use of its data. Academic research remains the main driver for collaboration with the Hub, which will leverage additional funding required to support long-term sustainability and provide evidence of the depth and richness of the data through publications.

INSIGHT has been approached by a range of SMEs and industry leaders, in the UK and globally, spanning both the technology and pharmaceutical industries, expressing interest in the Hub and seeking to explore potential areas of mutual benefit.^{1,5} For example, a UK tech SME is meeting regularly with the INSIGHT team to design an Oculomics project to detect cardiovascular risk through high-street optician scans. They are expanding the development of the Hub's sustainability model to incorporate alternative funding models to engage with SMEs, such as intellectual property or equity share arrangements. This recognition of segmentation of user requirements also reflects INSIGHT's strategy of supporting all sectors and achieving a balance of early, medium and longer-term value return under the Hub's forward plans towards Milestone 3.

PIONEER led 12 projects across NHS, academia and industry, including designing and implementing a fully-electronic COVID-19 screening and management system and real-time COVID-19 dashboard.^{1,7} The dashboard provided a fast-track screening tool for nationally prioritised clinical trials, supporting and reporting >10,000 recruitments for the NIHR Clinical Research Network. The PIONEER model is a scalable, interoperable system that can be federated to deliver high flexibility and agility both regionally, nationally and internationally. Sustainability for PIONEER is a mixed model of academic and commercial income, including partnership with the UK life-science sector. They have created an SME playbook, outlining how PIONEER can help and developed a flexible costing model which can include up-front costs, support in funding applications, feasibility testing or profit share/early adoption benefits.

BREATHE has enabled Hub services through developing datasets and phenomics, the latter initially focused on COVID-19 and now being expanded to other respiratory disorders. Creating curated national data assets has given BREATHE the opportunity to develop its Hub services and business model. For instance, working with a national e-cohort of 190,000 patients with asthma, BREATHE generated important whole country outputs on the association between socioeconomic deprivation and asthma care, outcomes and deaths. Having a major live national disease specific data asset has enabled the Hub to develop a rate card for services, which is now being road-tested with different potential data users. They have also been able to develop partnerships with new respiratory and data science industries and have worked with these partners on new revenue - raising

projects. Such partnerships and projects increase data access and utility, and opportunities for income generation, which is crucial to their longer-term sustainability.

Find more information about these examples in the Appendices.

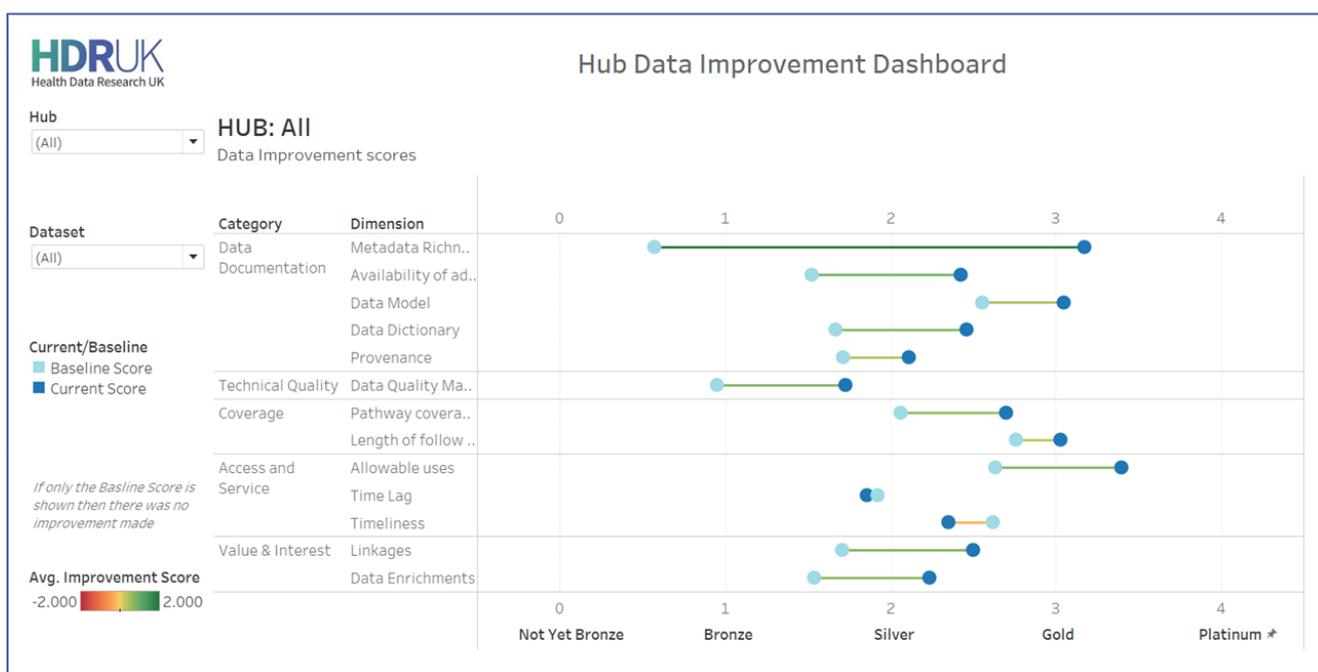
Improvements to Health Data

Collectively, the Hubs have made 157 datasets available via the [Hub Collections](#) on the Gateway, a portal that allows researchers and innovators to find and request access to datasets.

The Hubs were tasked with improving the data that they hold to enhance its research potential. To review the data improvement over the past 18 months, HDR UK created a new and innovative [data utility framework](#) to gather baseline information at Milestone 1 and objectively compare improvements at Milestone 2. The data utility framework is the first of its kind and was developed by HDR UK to review the Hubs' progress and to be made available for others to adopt and enhance the approach. It is based on input from users of health data as a means of understanding whether a dataset would meet specific research needs. The framework is composed of five categories, with multiple elements in each category, including documentation, technical quality, coverage, access & provision and value & interest. This is the first opportunity to use the framework not just for guiding users but for reviewing the improvements made to datasets.

There are measured improvement in data across all of the Hubs in various different data utility categories (see figure below). Of the categories in the utility framework, documentation saw the most consistent increase, driven primarily by improvements in metadata richness. The technical quality and value and interest areas also saw large increases, as Hubs developed data management plans and increased the linkage potential for their datasets. The category of coverage, relating to the representation of the clinical pathway and the length of follow-up did not increase as significantly, although many of the Hubs were already scoring highly in this area.

The Hubs took different approaches to improving data. While they focused on a broad range of improvements, there were some specific areas of attention. For example, INSIGHT focused on two datasets that were significantly improved, alongside creation of new datasets, whereas BREATHE focused on a much broader range of datasets, which saw more gradual improvements. Discover-NOW improved all of their datasets on coverage or timeliness, and PIONEER covered data management processes, documentation and linkage opportunities. Interestingly, there were examples where the timeliness of the data decreased, likely as a result of increasing demand and responding to the COVID-19 pandemic and more stringent standards.



Case study highlights

Within the first 18 months, **BREATHE** made respiratory data more discoverable through listing 50 diverse data assets on the Gateway and including >90 datasets in the BREATHE Collection.^{1.1} Through the citizen science [COVID-19 Symptom Study](#), in partnership with ZOE and Kings College, they urgently provisioned raw and curated data on COVID-19 symptoms to a range of users, which fed into local and national responses to the pandemic and now supports 54 live projects. BREATHE also created a national data infrastructure - [Eave II](#) - which led to a report of the [first national estimates on vaccine effectiveness](#) against COVID-19 hospitalisations; have undertaken urgent analyses at scale that have informed policy deliberations both nationally and in Europe; and created an environment that is enabling data agreements, wrangling, processing and analysis across six continents.

Discover-NOW has rapidly increased and improved its existing data – already the largest linked longitudinal health record in Europe – to meet the needs of the Hub’s wide multiple sector users responding to the global crisis. This is demonstrated across two main areas:

- Improving data to help prevent COVID-19 – such as through linking COVID-19 testing data with all haematology, biochemistry, immunology, microbiology and therapeutic monitoring results for the 2.5m patient records available in Discover-NOW
- Improving data to help manage the impact of COVID-19 – including working with hospital business-intelligence teams to link daily patient-level data on admissions, intensive care bed use, and high dependency bed use into Discover data, enabling forecasting of hospital and critical care demand and capacity for North West London.

As well as NHS and Public Health, SMEs are also using COVID-19 data to create performance improvements tools. Dr Foster, a leading data analytics SME and Hub partner, is building COVID-19 risk models to identify risk factors of death and a long length of stay in COVID patients. The Discover-NOW data is being used to inform national benchmarks for trusts to better understand their performance considering patient case-mix, and highlighting opportunities to improve patient care and outcomes.

PIONEER highlighted significant improvements to its datasets, informed by user need/requirements, increasing data uses across multiple sectors including academic, SMEs, NHS, Government, international pharmaceuticals, health technology and healthcare consultancy.^{1.7} The Hub has facilitated data collection and sharing across four regional NHS hospitals by designing and implementing a fully electronic COVID-19 screening and management system and real-time COVID dashboard. The dashboard provided a fast-track screening tool for nationally prioritised clinical trials, supporting and reporting >10,000 recruitments for the NIHR Clinical Research Network.

All of **INSIGHT’s** datasets have been updated and improved with the addition of a wide range of clinical metadata. Additionally, by adapting and improving cloud-based storage technology for raw clinical and imaging data, the lead-in-time necessary to generate new datasets has been reduced, thereby enabling the Hub to be agile in response to user requirements. This first bespoke dataset generated through INSIGHT supported the analysis of the impact of COVID-19 on patients with age-related macular degeneration (AMD), one of the leading causes of blindness.^{1.5} The project – initiated by Moorfields Eye Hospital and University Hospitals Birmingham with data science and analytical expertise from Roche-Genentech – provided the first reliable estimates of the scale and severity of the vision loss arising from delays in treatment for newly-diagnosed ‘wet AMD’ during the COVID-19 period, informing NHS (and industry) providers on strategies to optimise care of patients during service recovery.

Find more information about these in the Appendices.

Engaging Patients and the Public

Involvement of patients and the public, particularly minority voices such as people with disabilities and from the Black, Asian and Minority Ethnic (BAME) and lesbian, gay, bisexual, transgender, queer and intersex + (LGBTQI+) communities, has positively impacted on many of the Hubs' processes and outputs. From representation on executive and scientific committees, to citizen deliberation on access requests, and involvement in the shaping of research projects, PPIE is changing the culture of health data research by making it more accessible for everyone. As highlighted by **DATA-CAN**, this is not an 'add-on', but an integral part of the core business.^{3.2}

Public oversight and development of policies has led to significant changes in the requesting and access of data, how data research is communicated with the public, and how data are used. Having a key role at a project level ensures the Hubs are meeting public expectations, with patient and public representatives asking questions related to governance, privacy and data protection. For example, in a **BREATHE** project looking at how artificial intelligence (AI) is enabling public opinion to shape the pandemic response in Scotland, patient representatives asked probing data security questions and suggested techniques to protect individuals' privacy. This delivered real impact as the team were better prepared to address anticipated public concerns when presenting to Scottish Government.^{3.1}

INSIGHT's Data Trust Advisory Board developed a set of access criteria which achieves appropriate scrutiny whilst being feasible, efficient and scalable.^{3.5} **PIONEER's** Data Trust Committee co-created a PPIE best practice document to be shared with requestors, highlighting the benefits of meaningful PPIE to the data requestor, and signposting mechanisms to interact.^{3.7}

Many of the Hubs' PPIE teams have been instrumental in improving accessibility of information about data research, particularly by ensuring scientific summaries are now written in clear, understandable language. For example, **Gut Reaction** will only consider data access after their PPIE representatives have reviewed and understood the summary.^{3.4}

The Hubs' novel, transparent ways of working with patients and the public is helping to develop data trust, remove barriers to public engagement, and make balanced commercial decisions. As a result of feedback revealing that language was too scientific and issues around data privacy misconceptions, courses and workshops were developed to support PPIE teams to understand the use of health data. **INSIGHT's** Sense about Science public workshops explained their work with industry partners and openly explored concerns and addressed them.^{3.5} **NHS DigiTrial's** PPIE team supported the design and delivery of public engagement events, including seldom heard communities' practical recommendations for the identification and communications services, in particular.^{3.6}

Discover-NOW undertook one of the most progressive large-scale public deliberations in the UK about the use of health and care data, which has been endorsed nationally as best practice. Recommendations by the 100 public participants included 'five safes' in the establishment of their TRE, to amendments to their commercial framework, and transparency of all projects accessing data.

The **NHS DigiTrials** Co-Development Panel have made invaluable contributions to the development of the commercial model including providing key insight on the pricing model being developed, including views on the fairness of differential pricing for researchers from different sectors.^{3.6}

As a result of the PPIE contributions, additional benefits included seeing the initiatives being replicated by other organisations and identified as a factor in data requestors choosing to partner with the Hubs.

Appendices

This section is a repository of health data research case studies from across the seven Hubs, highlighting the number and diversity of health data uses. HDR UK aims to make these transparent and accessible to show the importance of health data research and stimulate new uses of data. This includes:

Data Improvement Case Studies: examples of how data improvements have generated opportunities for users and made an impact on Hub sustainability.

Data Utility Breakdowns: improvements made by the Hubs to datasets provided at Milestone 1, using the Data Utility Evaluation Framework.

Patient and Public Involvement and Engagement Case Studies: highlights from the Hub PPIE activities and how they have delivered real impact.

Patient and Public Involvement and Engagement plans: the Hubs' refreshed PPIE strategies.

Hub Projects: key projects the Hubs have been commissioned to deliver and how they have impacted on Hub services and sustainability.

Appendix 1 – Data Improvement Case Studies

Hub case studies focusing on how data improvements have generated opportunities for users across multiple sectors and the impact of the improvements to data value and Hub sustainability

1.1 BREATHE

Data Improvement Use Case 1 - How are improvements to the data increasing opportunities for Hub users? Do users of the data span multiple sectors?

Introduction

BREATHE's *raison d'être* is to improve respiratory health through enabling the safe and trustworthy use of respiratory health data at scale. BREATHE is working towards this vision by enhancing data utility through making the UK's outstanding respiratory data assets Findable and improving their utility through focusing attention on critical issues of data Accessibility, Interoperability and Reusability (FAIR).

At the time of writing, BREATHE has published metadata on 50 studies on the HDR UK Innovation Gateway with >90 datasets listed in the BREATHE Collection. These datasets span respiratory diseases (communicable and non-communicable), clinical settings (primary and secondary care), age groups (children and adults), sectors (public, industry and third sector) and geographies (four UK nations and extending to low- and middle-income countries (LMICs)).

Following discussions with HDR UK, BREATHE has focused activity over the last year on supporting the national pandemic response. COVID-19 is therefore a recurring theme throughout our case studies. That said, we have not lost sight of our wider mission to support efforts to improve access to and enhance the quality of respiratory data in general to increase opportunities for multiple Hub users.

Improving the ZOE Symptom Tracker app data

In March 2020, BREATHE was approached by HDR UK to facilitate data hosting for the ZOE COVID-19 Symptom Study. This comprised of data derived from the citizen science ZOE COVID-19 Symptom Tracker app and was, at the time, one of the only population-based (>4 million people) sources of data on the geographical distribution of COVID-19 across the UK. Working with our Trusted Research Environment (TRE) SAIL Databank, BREATHE was able to provide – at speed – a platform to receive, curate, and securely disseminate symptom data to diverse stakeholders. This involved creating a governance mechanism for managing data requests, as well as an analytical platform for provisioning data to a variety of data users.

BREATHE rapidly scaled technology and applied automated services on demand to deal with the daily ingress of data before curation and onward provisioning to multiple sector users, including central government analysts, government COVID-19 bodies (e.g. New and Emerging Respiratory Virus Threats (NERVTAG), Scientific Pandemic Influenza Group on Modelling (SPI-M), Chief Medical Officers COVID-19 Advisory Groups), NHS Trusts, local authorities, public health agencies, and academic groups across the UK. There are currently 54 live projects using ZOE data provisioned by BREATHE spanning multiple sectors.

BREATHE has also anonymously linked the ZOE data to nearly 100,000 participants in Wales. This has allowed parallel analysis of the ZOE symptom reporting with Welsh healthcare records in validating symptom reporting through to healthcare episodes and COVID-19 test results, which have been used to support modelling COVID-19 prevalence.

BREATHE's role in facilitating access and use of the ZOE data has led to creation of numerous COVID-19 heat maps and several academic outputs, including:

- [Current smoking and COVID-19 risk: results from a population symptom app in over 2.4 million people](#)
- [Real-time spatial health surveillance: Mapping the UK COVID-19 epidemic](#)
- [Age-dependent and Independent Symptoms and Comorbidities Predictive of COVID-19 Hospitalization](#)

Developing a national surveillance platform for COVID-19

Commissioned by the Scottish Government, BREATHE facilitated creation of EAVE II as a national COVID-19 surveillance platform covering ~99% (5.4 million) of the Scottish population. This has involved linking data from 940 general practices to testing, vaccination, hospitalisation, intensive care unit (ICU), and mortality data to create the world's only end-to-end national COVID-19 platform. EAVE II is in use for the epidemiology of COVID-19 informing nowcasting/ forecasting efforts, enabling risk stratification to identify those at increased risk of serious COVID-19 outcomes, and estimating vaccine effectiveness.

The epidemiological work regularly feeds into SPI-M and the Scottish Government's Chief Medical Officer's COVID-19 Advisory Group. A recent example of this was identifying increased risk of serious outcomes in those with the S gene dropout, which was recently presented to the Scottish Government's Resilience Room (SGoRR) chaired by the First Minister. The risk stratification work has been used to validate the QCOVID algorithm for the Scottish population, which BREATHE has made available to the Scottish Government to inform their deliberations on shielding and vaccine prioritisation.

Working with Public Health Scotland (PHS), BREATHE recently reported the first national estimates on vaccine effectiveness against COVID-19 hospitalisations for the Pfizer-BioNTech and Oxford/AstraZeneca vaccines. This was made possible by linking GP held data on vaccination status to the records held by PHS on vaccines administered in vaccination centres, thereby providing real-time estimates of vaccine coverage. This provided important national reassurance that the vaccination programme is having the desired impact of protecting the most vulnerable.

The analysis also showed that the vaccines were equally effective in older people in whom there were previously concerns about suboptimal vaccine responses. PHS has also undertaken the first vaccine stratified estimates in older people, with these results communicated in advance of their release to the Scientific Advisory Group for Emergencies (SAGE) co-chairs, governments and the World Health Organization (WHO). This has already led to major policy changes in relation to the Oxford-AstraZeneca vaccine being licensed in older people in France, Denmark and Germany.

Building on the BREATHE-enabled EAVE II vaccine effectiveness work, urgent related work is ongoing in England, Northern Ireland, and Wales working closely with the respective national public health agencies. This work is feeding into the [COVID-19 National Core Studies](#) urgent programme of work. Data and Connectivity Vaccines Pharmacovigilance (DaC-VaP) brings together vaccine data linked to routinely collected data from general practices, hospitals, death registries, and laboratory tests in the four UK nations. Proactive sharing of code lists, statistical analysis plans, analytic code, data, and outputs are occurring through the BREATHE Hub, thus facilitating an efficient approach to vaccine analysis across the UK.

International COVID-19 Alliance

BREATHE is represented at the Executive Leadership Team of the International COVID-19 Data Alliance (ICODA). Convened by HDR UK, ICODA has a particular focus on improving access to high quality data from lower and middle income countries (LMICs). BREATHE, in partnership with its TRE SAIL, is providing platform hosting services for ICODA, including technical and governance support for the programme. ICODA has worked with the International Perinatal Outcomes in the Pandemic (IPOP) consortium to engage with 86 data contributors from

40 countries to investigate perinatal outcomes during the pandemic. Importantly, this infrastructure is providing opportunities, for the first-time, for teams of analysts from different countries to simultaneously investigate the impact of COVID-19 lockdowns and disruption to health services on perinatal outcomes across six continents.

Cystic Fibrosis Trust Partnership

BREATHE has partnered with the Cystic Fibrosis Trust (CFT) to provide UK-wide data linkage for their entire CFT registry data of >12,000 people (representing ~99% of people with CF in the UK). BREATHE will shortly begin facilitating consented UK-wide linkage and sharing of this unique, strategically important dataset. We are currently planning a joint BREATHE-CFT roundtable meeting to showcase the dataset to industry colleagues and elicit industry input into how we could further develop the dataset and associated opportunities for Hub users.

UK-wide data curation

Linking to work undertaken by various BREATHE sites in advancing clinical phenomics (Imperial, Leicester), BREATHE has initiated further work on UK-wide data curation to increase the utility and research efficacy of data found within the NHS. Standardising coding and linkability of datasets are key to creating 'respiratory-ready' datasets within GP records and hospital episode data, seeking optimisation of data for combinatory and comparator analyses. By providing standard sets of phenotypes across key respiratory conditions, BREATHE is taking steps to provide opportunities for the use of electronic health record (EHR) data to enable high quality reproducible health research by diverse Hub users.

Such work is, for example, occurring within the aforementioned DaC-VaP study. An integral part of our data utility work is the BREATHE curation and use of data for drug discovery taking place in our industry collaborations. The relevant studies are collaborations looking at genomic analysis of chronic obstructive pulmonary disease (COPD) risk and severity; collaboration in the genetics of smoking cessation; phenome-wide association studies; interstitial lung disease genomics; and PhD training collaborations.

Summary

Within the first 18 months of existence, BREATHE has made respiratory data much more discoverable through listing 50 diverse data assets on the HDR Innovation Gateway and including >90 datasets in the BREATHE Collection. Our case studies illustrate how we have urgently provisioned raw and curated data on COVID-19 symptoms to a range of users across multiple sectors, created a new national data infrastructure, undertaken urgent analyses at scale that have informed policy deliberations both nationally and in Europe, and created an environment that is enabling data agreements, wrangling, processing and analysis across six continents. Our work has been fundamental to the national data-enabled COVID-19 response. As we emerge from the pandemic, we hope to accelerate data-enabled responses to other important respiratory disorders.

Data Improvement Use Case 2 - *How has the data been used to enable Hub services, income generation opportunities and develop Hub sustainability?*

BREATHE has generated several successful examples of deploying Hub services, using existing data assets and through clinical and technical expertise, to help data providers, researchers and innovators from multiple sectors to advance and accelerate their aims. Focusing on data assets which have wide appeal, BREATHE has been able to deploy its initial iteration of a rate card for service charges both in relation to research grant applications and services offered to industry.

Wales Asthma Observatory

Creating curated national data assets offers the opportunity to develop our Hub services and business model. The Wales Asthma Observatory is a national e-cohort of patients with a history of asthma derived from

routine primary and secondary care records. This open cohort contains ~190,000 patients with currently treated asthma. BREATHE has enabled the operational research lead for the Wales Asthma Observatory to generate important whole country outputs such as [this paper](#) investigating the association between socioeconomic deprivation and asthma care, outcomes and deaths in Wales. This dataset is continually updated, and parallel methodological work is ongoing to investigate the opportunities offered by Natural Language Processing (NLP) to interrogate free text records and further enrich the core dataset. Having a major live national disease specific data asset has enabled us to develop our rate card for services, which is now being road-tested with different potential data users.

The continued maintenance of this unique national data asset – for one of the commonest chronic disorders in the UK – is a key focus for BREATHE to establish use cases for academic groups, NHS and public health agencies, industry and charities who may wish to commission projects in population-representative data. We are also seeking external funding to enable reuse of the data, fund further development and insights, and to mature our rate card for analytical, clinical, and operational services whilst maintaining integrity of data governance for public benefit.

Clinical Phenomics

In enabling Hub services, BREATHE has brought together a set of phenotypes covering a number of key respiratory conditions. This is not intended as an exhaustive list of codes or mandatory, rather a guide to aid researchers in formulating appropriate code lists for their analysis. By creating such a resource and encouraging publication of code lists, we also hope to make respiratory research more reproducible. This work has been achieved through working closely with the HDR UK phenomics implementation project, with said phenotypes and code lists being uploaded to the Phenomics Portal. The respiratory aspects of this work began with codes for COVID-19 and is now being extended to a number of other respiratory conditions.

Salford Lung Study

As part of our goal to form strategic data partnerships, BREATHE recently forged relationships with a multinational pharmaceutical company as lead for its Salford Lung Study (SLS) dataset, a landmark real-world evidence trial, in order to help make this prize dataset available to the wider respiratory science community. This is, as far as we are aware, the first industry dataset to have its meta-data listed on the Gateway. We are now working with the company on approaches to operationalising access to this dataset. Furthermore, we hope to be able to enrich the dataset through enabling data linkage thereby further increasing the utility of this dataset to multiple users. For example, by working alongside another BREATHE partner, BREATHE will provide access to and support for electronic health record data linked to SLS.

Going forward, we hope to build on this success to make additional industry datasets discoverable through the Gateway and available via BREATHE. We are, for example, currently in discussion with a drug discovery company in relation to its interferon beta trial for COVID-19.

Chronic obstructive sleep apnoea

Building on our acknowledged expertise in respiratory and health data science, BREATHE has set up a project with a medical equipment company who are investigating clinical care pathways and accompanying health economic implications of the care of patients with chronic obstructive sleep apnoea (OSA).

Initially coming to BREATHE with a generic proposal for data access and analysis, BREATHE has worked with the company to shape their proposal into a more streamlined set of questions that are best answered through a phased approach. These involve an initial project to scope OSA coding and patient presence within national

datasets within Wales, before proceeding to a full healthcare pathway analysis involving data from Wales and Scotland. BREATHE provided recommendations as to the feasibility of data availability and is supporting analysis to provide the company with key data to inform their strategic plans.

As an industry applicant, BREATHE has been able to use an early iteration of its rate card, to charge appropriately for the resources put towards this project – with the project serving as the first project to drive income generation for both scientific and data services for the Hub. As this is a collaborative project, BREATHE is also facilitating patient and public engagement for the company in order to ensure the project is undertaken with the support of public and patient groups. BREATHE and the company additionally plan to have a joint publication strategy.

Predicting COVID-19 severity with AI

BREATHE has partnered with a medical Natural Language Processing (NLP) software provider and industry partner, on a COVID-19 initiative. This is a multicentre, non-interventional study using data captured in EHRs of participating hospital sites. Thus far the project has recruited 15 sites with due diligence being undertaken on a further 32 sites expressing an interest.

Using the software provider's technology, we will develop a model that predicts risk of deterioration and COVID-19 severity using Artificial Intelligence (NLP techniques applied to anonymised EHRs). This unique approach incorporates all EHR data from a hospital (e.g. inpatient, outpatient, emergency room, pathology, etc) regardless of the management system in which they operate. This allows previously inaccessible free text, in combination and linked with standard reporting methodology, to provide a much richer picture of individuals care and outcome than was previously possible.

BREATHE will provide information on patients who may be eligible for specific treatments or follow up who may otherwise be missed and monitor outcomes in specific subgroups of patients (e.g. black and minority ethnic groups). The project directly links to BREATHE's data utility mission in terms of improving access to rich data sources for research use and to support an industry partner to work in a UK data environment to deploy novel technology. The software provider is funding the project with direct cash payments to participating partners to cover participation costs.

Creating synthetic assets

BREATHE is committed to working with our industry partners on specific projects and making use of their expertise. BREATHE has recently partnered with an AI software provider and will make use of their software, provided to BREATHE in-kind as part of their partner contribution.

BREATHE will create synthetic data assets based on real data structures and content, permitting scientific feasibility and training to be deployed in a more flexible manner whilst very substantially reducing the risk of inadvertent disclosure of protected data. Whilst governed projects are commissioned for technical teams to map the software to real data, such as mapping to the Wales Asthma Observatory, the software creates new datasets based on the structure and format of real data.

We see particular value in initially focusing on generating data assets in three main areas: 1) synthetic health records 2) synthetic longitudinal cohort data; and 3) synthetic clinical trials data. The creation of such synthetic data assets will facilitate significant opportunities to open data access to far wider sectors of researchers than is otherwise possible, as well as providing BREATHE with an option to add these as part of our sustainability model moving forwards.

Summary

This case study illustrates how BREATHE has enabled Hub services through developing datasets and phenomics, the latter initially focused on COVID-19 and now being expanded to other respiratory disorders. We have used our data infrastructure to develop an initial Hub services rate card, which is being iterated on with partners and through early experiences. We have been able to develop partnerships with new respiratory and data science industries and have worked with these partners on new revenue raising projects. The value of the exemplar projects total in excess of £1m in actual and in-kind contributions to BREATHE. Such partnerships and projects increase data access and utility, and opportunities for income generation, which is crucial to the longer-term sustainability of BREATHE.

1.2 DATA-CAN

Data Improvement Use Case 1 - *How are improvements to the data increasing opportunities for Hub users? Do users of the data span multiple sectors?*

Summary

COVID-19 has brought into sharp focus the need for timely intelligence to inform urgent decision-making, when dealing with a virus that spreads rapidly through a population and places huge demands on health services. The UK has a highly regarded public health system, particularly in cancer, with long established disease registries and health data reporting. However, it became apparent that the system was unable to determine the direct and indirect impacts of COVID-19 on cancer patients in a timeframe to underpin rapid action. This was primarily due to the lag time between health data collection, curation and being made available. DATA-CAN identified this gap and provided real-time data from NHS trusts to demonstrate to physicians, policy makers and governments the significant impact that COVID-19 was having on cancer services UK-wide.

This case study highlights how real-time data is deployed to provide crucial intelligence that informs action, underpinning our aspiration to establish a Real-time Data Network (RTDN) that catalyses the use of data to inform better outcomes for cancer patients.

COVID-19 real-time data research project

During the first lockdown (March-May 2020), DATA-CAN and UCL's Institute for Health Informatics worked with UK cancer centres to collect and analyse real-time hospital cancer service data. They looked at two measures to delineate the pandemic's impact – '2 Week Wait' (2WW)/ 'red flag' referrals from GPs for people with suspected cancers (a reliable measure of the pandemic's impact on diagnostic pathways) and chemotherapy appointments (a reliable measure of impact on therapeutic pathways). Bringing these data together improved their quality, creating new aggregated datasets that indicated, for the first time in the UK, the impact of COVID-19 on cancer services.

Curation and analysis of the data found that, compared to pre-COVID-19 levels, there were significant drops in urgent referrals (70% decrease) and chemotherapy treatment attendances (40% decrease).

They employed these results to inform different modelling scenarios of the impact of COVID-19 on excess death in people with cancer, in a cohort of ~3.8 million patients with linked primary and secondary care data (the CALIBER dataset).

Their expertise in curating and analysing these data on 24 cancers and 15 co-morbidities enhanced the quality of this dataset and permitted, for the first time, prediction of the magnitude of the pandemic's impact on excess cancer deaths.

The analysis suggested that the COVID-19 pandemic could result in 7,165 more deaths in people with newly diagnosed cancer. This could rise to 17,910 additional deaths if all people currently living with cancer were considered, with 78% of excess deaths occurring in patients with >1 comorbidity.

From a data improvement perspective, their expertise and research activities helped create new aggregated datasets that did not exist before (2WW and chemotherapy attendance datasets), enhance existing datasets (CALIBER dataset) and deploy these improved datasets to precisely delineate for the first time the impact of COVID-19 on cancer services and cancer patients in the UK.

Impact of the research

Their rapidly assembled [MedRxiv pre-print](#) was shared with the UK's four chief medical officers, the National Clinical Director for Cancer for England and SAGE, contributing to decisions to restart cancer services. The pre-print received >22,000 reads with >80 citations (Google Scholar). The full research paper is available in [BMJ Open](#). Significant media coverage resulted with over 400 stories in print and broadcast media, with significant impact on public opinion.

DATA-CAN findings were quoted in the Health and Social Care Select Committee report '[Delivering core NHS and care services during the pandemic and beyond](#)' (Oct-20) and during a Westminster Hall debate.

Internationally, study data were presented to the Board of the European Cancer Organisation (ECO) and to the 2020 Session of World Health Organisation Europe. DATA-CAN's scientific lead was invited to co-lead ECO's Special Network on COVID-19 and cancer. This Network launched a '[7-Point Plan to Build Back Better](#)', which received significant international attention. The data were also presented to the European Beating Cancer Committee in the European Parliament and will front a pan-European COVID-19/cancer campaign.

DATA-CAN has co-authored a series of papers on COVID-19 research published in [Gut](#), [Lancet Oncology](#), [European Journal of Cancer](#) and [Lancet Gastroenterol Hepatol](#) during 2020/21. DATA-CAN data was extensively quoted in research by [CF Consulting](#) and the Institute for Public Policy Research.

Their research was recognised by Health Data Research UK as 'Highly Commended' for the Impact of the Year 2020 award.

They presented data to the Clinical Research Coalition chaired by Baroness Blackwood, which made a series of recommendations to Life Sciences Minister Lord Bethell and resulted in this [white paper](#), which highlighted how real-world evidence (RWE) can be used for the benefit of patients, recognising that initiatives like DATA-CAN can help the UK to leverage RWE, become a world leader and set best practice globally.

This project, the enhancement of the datasets achieved and its national and international impact, has highlighted to DATA-CAN the absolute primacy of near real-time data, underpinning the development of a Real-time Data Network (RTDN) for cancer data. It is also helping engender a culture of enhanced data sharing within NHS trusts.

Current work: COVID-19 and real-time data

DATA-CAN's Real-time Data Network (RTDN)

The impact and importance of this rapid real-world data led the team to develop DATA-CAN's RTDN, extending to other NHS organisations UK-wide. NHS trusts will share an agreed minimum non-identifiable aggregate weekly dataset of pre-existing national data items at minimum-agreed quality. These data will be available to the trusts for service review and research, making it attractive for hospital trusts to be RTDN members and gain new insights from the aggregated data. There will be an option to share deeper datasets dependent on members' digital maturity, patient choice, information governance and organisational priorities.

During the second wave of COVID-19, DATA-CAN is collaborating with the cancer alliances of North Central London, West Yorkshire and Harrogate and the South West to contribute real-time data to monitor impacts on cancer services and patients. This significantly increases the number of NHS trusts contributing to the project and extends the reach/influence of real-time findings into cancer service networks. Additionally, contributions from Wales, Scotland and Northern Ireland (where all five hospital trusts participated in the initial study) make this a true UK-wide collaboration.

Members of DATA-CAN's PPIE group have visibility of all aspects of the Network and a central role in governance.

DATA-CAN has engaged the Yorkshire and Humber Care Record cloud-based platform, (YHCR's 'Data Ark'), to safely and securely host real-time data from NHS trusts.

There has been significant interest in the RTDN from the Life Sciences industry, with £65,000 in medical educational grants from Roche Pharmaceuticals and Janssen Pharmaceuticals to support their work and many further enquiries.

DATA-CAN is also now working with Roche UK and Deloitte to consult with service managers and design a live publicly available dashboard of data from the RTDN for May 2021 to help with cancer service monitoring and recovery. The work is funded by Roche and Deloitte are contributing in-kind resources.

Conclusion

Their work has been impactful in multiple sectors. For the NHS, it highlighted for the first time the impact of COVID-19 on cancer services and demonstrated how different trusts could work together to aggregate data for clinical and patient-focussed insight. For industry, it has gained significant support, with the Association of British Pharmaceutical Industries highlighting it as an example of how UK datasets and data research can underpin relevant insights to industry. DATA-CAN researchers have delivered keynote addresses at industry-focussed fora and contributed significantly to white papers/reports. From a policy perspective, it helped underpin restoration of cancer services. From a patient/public perspective it was the lightning rod that first highlighted the negative impact of COVID-19 on citizens/patients. From an international perspective, it underpins ECO's European COVID-19 campaign with impacts at European Parliament level.

The efficient delivery of accurate near real-time cancer intelligence has highlighted the challenges that COVID-19 raises for cancer services and patients UK-wide and has helped inform solutions to these challenges. DATA-CAN's ability to capture and link different datasets, through their RTDN, safely, efficiently and responsibly, allows critical intelligence to be rapidly gained, underpinning research to improve cancer services and enhance patient outcomes, both now and post-COVID-19. Displaying data through a public-facing dashboard affords us the opportunity to build trust with patients and the public through transparency.

1.3 DISCOVER-NOW

Data Improvement Use Case 1 - *How are improvements to the data increasing opportunities for Hub users? Do users of the data span multiple sectors?*

Discover-NOW has pivoted data improvement efforts to provide more opportunities for all sectors to utilise the power of data in their fight to understand and overcome COVID-19. We have been able to rapidly increase and improve the existing data, already the largest linked longitudinal health record in Europe, to meet the needs of our wide multiple sector users responding to the global crisis. This is demonstrated across two main areas below.

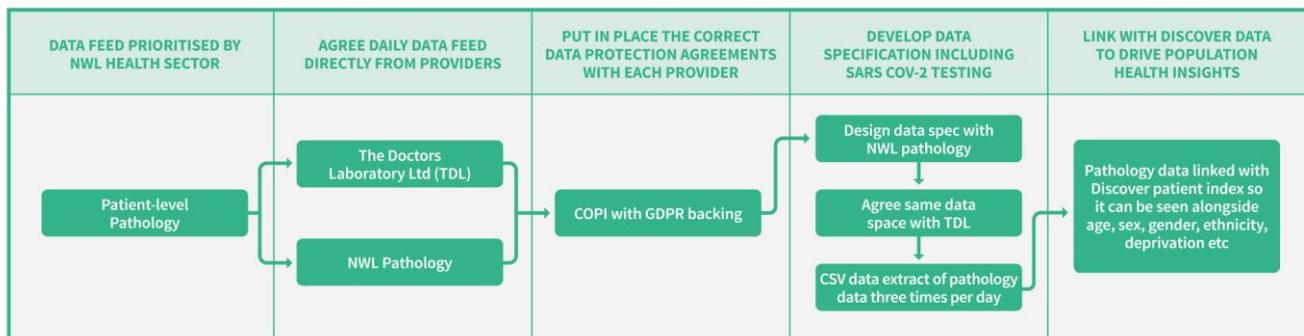
1. Improving our data to help prevent COVID-19

1. Linking COVID-19 Testing and Pathology Data

While working alongside NHS, Local Authority teams and Hub users, it became apparent that understanding the spread of COVID-19 at a granular level would be key to targeting interventions, preventing local breakouts and keeping vulnerable people safe. It was also clear that the ability to track infection rates would be essential for developing and rolling out targeted vaccination plans and proving their efficacy in the future.

The Discover-NOW team worked at pace to safely integrate patient-level laboratory data (see Figure 1) to enable infections to be tracked and to predict COVID-19 healthcare demands. The data sourced contains all lab pathology data including Pillar I COVID-19 test results and all haematology, biochemistry, immunology, microbiology and therapeutic monitoring results for the 2.5m patient records available in Discover-NOW. This data improvement significantly increased the novel research opportunities previously not feasible due to the complexities of collecting so many clinical measures.

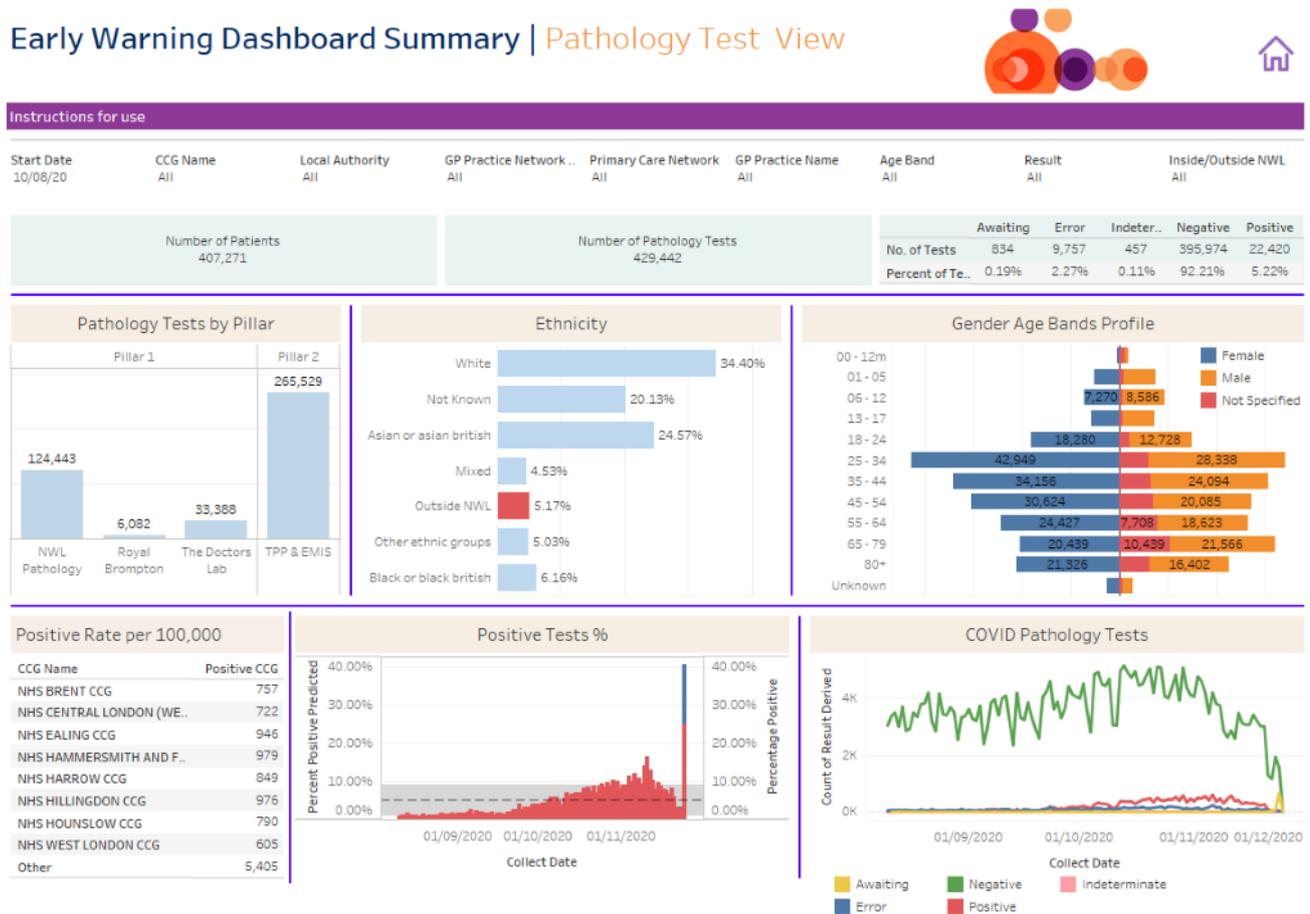
Figure 1: Integrating Patient-Level Laboratory Data into the Discover Data



1.2 New Opportunities for Users from COVID-19 Testing and Pathology Data

Working with Hub users such as NHS colleagues in the COVID-19 Response Command (Commissioners, Clinicians and Trust Executives), and Public Health specialists, the Discover-NOW team used COVID-19 testing data to create a COVID-19 Early Warning Radar (see Figure 2). This suite of tools helped to identify local outbreaks, identify patients at higher risk of contracting COVID-19 (e.g. those in densely populated households) and to target risk-reducing interventions.

Figure 2: COVID-19 Early Warning Radar Dashboard Example



Discover-NOW was also able to improve the timeliness of Primary Care data (the foundation of data linkage) from a two-week lag to two days. This was achieved through moving from using a third-party primary care extractor to sourcing data from the emerging One London Local Health and Care Record Exemplar (LHCRE), a partner of Discover-NOW. A more near-time view of what is happening is now possible and interventions can be put in place more quickly.

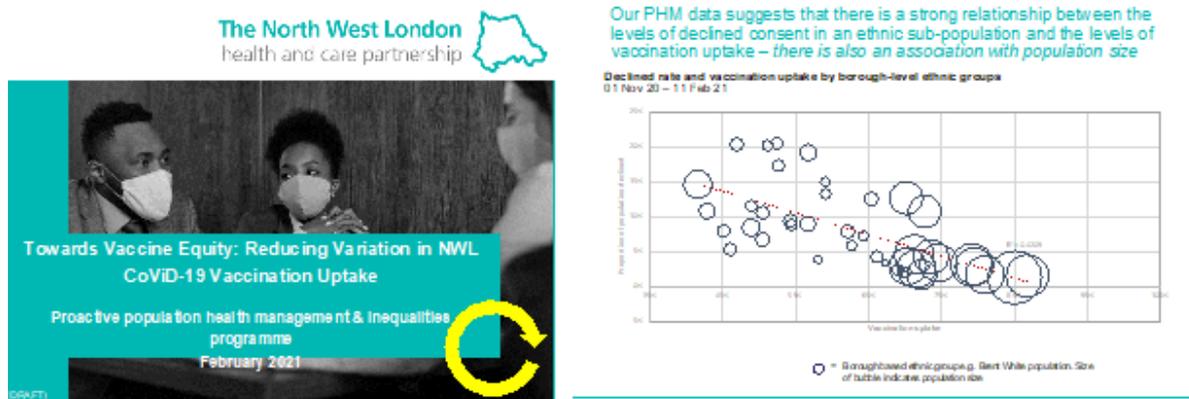
As well as NHS and Public Health, SMEs are also using COVID-19 data to create performance improvements tools. Dr Foster, a leading data analytics SME and Hub partner, is building COVID-19 risk models to identify risk

factors of death and a long length of stay in COVID patients. The Discover data is being used to create national benchmarks for Trusts to better understand their performance considering patient case-mix, and highlighting opportunities to improve patient care and outcomes.

1.3 Opportunities and progress with linked COVID-19 vaccination data

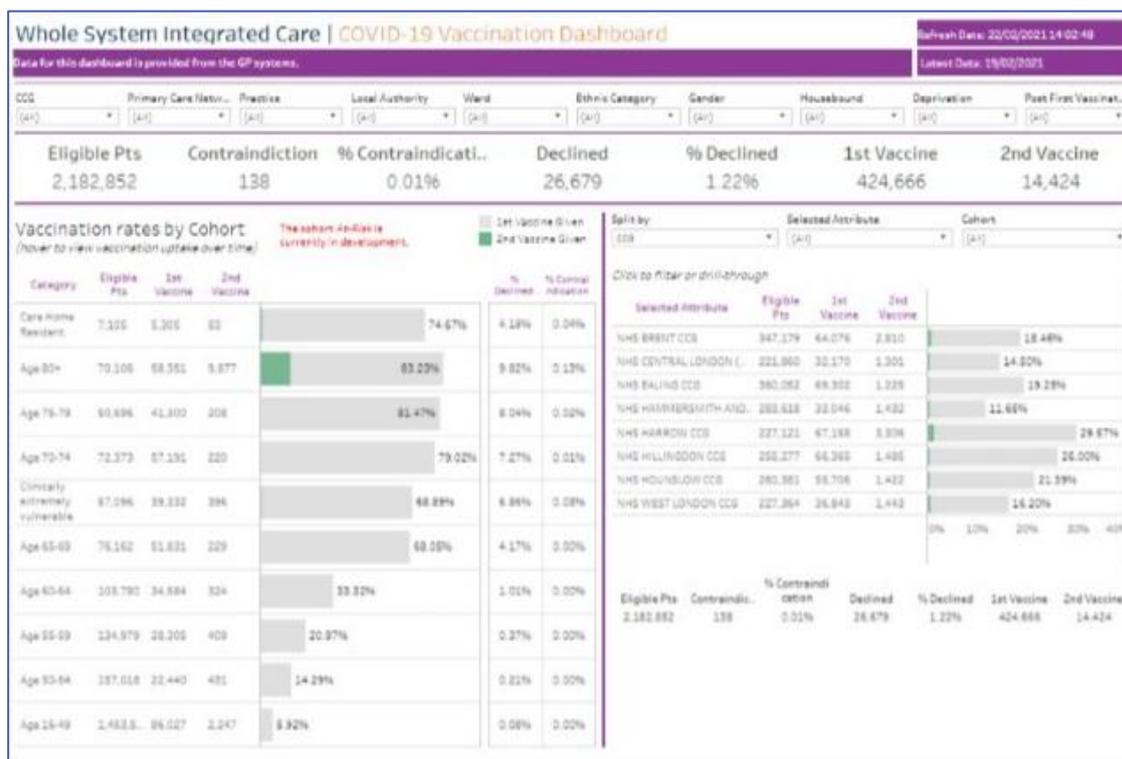
Discover data has also been improved to link with data that includes details of people vaccinated, by vaccine, allowing breakdown by age, gender, ethnicity, indices of deprivation. For local healthcare and public health users, this data linkage has enabled a novel and detailed view of the vaccine rollout ensuring data insight driven resource management (see Figure 3). This has ensured that vaccines are getting to the NWL populations that need them the most, tracked against the Joint Committee on Vaccination and Immunisation Priority Lists for vaccination to ensure those at risk are protected.

Figure 3: Example of Dashboard Displaying the Linked Vaccine and Discover Data



For commissioners and health care professionals, the linked data has also enabled the increased understanding and management of inequalities in vaccination take-up using a population health management approach. The Discover-NOW team achieved this through the development of a COVID-19 vaccination dashboard (see Figure 4) for use across the health care system to monitor population take up and to target populations at risk.

Figure 4: Example of COVID-19 Vaccination Dashboard



1.4 New opportunities for users from COVID-19 vaccination data: Industry and public-private partnerships

In addition to population health management benefits, we have also identified the opportunity for global vaccine manufacturers to use Discover-NOW data to monitor the safety and efficacy of their vaccines in the real world.

Discover-NOW is in discussion with a pharmaceutical company to understand their COVID-19 vaccine effectiveness. With a linked longitudinal record, individuals can be retrospectively monitored according to COVID-19 vaccine. Discover-NOW is able to quantify vaccine effectiveness based on rates of infection, hospitalisation, major side effects and mortality as captured in real-world healthcare settings. The Discover-NOW service model enables a rapid study start-up, combined with agile data curation and timely data at scale; a service unmatched elsewhere and perfect for urgent and deep real world data research.

The improved data has also enabled Discover-NOW to engage with a European public-private partnership (COVIDRIVE) to collaborate on vaccine studies. Discover-NOW has been selected as one of five pan-European sites to analyse vaccine effectiveness by looking at vaccinations against hospital admissions. This would support COVIDRIVE to conduct two pan-European studies to be submitted to the EMA as evidence for regulatory approval.

2. Improving data to help manage the impact of COVID-19

As well as improving data to help prevent COVID-19, Discover-NOW has also been working hard to improve data to help clinicians, operational managers, academics, researchers, and industry to work together to better understand and mitigate the impact of COVID-19.

2.1 Improvements to admissions and bed use data for COVID-19

The Discover-NOW team worked with hospital Business-Intelligence teams to link daily patient-level data on admissions, ICU bed use, and HDU bed use into Discover data, enabling forecasting of hospital and critical care demand and capacity for NWL. The NWL ICS now relies on this data to plan hospital capacity and patients transfers across the system. In the future this data will also be used for economic analysis and pandemic research.

2.2 Enabling near patient testing and remote monitoring capability and data

The pandemic placed a huge pressure on hospital capacity resulting in the rapid uptake of remote monitoring solutions to help keep patients safe at home. This saw Discover-NOW repurpose established work with Hub partner, AstraZeneca, on integrating innovative remote monitoring and near-patient testing into existing pathways for patients with Type 2 Diabetes, to be used to support NWL COVID 'Hot Hubs'.

AstraZeneca, through working with NHSX and SME Huma, also a Hub Partner, enabled oxygen saturation and breathlessness scores of patients with mild to moderate COVID-19 to be quickly and efficiently remotely monitored by GPs in the 'Hot Hubs'. The remote monitoring data captured within Huma is now flowing (initially via CSV extracts ahead of the API development) into Discover data, allowing these new cutting-edge pathways to be evaluated for safety and effectiveness.

2.3 Improving and curating existing data to generate a shielded patient list

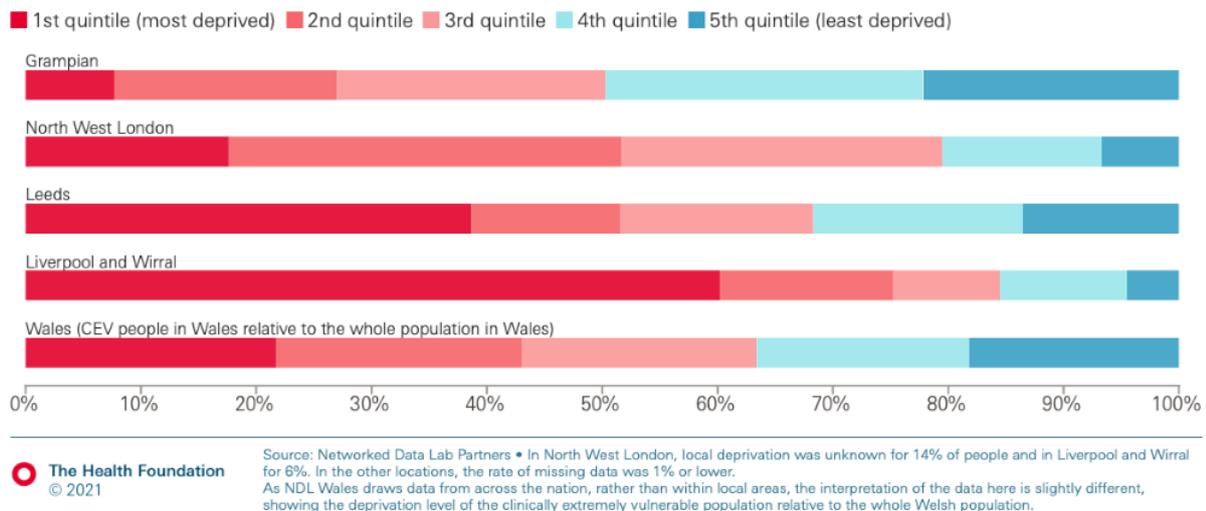
Discover-NOW through linking NHS Digital shielded-patient criteria to Discover data has been able to achieve a richer linked integrated care record than ever before possible. This linkage has enabled NWL to identify and flag those at-risk of worse outcomes from COVID-19, and ensure appropriate measures are put in place to reduce risk. Discover data is ready to be used in future research into the impact of the shielded patient processes to inform any further waves or pandemics.

This data improvement also enabled the opportunity for Discover-NOW to work with *The Health Foundation's [Networked Data Labs](#)* to contribute to analysis of demographic variation in those advised to shield; as published by The Health Foundation (See Figure 5).

Figure 5: Example of The Health Foundation Analysis of Demographic Variation

There is substantial variation across regions in the proportion of the clinically extremely vulnerable population living in areas of high deprivation

Percentage of clinically extremely vulnerable people in each quintile of the Index of Multiple Deprivation (English, Scottish or Welsh 2019 IMD)



2.4 Improved coding and templates to understand and manage post COVID-19 syndrome

Working with primary and secondary care, the Discover-NOW team are implementing standard templates into Electronic Health Records in all care settings in order to better monitor, manage, and research the emerging post COVID-19 syndrome. By using clinically agreed SNOMED coded templates, data quality flowing back to Discover data is improved, enabling better service planning, research, and evaluation of interventions for Hub users.

2.5 Enabling research opportunity through improvement to data access

To make the most of opportunities for users across all data improvements, we know it is essential that data can be safely accessed and used within an appropriate timescale, especially in the fast-paced world of COVID-19.

Discover-NOW has enabled timely review of research opportunities to access the rapidly improving dataset during the crisis, through supporting the establishment of the [COVID-19 NWL Data Prioritisation Group](#). This group met weekly to review operational and research requests for COVID-19 related issues to access the Discover data and to date approving over 30 research applications from a range of sectors including academic, SMEs, national bodies (ONS) and the NHS.

Reflections and looking forward

Discover-NOW has made a number of significant improvements to the Discover data in response to the COVID-19 pandemic that have been pivotal to support the local health and care response.

By applying the data improvement process to clinical, operational, and population health management data, Discover-NOW has generated a research data set that is pioneering new research and understanding of COVID-19 for all uses.

Data Improvement Use Case 2 - How has the data been used to enable Hub services, income generation opportunities and develop Hub sustainability?

Discover-NOW has built a strong service offer with a firm track record and is creating a unique testbed approach for innovative data and digital services. This is enabling more innovative work and significantly increasing Hub income seen through our partnership with AstraZeneca. We are investing in a TRE with Hub partner, IBM that will underpin Hub sustainability plans.

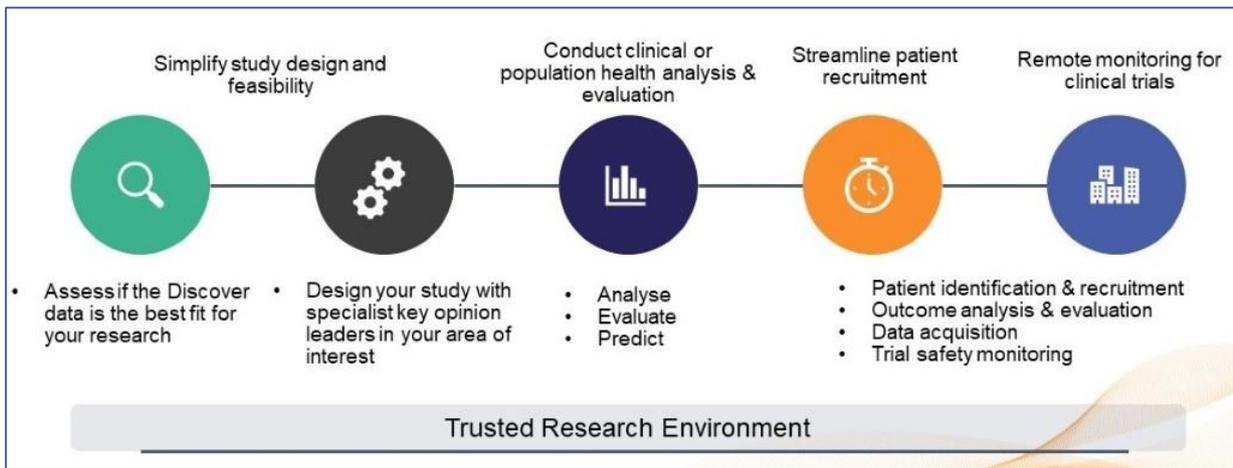
Discover-NOW service offer

Our service offer has been developed and refined in collaboration and through engagement with our stakeholders across industry, large, SME, NHS and academia on what they aim to achieve with Real World Evidence and how the outputs can support the health system. This insight alongside our PPIE has allowed us to continually improve the quality and agility of our approach to service development.

Our offer is spread across high demand research areas and is under-pinned through the implementation of a high-performance TRE (see Figure 1).

Examples of the studies enabled through our offer and the Discover curated data are shared publicly across our [research tracker](#) and [COVID-19 research tracker](#). A subset of 20 projects underway are highlighted in the KPI submission for this Milestone.

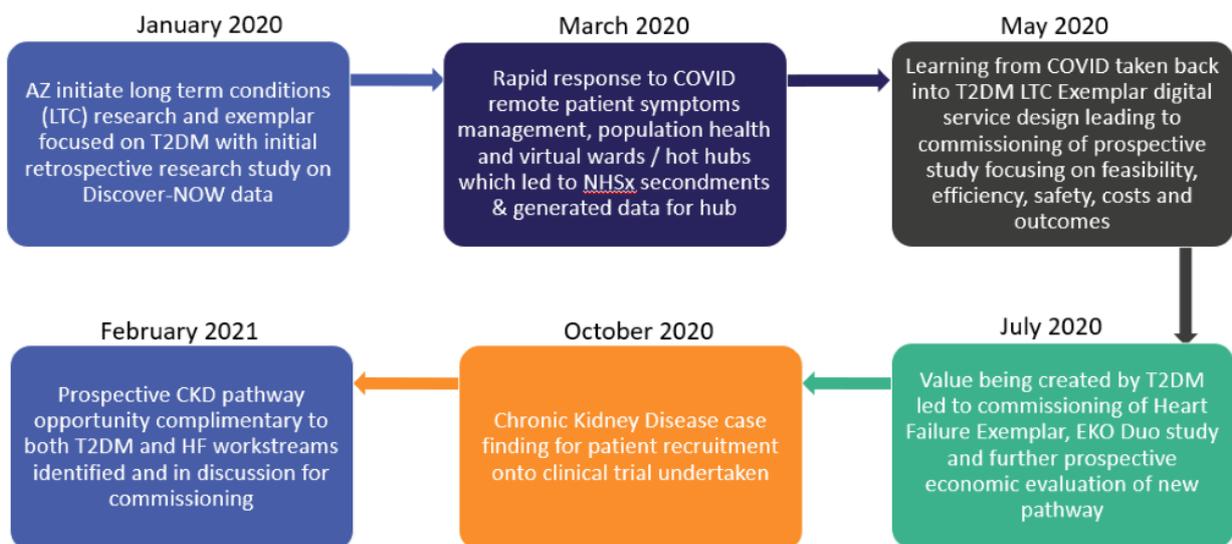
Figure 1: Summary of the Discover-NOW service offer



Data enabling a digital test-bed service and income generation

COVID-19 challenged Discover-NOW and AstraZeneca with the opportunity to provide and deploy new digital and data enabled services which had not previously been implemented at such scale. With our service offer at our foundation, we addressed this challenge through the development, deployment and evaluation of a digital testbed approach. This has since been scaled to other projects and disease areas to bring new income to the Hub and boost our sustainability plans (see Figure 2).

Figure 2: Demonstrating the Timeline for the Digital Test-Bed Service Development



During COVID-19 NHSX reached out for support with the national response for COVID-19 remote monitoring. This was as a direct result of our collaboration with AstraZeneca in building digital remote monitoring solutions

for patients with T2DM and service offer expertise. Team members seconded to the response and were able to support clinical outcomes through expertise and use of our Discover high-quality near-time data.

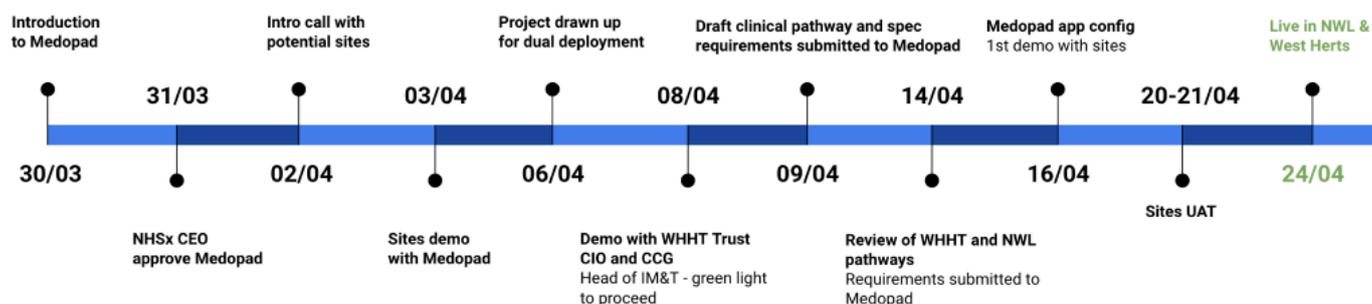
Working alongside clinicians, we were able to digitally enhance clinical practice through innovative remote monitoring and near-patient testing service. This remote monitoring setup by Hub partner, HUMA, was fast, flexible, and simple to use, and the data generated from the remote monitoring platform is linked into the Discover data enhancing our service offer.

Figure 3 shows the fast set up of the service, within two weeks from launch, at the peak of COVID-19 in London, the service had:

- Over 150 patients on the virtual ward
- 96% of patients using the service as prescribed
- Technology and implementation appeared inclusive; oldest user, 80 years old
- In secondary care, the service increased secondary care efficiency by approximately 500%, based on HCPs making over 150 calls to patients per day, reducing to 30 calls per day after introduction

This is understood to be the fastest ever implementation of a digital service in the NHS; start to go-live in under four weeks. Several thousand patients have to date have been discharged from the pathway. [Channel 4 recently reported on the service.](#)

Figure 3: Implementation timeline



This digital innovation testbed model is increasing the value of the Discover-NOW service offer through its use in further commissioned projects and across multiple disease areas and this includes income generating studies with AstraZeneca.

Example 1: Astra Zeneca, Huma, and T2DM

A retrospective analysis of pathways, resource use and outcomes in T2DM is underway using the Discover data and in collaboration with Imperial College London and Imperial College Health Trust.

AstraZeneca partnered with Imperial College London and the NHS in NWL to improve T2DM care pathways and assess how new technologies could be used to improve care for high-risk patients. The Discover-NOW team have developed a new model of personalised, remote care allowed high risk patients to be identified and treated 33 months earlier.

Discover-NOW is also collaborating with ICHP to undertake an economic and outcomes evaluation of this new T2DM digitally enhanced pathway, this involves recruitment of patients from the NWL Health Research Register.

This example shows the unique digital and data innovation test bed service offer in development: from service baseline, to pathway redesign, pilot and evaluation.

Example 2: Astra Zeneca, Eko Duo, and Heart Failure (HF)

This work has enabled research and evaluation into other diseases and treatments such as new HF digital interventions and pathways through collaborating with AstraZeneca, Imperial College London, Eko Duo, and the NHS.

The study aims to baseline treatment pathways and quantification of the care delivered for HF and compare to national predefined standards. The outcomes of this study will directly impact the NHS by identifying variation, good practice, and targeting opportunities to improve care.

Next, we are undertaking an economic analysis on a medical device product for HF patient pathway improvement, Eko Duo. Eko Duo should improve accuracy of HF diagnosis and improve pathway efficiency. Discover-NOW is working with the NHS to design and implement this tool and use the data generated to evaluate impact.

Discover-NOW is now developing a new exemplar programme in Chronic Kidney Disease with AstraZeneca using this digital and data innovation test bed service offer.

Example 3: BMS and AF

Taking this testbed for digital innovation approach, Discover-NOW can also help [to train and test AI and machine learning algorithms.](#)

The Bristol Myers Squibb-Pfizer Alliance has created an algorithm to predict who may be at risk of developing atrial fibrillation (AF) using CPRD, a national primary care data set.

Once an algorithm has been trained, it needs to be tested in a different environment to make sure it still works. Discover-NOW has provided this agile testing environment, applying the same AF algorithm to the Discover data. The Discover scale and breadth is unique, providing an ideal algorithm training and testing environment, this retrospective study has since been published in the European Journal of Preventative Cardiology (Sekelj, 2020).

The Alliance is now undertaking an economic evaluation of the algorithm as a further phase of the project with Discover-NOW.

Example 4: Amgen and the High Impact Intervention Tool for CVD

Our approach has also enabled the creation of a High Impact Intervention Tool with Amgen.

Featured in The [Guardian newspaper](#), our collaboration is creating a tool to quantify care gaps in terms of detection, treatment and management of patients with high cholesterol levels. The tool can be used to target high impact interventions such as lifestyle interventions and reviewing lipid levels. It will use predictive modelling using a leading enterprise AI technology embedded into the Discover data, DataRobot, to establish the impact of addressing these gaps on reduction of CVD events and costs.

Example 5: Medical Device Company and CVD

The flexibility of our approach is such that cohorts can easily be defined and followed up, our service made even more powerful through Discover data Consent to Contact flags of the NWL Health Research Register (an investment by Discover-NOW).

A medical device start-up has used this service to recruit hypertensive patients for a clinical trial. An initial feasibility assessment was undertaken and Discover-NOW made it possible to recruit and follow-up patients for the trial.

Using the representative pre-consented population, the client has been able to quickly identify and on-board eligible patients, a target they struggled to reach prior to working with Discover-NOW.

Example 6: Hypercholesterolaemia and CVD

A pharma company is using the Discover-NOW testbed service to create machine learning algorithms and build a robust evidence-base for change, specifically for cardiovascular disease and hypercholesterolaemia pathways.

By identifying patient cohorts most at risk and amenable to interventions, clinicians and commissioners will be able to better target resources to reduce cholesterol levels supporting NHS England Long Term Plan aims of saving lives by reducing heart attacks and strokes.

Forward look and reflections

We are actively making strategic investments across our service offer to secure additional long-term revenue streams. The next year will involve developing the digital and data innovation testbed approach into a viable and repeatable service offer for other clients alongside continuing to deliver our core service offer.

Sustainability requires better data and a data environment that further enables our service offer, as such, Discover-NOW is investing in a TRE and will have a Minimal Viable Product in Spring 2021. The TRE will be developed through agile sprints led by user needs and informed by our PPIE including our Citizen Advisory Group (see PPIE case studies) guiding the technical and access controls to be put in place to enable non-NHS users direct TRE access.

In collaboration with the NHS and North West EHealth (NWEH) a Hub partner, Discover-NOW has now implemented patient recruitment technology, FARSITE, across 360 GP practices with over 200 practices already activated. FARSITE enables the streamlining of feasibility and patient recruitment which we are developing as part of our service offer in collaboration with NWEH and with large clinical research organisations.

Discover-NOW is on track to continue to build on our services and income generation to reach the goal of being sustainable, returning maximum benefit back to the healthcare system and ultimately improving patient lives.

1.4 GUT REACTION

Data Improvement Use Case 1 - *How are improvements to the data increasing opportunities for Hub users? Do users of the data span multiple sectors?*

Demonstrating utility from 'Improving the data'

Inflammatory Bowel Disease (IBD) includes Crohn's disease and ulcerative colitis. Together, these immune-mediated conditions affect around 500,000 people in the UK and are characterised by chronic debilitating abdominal symptoms that can have a major adverse impact on health, employment, relationships, and quality of life. Treatment increasingly focuses on 'biologic' antibody therapies, but individual responses vary substantially for poorly understood reasons.

IBD and COVID-19

Conventional immunosuppressant and newer biologic therapies are helpful in the treatment of IBD because, as the name suggests, they suppress the body's immune system to stop it attacking the gut. Adults prescribed these drugs for IBD are deemed 'clinically extremely vulnerable to COVID-19'.

There is a time-critical need for patients and clinicians to understand their level of risk with a view to shielding, and to assess vaccine efficacy for patients taking immunosuppressant therapies.

As an immediate response, in just eight days in March 2020, the IBD Registry developed a [COVID-19 IBD Risk Tool](#) based on the British Society of Gastroenterology (BSG) Risk Grid to allow patients to self-assess their risk.

Liz Dobson, CEO of the IBD Registry presented the tool to the [HDR UK Alliance Symposium in December 2020](#), winning the Lightning Talk award in recognition of its impact and benefit for patients.

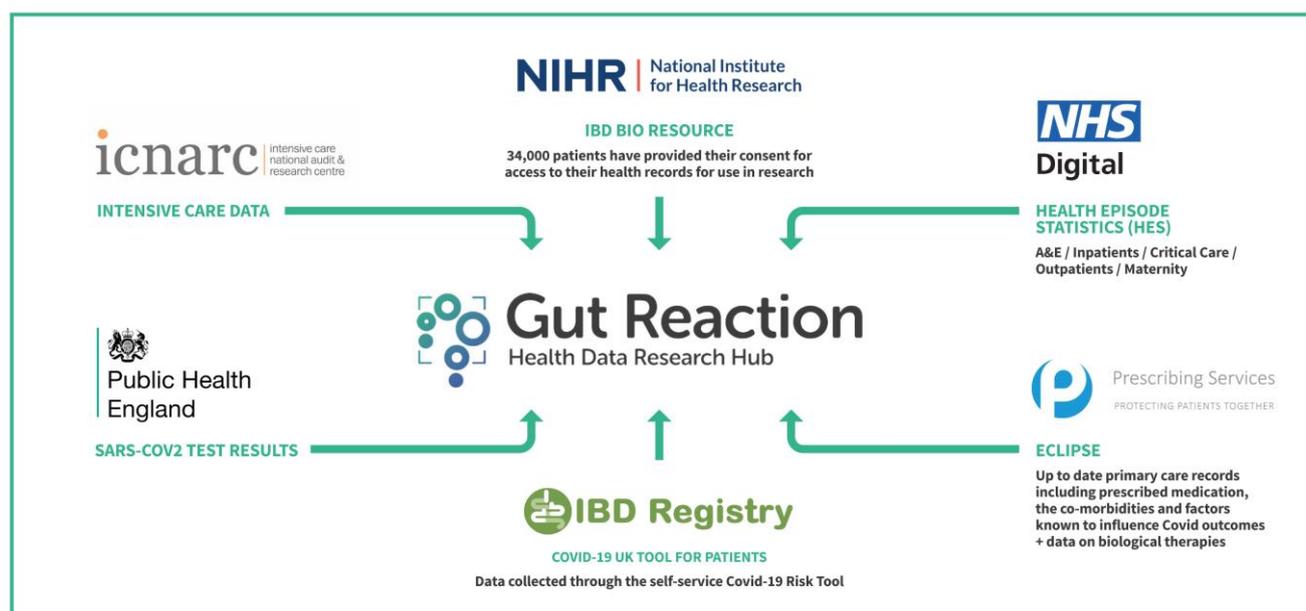
Bringing new datasets to the Hub

To further understand the risks of acute covid infection, and future vaccine efficacy for patients, the Gut Reaction hub has brought together multiple datasets (Figure 1, below) and made the integrated data available within a secure Trustworthy Research Environment (TRE) hosted at AIMES.

Clinical phenotype (including secondary care prescription) and demographic data from IBD BioResource have been loaded into the TRE alongside COVID-19 test result data from Public Health England (PHE) and Intensive Care National Audit & Research Centre (ICNARC) data, both obtained under the Control of Patient Information (COPI) notice supporting information sharing to manage the spread and impact of the current coronavirus pandemic.

A separate application to NHS Digital for Hospital Episode Statistics (HES) and Office for National Statistics (ONS) mortality datasets was approved for Covid-related studies. These datasets have been linked at record level in the TRE.

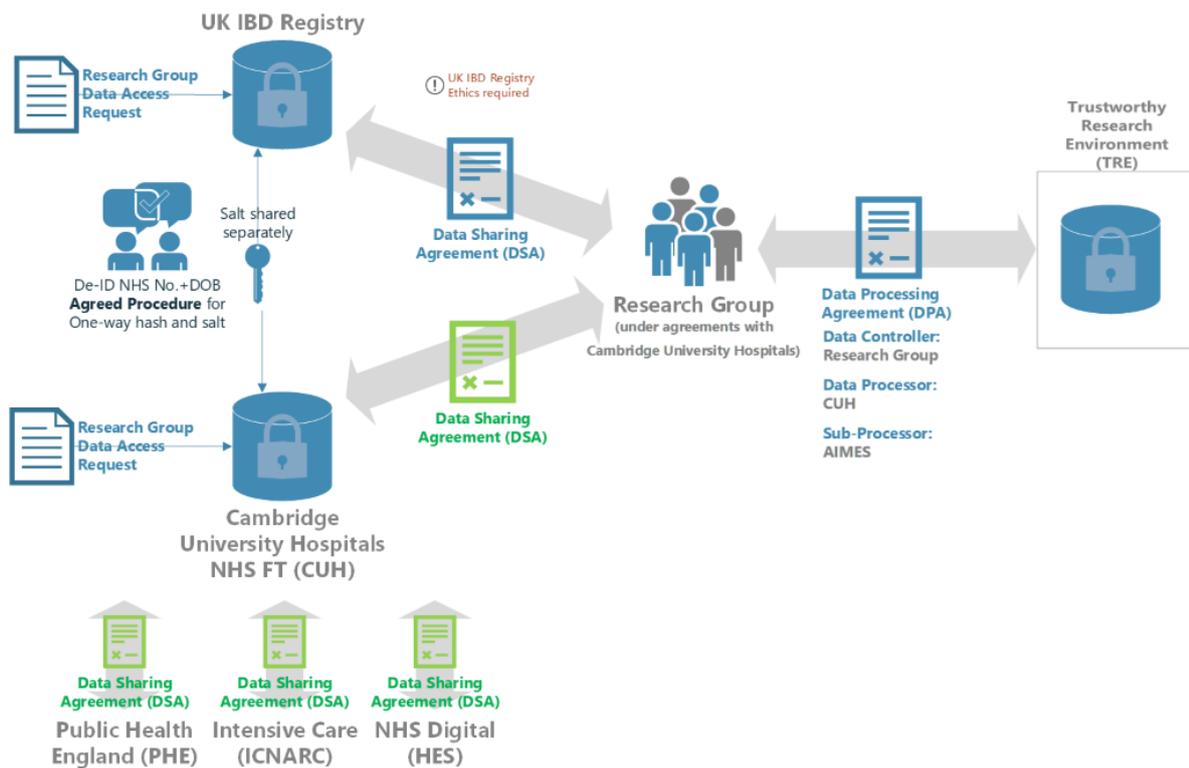
Figure 1: Understanding COVID-19 outcomes for patients on immunosuppressant therapies for IBD (Data Sources)



The above datasets are already being used in the TRE with the exception of the UK IBD Registry Covid-19 Risk Tool and primary care prescribing datasets which will be added soon.

The governance arrangements allowing these datasets to be integrated provides a useful and replicable model for bringing additional data (under different data controllers) to the hub for linkage and analysis alongside Gut Reaction datasets (Figure 2 below).

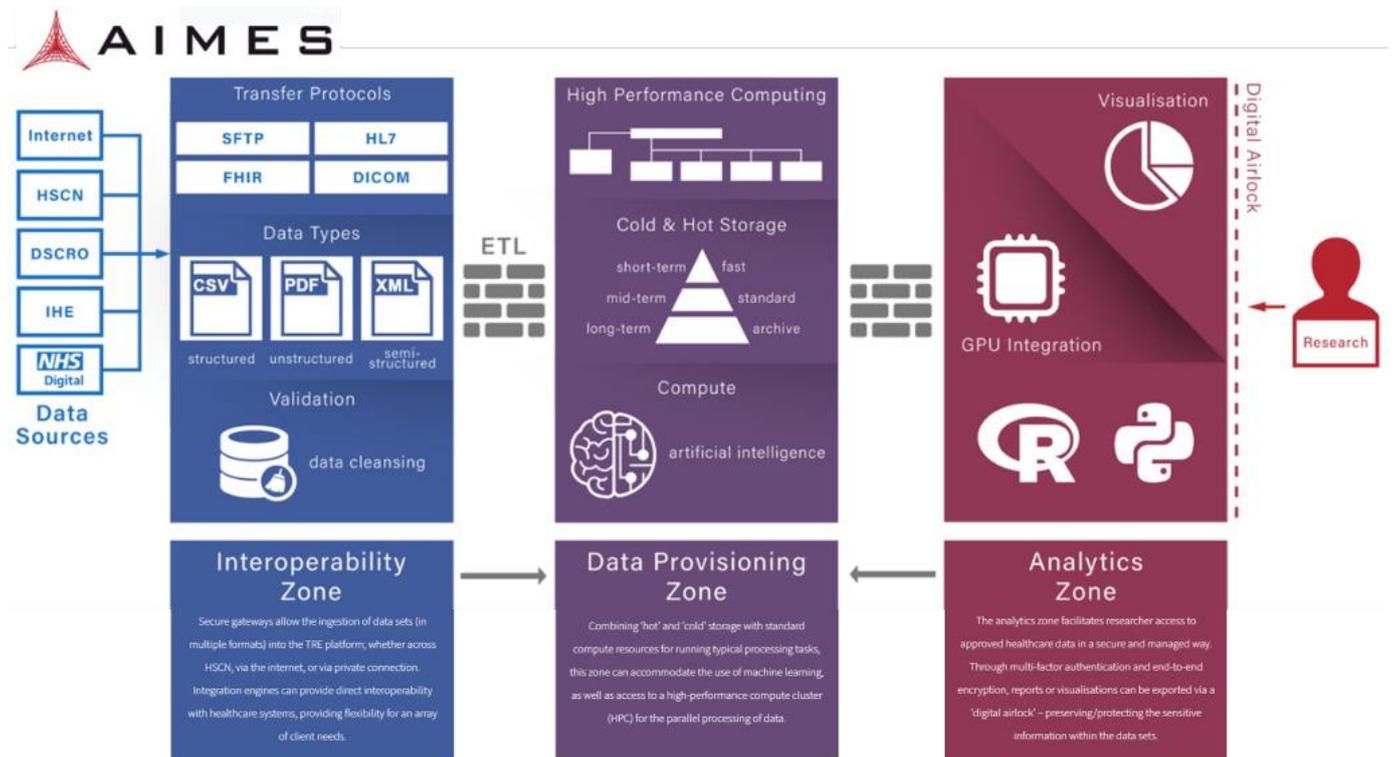
Figure 2: Information Governance arrangements to support integration of the IBD Registry COVID-19 Risk Tool data



Trustworthy Research Environment (TRE)

Data is accessed within the AIMES TRE, providing researchers with a secure environment to analyse approved data using an agreed (customisable and security checked) tool set. All activity within the TRE is recorded and fully auditable, and a digital 'airlock' means outputs or reports/findings may be exported from the TRE, but the underlying data cannot.

Figure 3: AIMES Trustworthy Research Environment (TRE) Framework



1.1.1 Findings & feedback... A researcher's perspective

“Using IBD BioResource clinical data and PHE, ICNARC and CHES [COVID-19 Hospitalisation in England Surveillance System] datasets integrated in Gut Reaction we are investigating if immunosuppressant therapies used in IBD increase COVID19 infection rates and adverse outcomes (hospital / ITU admission or death).

In Feb 2021 we analysed data from >1,600 IBD patients with positive COVID tests, and 7,000 paired COVID19 negative patients. Current results suggest that IBD therapies minimally impact risk of infection. Moreover, hospitalisation and death rates in IBD patients are equivalent to the rates reported in the age-matched UK population. If confirmed in the pending IBD Registry analysis, we will hasten to make these unequivocally ‘good news’ results available to the worried patient community and their clinicians.

We have observed incomplete overlap of death records between PHE, ICNARC and CHES, suggesting missing records - hence are contacting our BioResource sites to validate hospital admission and death data. The addition of new datasets (ICNARC and CHES) ran smoothly and used the same ID scheme, suggesting linkage to other data sources will be straightforward.

All datasets are analysed within the AIMES trusted research environment (TRE).

Observations:

- TRE onboarding by AIMES included basic training (software and authentication setup, basic environment, airlock).
- Requested software (Python, Jupyter Notebooks environment, statistical analysis packages)
- AIMES set up Gitlab to support collaboration.
- the TRE works efficiently but current work is not computationally intensive. Future projects using genomics data will test the AIMES HPC in a separate Linux software environment

Overall, I came to this project worried that the TRE would make analysis work overly cumbersome. This was not the case. I completely appreciate that the project required accessing several databases with sensitive patient records, mandating high levels of security and assurance. We got access within 4 months of the project start, which would normally be impossible. If doing such analyses requires dealing with TRE constraints, it is a small inconvenience.”

Alex Sazonovs (Postdoctoral fellow, Sanger Institute)

Benefits

Data Improvement and increased opportunities for hub users

Since commencement of the Gut Reaction Hub programme,

- Data from NHS Trusts is now flowing: Despite the significant challenges of coping with COVID, nine of the ten Trusts engaged have approved the Information Governance (IG) framework, sharing agreements are in place. Eight Trusts have provided at least some data to the hub including new types of data (e.g. digital pathology images) – a major achievement in the circumstances, and a clear demonstration of the value of the hub to collaborating Trusts and research leads.
- Complete datasets have now been received for just over 50% of these participants
- Increased participation: NIHR BioResource has increased recruitment of IBD participants from ~30,000 to 34,600
- New sources and modalities of data secured (HES, ICNARC, PHE, NHS Trusts) with more to follow soon

- **Curation**

- Additional datasets from Hub partners (IBD Registry, Wellcome Sanger Institute) listed on the HDR UK Innovation Gateway
- Additional linkages have been established and flagged on the Gateway for these new Datasets (ICNARC, HES & PHE)
- Datasets and linkages that were previously unavailable to researchers are now being actively used.

- **Trustworthy Research Environment (TRE)**

- AIMES TRE is now established and being used

- **Models for replication and scale**

- The governance model developed to bring data (under different data controllers) to the hub provides a valuable blueprint for future use.

- **Data Discovery**

- Dataset Discovery – the number of Gut Reaction datasets on the HDR UK Innovation Gateway has been increased from four to eleven, with automated processes established to load and update metadata. Data utility ‘medallion scores’ have also improved significantly.
- Cohort Discovery – Cohort discovery tool is now available for demographic and phenotypic datasets.

These data would not have been brought together, linked and made available for analysis without the Gut Reaction Health Data Research Hub programme.

Next steps

1.1.2 Interest from SMEs to large international organisations across IT, MedTech & Pharma

Several exciting possibilities are now emerging, generating significant interest from industry. There has been a marked increase in commercial organisations applying for data access in the last few months and several companies are now developing research questions that would not have been feasible previously.

The planned additional data resources, particularly primary care prescribing, will unlock further exciting opportunities for novel research, subject to individual Clinical Commissioning Group (CCG) agreements. Janssen and Takeda have expressed interest in these capabilities.

High Performance Computing (HPC) will mean the capability to analyse linked genomic data in the AIMES TRE. Data to support investigation of the relationship between Smoking and IBD that includes genomic (whole exome sequence (WES)) data has been loaded in readiness for this capability going live at AIMES in March. This study will be undertaken by Dr Carl Anderson and his team at the Wellcome Sanger Institute.

LARGE PHARMA JANSSEN IMMUNOLOGY



“Gut Reaction Hub provides an invaluable resource for conducting Digital Health (DH) research in gastrointestinal diseases. Specifically, the number of subjects, the wealth of genetic data and connection to patient NHS data, enables the identification of patient subjects of interest based on genotypes or phenotypes. Furthermore, the ability to recall patients and enrol them in downstream DH studies facilitates the identification of digital signals for no-invasive and remote patient stratification and disease monitoring.”

DR YANNIS PANDIS
Digital Health Science & Technology Lead
Janssen Immunology

LARGE PHARMA TAKEDA - VEDOLIZUMAB



“The power of the Gut Reaction Health Data Research Hub programme will only grow as links are established to outcome data (both from HES and the IBD Registry), pathology data, imaging reports and outpatient / discharge letters. It is encouraging to see that the proposed review and approval pathway for studies proposed by pharmaceutical companies has already been established and includes a tight timelines. At this stage Takeda UK are working with the Hub team to establish the feasibility of using this dataset to answer a range of clinically important questions related to our IBD drug, vedolizumab.”

DR SIMON MEADOWCROFT
Medical Director
Takeda UK & Ireland

Service Development: Specialist data scientists are now in post to develop services to support data analysis, curation and annotation using Machine Learning (ML) and Artificial Intelligence (AI) with diagnostic images and alongside Natural Language Processing (NLP) for unstructured text. This is generating interest from data science SMEs such as Pangea AI, and in MedTech with Ludger and Avenna.

SME SPECIALISING IN MACHINE LEARNING & AI PANGAEA AI



“...availability of such data through the Gut Reaction hub will open opportunities for us (Pangaea) to collaborate and realise the aforementioned benefits in the context of gut related diseases. This will also help us achieve our core mission to improve patient outcomes while attracting more research investment to the UK through a sustainable business model.”

DR VIBHOR GUPTA
Founder
Pangaea Data

MEDTECH - 'PROBLEMS' ADDRESSED FOR LUDGER & AVENNA
POOR BIOMARKERS FOR CHRONIC INFLAMMATION (CL)
AND IMMUNOFRAILTY (IF)



“Our proposed project using Gut Reaction Hub resources is to conduct detailed N-of-1 lifecourse studies on causal mechanisms of immunofrailty in individuals living with IBD. This would include IBD patients who have or will contract COVID-19. The idea is to map changes in their blood glycomics patterns that could be potential prodromal markers of future alterations to their state of immunofrailty, to update our general IF Model and further refine our models for prognosis and prediction of inflammatory events in IBD patients.”

DR DARYL L FERNANDES
Chief Executive, Ludger Ltd
Chief Science Officer, Avenna Ltd

Data Improvement Use Case 2 - How has the data been used to enable Hub services, income generation opportunities and develop Hub sustainability?

Introduction – working towards a sustainable model

Achieving sustainability for Gut Reaction, the Health Data Research Hub for Inflammatory Bowel Disease (IBD) will be dependent upon meeting health data research needs across the NHS, academia and industry in a way that is supported and delivers value to patients, the public, the NHS and Social Care.

The natural tendency for any stakeholder involved in a transaction is to consider “What’s in it for me? How will I, or my family, friends or organisation benefit most?” An organisation generating or controlling data may take

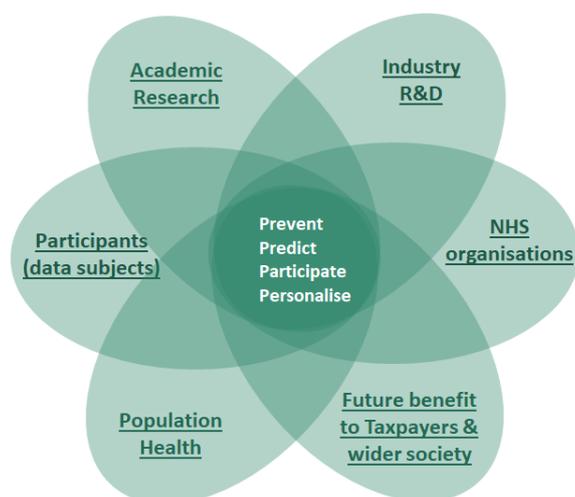
the view that best deal for its data is securing the highest value the market will bear. A commercial organisation accessing data may want to take away and use as much as it can at the lowest cost without restrictions on how it is used. Attitudes of a person who is generally well, may be different from someone with a long-term illness.

‘Sustainability’ means meeting our own needs without compromising the ability of those that follow to do the same. Keeping the benefits for patients and society at the core of what we do is essential, and maximising those benefits means working in partnership between the NHS, and researchers, charities and industry in way that is fair, transparent, and open.



Value proposition

Balancing the needs of key stakeholders



Our core offer

to Academic & Industry Researchers

- Dataset discovery via HDR UK National Innovation Gateway
- Transparent & efficient data access application process via the NIHR BioResource
- Service development:
 - Sourcing and developing new & relevant data resources
 - Increasing participation from currently under-represented groups
 - Cohort discovery self-service tool
 - Trustworthy Research Environment (TRE)
 - Bespoke data curation/transformation
 - Machine learning and AI capabilities including Natural Language Processing (NLP) & Imaging analysis
- Transparent pricing model

Industry Engagement

Working with ABPI to develop the Gut Reaction ‘Researcher Engagement Pack’

In May 2020 we started working with the team at The Association of the British Pharmaceutical Industry (ABPI) to inform development of an information pack being prepared for researchers interested in accessing linked de-identified data for patients with IBD. The document set out the remit of the Gut Reaction Hub, its relationship to the NIHR BioResource, UK IBD Registry and other partners, as well as the data sources, types of data available both now and planned, the application process for accessing data and crucially, setting out a model for data access within a secure cloud Trustworthy Research Environment (TRE), rather than data being handed over to take away and analyse elsewhere.

Key messages conveyed in the Researcher Information Pack:

- Access to data is contingent upon a clear articulation of patient/public benefit
- Only data required to answer the specific approved research question(s) will be provided
- Data access will be via the TRE

- The principle applies that any downstream commercialisation of intellectual property (IP) generated from insights enabled by data provided should see a fair share of any revenue or cost savings generated being returned to the hub and participating NHS organisations for re-investment in frontline services

Feedback received from ABPI Members

The engagement from ABPI members was overwhelmingly positive. At a follow-up event in October 2020 participants were generous with their time and feedback.

“Our members have greatly appreciated being involved in the development and opportunity to comment on the Researcher Information Pack. They believe that collaboration with industry will significantly advance understanding of IBD and contribute towards the long-term goal of improving available treatment options as well as the quality of life of IBD patients.

“... members felt that it is a useful document, setting out helpful context and answers to most of the key initial questions companies would have, including around the integration with other resources such as the IBD Registry and IBD Bioresource. This will help industry to better understand and plan for future collaborations with Gut Reaction and the NIHR IBD BioResource.”

Carolyn Heaney | NHS Engagement Policy Partner,

The Association of the British Pharmaceutical Industry

ABPI members were supportive of the data access model described and understood that data sharing was more likely to be supported and endorsed by patients where researchers were given access to only the data required to answer approved questions within the TRE.

Given the data is not openly available to take away and ‘mine’ to inform research opportunities, it was clear that researchers need a different way to be able to access sufficient detail about the data available to help shape research questions, data access applications and identify viable study cohorts.

Dataset discovery → Cohort discovery

The HDR UK Innovation Gateway facilitates discovery of datasets, tools, papers and other health data research resources from across the UK. There are eleven Gut Reaction Datasets available on the Gateway so far. Metadata provided contains information about each dataset, fields available, levels of data completeness, overall volumes, and how to apply for access.

Whilst it is useful to know that a dataset exists and what it contains, it is not possible at this stage to query the data to find out how many records might satisfy a particular set of conditions.

Users are invited to make a data access application to the relevant organisation to be able to investigate further.

Cohort discovery tool

The next step considered locally has been the implementation of software that enables datasets to be interrogated to ascertain how many records meet a particular set of criteria without accessing the underlying (record level) data.

The tool selected is ‘Informatics for Integrating Biology at the Bedside’ or i2b2 which has been configured as a self-service tool to enable researchers to query de-identified/aggregated patient data gathered from multiple sources.

Two datasets have been configured and loaded, so far:

- Case Report Forms (CRFs), phenotypic data collated and recorded by the NIHR BioResource Specialist Research Nurses (SpRNs) when patients are enrolled in the BioResource, and
- Lifestyle questionnaire data provided by the patients themselves.
- The key steps in this process have been:
 - Over a thousand data items have been rationalised to a manageable aggregated level
 - Metadata has then been catalogued and mapped to ontology codes
 - Data has been configured for automated upload to the tool

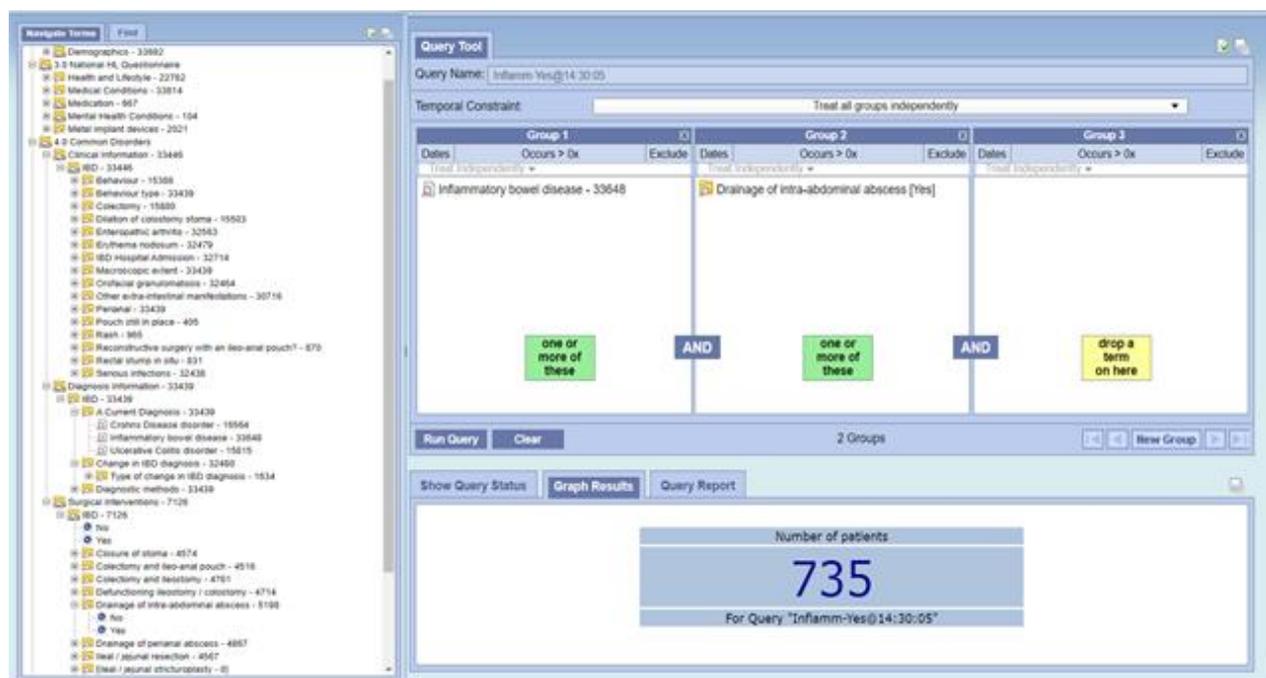
The i2b2 tool has been made available in a standalone Protected Data Domain (PDD) within the AIMES TRE, and a data scientist from GSK has supported a review of the tool to test access arrangements and utility for cohort discovery in practice.

An important feature of using this tool as configured within our TRE is that once a cohort of patients has been selected, the prospective researcher will simply see the number of records that meet their selected criteria.

Figure 2 below is a screenshot of the Gut Reaction cohort builder in use, to illustrate the fields available, the query builder, and outputs as a 'cohort count'.

Using the 'airlock' functionality of the TRE means that the hub data management team can export patient identifiers for the exact cohort selected without the underlying data ever being exposed to the user.

Figure 2: Gut Reaction - Cohort Discovery Tool (i2b2) - screenshot



Feedback from initial test users

- For cohorts selected by phenotype, it would be useful to show what other resources (e.g. NHS records, genetics, samples) are available
- It would be useful to have a roadmap of data expected
- Significant interest in the development of synthetic data

Next steps

- Engagement with the Patient Advisory Committee - June 2021
- Simplified data access application & authorisation arrangements - July 2021
- Cataloguing 'new' data received from NHS Trusts – July 2021
- Planned launch in September 2021
- Further testing, feedback, and refinement throughout 2021
- With additional datasets to follow in due course
- Discussions are also underway with HDR UK to explore:
 - what might be required to standardise metadata across Health Data Research Hubs or Alliance members to work towards federated cohort discovery across different TREs
 - development of synthetic data (made up data with real-world characteristics).

Wider Industry Engagement

The highlight of the first 18-months of the Gut Reaction programme has been the commitment of commercial companies to long-term sustainability. Two major pharma companies have agreed to provide more than funding over five-years to support our activity - demonstrating confidence in the work of the hub and the long-term use of its data.

Data access applications from industry are rising and three were approved at the most recent Steering Committee meeting in February.

Despite the delays to sourcing data from NHS Trusts due to COVID-19, the broad range of very different studies already undertaken provides evidence of the utility of the data.

- Studies with large pharma show that the Gut Reaction data can be used to recall participants with an identified genetic characteristic for clinical trials through the NIHR BioResource. We have also shown that the use of hub data can be used to characterise phenotypes and determine the prevalence of subtypes
- With large pharma collaborators we are also investigating the use of data to characterise genotypes
- With SMEs, the data have been used to generate research tools

Academic research remains the main driver for collaboration with the hub, which will leverage additional funding required to support long-term sustainability and provide evidence of the depth and richness of the data through publications.

1.5 INSIGHT

Data Improvement Use Case 1 - *How are improvements to the data increasing opportunities for Hub users? Do users of the data span multiple sectors?*

Clinical data improvements:

In successfully achieving Milestone 1 (December 2019), INSIGHT provided two datasets: UHB Eye Image Dataset Release 001 and MEH Eye Image Dataset Release 001. In June 2020, we created a bespoke COVID-19 dataset to inform NHS service recovery, planning and clinical audit (*Project Reference 1, KPI worksheet*). Approaching Milestone 2, INSIGHT added two further datasets covering the leading causes of blindness, Age-related Macular Degeneration (AMD) and Diabetic Retinopathy (DR).

INSIGHT's baseline datasets comprised single tables with 16 columns (data points) per record and image data captured in basic formats with only partial image metadata available. By virtue of the capability established through the Hub, all datasets have been updated and greatly improved with the addition of a wide range of clinical metadata. Additionally, by adapting and improving cloud-based storage technology for raw clinical and imaging data, the lead-in-time necessary to generate new datasets has been reduced, thereby enabling the Hub to be agile in response to user requirements. As a result, INSIGHT offers high-quality, well-curated, large, longitudinal datasets of linked retinal images to clinical outcomes in ophthalmology. The expansion in clinical metadata coupled with image data enables enhanced discovery and permits prospective applicants to the Hub - from major Pharma and Tech companies to SMEs, academia and the NHS - to address a variety of research questions across different domains, including disease detection, progression, outcomes, treatment burden, treatment response, and modelling/prediction.

Advised by clinical subject matter experts and in collaboration with industry partners, INSIGHT harmonised data across the partner NHS Trusts and created a scalable multi-table database schema, which includes 183 columns (data points) per record. The schema has the potential for UK-wide use having already achieved interoperability across the two leading ophthalmic Electronic Health Record (EHR) systems (some 80 percent UK coverage). The schema harmonises with existing standards for *systemic health* (avoiding redundancy/duplication) but provides the first such standards for *eye health*. These have been developed in discussion with international collaborators, including regulators. The data dictionary, including data definitions and lineage information will be publicly available. Additionally, we are mapping to Fast Healthcare Interoperability Resources (FHIR) in order to improve international accessibility and interoperability.

Image data and metadata improvements:

INSIGHT has invested expertise and resources to build tools and processes with the ability to read images from the six major UK source systems, each with their own proprietary image file format. INSIGHT has created a standard process to store the image metadata, which includes harmonising data from different devices, enhancing INSIGHT's capability of producing multiple outputs per image. Uniquely the Hub can provide three different imaging formats according to user and sector need:

1. Standard proprietary format: the imaging vendor's default output;
2. Raw image format: of particular value for machine learning/Artificial Intelligence (AI) applications;
3. DICOM format - the Digital Imaging and Communications in Medicine (DICOM) standard is the international standard for imaging and related data.

The proprietary conversion process and logic that INSIGHT created for these transformations was reviewed, refined, and validated in partnership with our industry partners leading to the ability to process into DICOM standard with a high level of quality and fidelity, supporting harmonisation and interoperability (including of derived technologies). Consultation with other international bodies and the major imaging vendors helped ensure alignment with industry/regulatory standards.

Processing performance has been accelerated with the Hub now able to process 9 million images (90TB) within 21 days - a 25-fold increase in efficiency since Milestone 1. This exceptional level of performance will be critical to successfully serving customers from different industries and sectors, in particular for bespoke image processing.

Infrastructure and tooling improvements:

The initial tools and applications comprising the INSIGHT data platform were built for cloud-first implementation, drawing from the hugely successful approach developed through the Moorfields Eye Hospital

and DeepMind (Google Health) Research Collaboration. During 2020, we modified and reconfigured some of these tools to support and enable on-premise installation for data processing, broadening participation of NHS Trusts that are not yet able to utilise cloud environments. Data from new Trusts joining the Hub will not only enrich the existing data, but also create more diverse datasets increasing the opportunities for INSIGHT users. Our infrastructure allows us to quickly and efficiently scale to seamlessly service multiple concurrent user requests.

A three-tiered data quality assurance process has been implemented, with built-in monitoring processes that automatically flag data quality issues and provide feedback loops at each level. The three tiers cover the data along its journey from the generation and collection at the individual Trust-level, the ingestion into and processing through the INSIGHT platform, and preparing and handing-off the data to users.

The result of our data improvement efforts is visible by the Platinum score achieved across all four INSIGHT datasets on the HDR UK Innovation Gateway:

Major case study: Assessing and responding to COVID-19 impact on eye services:

This first bespoke dataset generated through INSIGHT supported the analysis of the impact of COVID-19 on patients with AMD, one of the leading causes of blindness. This case study demonstrated how INSIGHT could uniquely address an urgent need through its dataset and analytical capabilities. The project - initiated by Moorfields Eye Hospital and University Hospitals Birmingham with data science and analytical expertise from Roche-Genentech - provided the first reliable estimates of the scale and severity of the vision loss arising from delays in treatment for newly diagnosed 'wet AMD' during the COVID-19 period, informing NHS (and industry) providers on strategies to optimise care of patients during service recovery. This case study exemplified the exceptional strengths of INSIGHT including:

1. Scale and diversity: combined anonymised patient data from the two NHS partners providing insights into key differences regarding the patient journey and timing between the two hospitals for over 2,200 affected patients from diverse populations.
2. Cross-sector understanding of user needs: the Hub's broad expertise from across NHS, Pharma, Tech, academia and the charity sector was evidenced in co-creation of this dataset, enhancing its quality, relevance, accessibility, analysis, and reporting.
3. Unique data: the Hub's dataset was combined with Roche's landmark MARINA Trial Data (Minimally Classic/Occult Trial of the Anti-VEGF Antibody Ranibizumab in the Treatment of Neovascular AMD) - one of the last studies to predate the widespread adoption of modern AMD therapies, and an ideal comparator to estimate the impact of loss of ophthalmic services.
4. Unique analysis: combining the expertise of Hub partners permitted the exceptional modelling of the number of patients who would have needed to initiate anti-VEGF therapy for wet AMD in a specific time window following UK lockdown where ophthalmic services were interrupted.

As a first use-case exemplar, this project was of direct value to the NHS Trusts. It furthermore enabled end-to-end testing of INSIGHT data collation, ingestion and curation systems, and established a consistent data treatment protocol to assure quality.

Additional applications and requests:

INSIGHT received and approved a second application for bespoke data requested by an NHS-led study investigating whether response to treatment for those with AMD is affected by having Diabetes Mellitus. Diabetes and 'wet AMD' are both associated with abnormal growth of new blood vessels within the retina. Although these diseases are common it is still unclear what impact diabetes has on AMD, and whether the

presence of diabetes may affect the treatment outcomes for AMD. The aim of this study is to help identify whether the presence of diabetes affects the response to treatment in AMD in terms of measured visual functions and structural changes, and it explores the potential for personalisation of medicine in this field.

The application for a third bespoke dataset focusing on 'wet AMD' data including image data, is currently under review with the Data Trust Advisory Board. A fourth application to INSIGHT for a dataset focusing on diabetic macular oedema has also been received. These two diseases are the leading causes of blindness in the UK, Europe and US, and we anticipate that these exceptional datasets will become a cornerstone to enabling academic institutions and industry to find solutions that address key health challenges and provide direct and tangible benefits to patients and the public.

INSIGHT has been approached by a range of SMEs and industry leaders (in the UK and globally), spanning both Tech and Pharma, expressing interest in our Hub and seeking to explore potential areas of mutual benefit. For example, a UK Tech SME is meeting regularly with the INSIGHT team to design an *Oculomics* project to detect cardiovascular risk through high-street optician scans. We are expanding the development of our Hub's sustainability model to incorporate alternative funding models to engage with SME's, such as intellectual property or equity share arrangements. This recognition of segmentation of user requirements also reflects our strategy of supporting all sectors and achieving a balance of early, medium and longer-term value return under the Hub's forward plans towards Milestone 3.

Data Improvement Use Case 2 - *How has the data been used to enable Hub services, income generation opportunities and develop Hub sustainability?*

Case study context:

Age-related Macular Degeneration (AMD) is the leading cause of blindness in the UK, Europe and the US. In the UK, AMD affects more than 600,000 people. The most progressive type of AMD is 'wet AMD' in which abnormal blood vessels grow into the retina. Without treatment it is a rapidly blinding condition, but the advent of intravitreal injection of vascular endothelial growth factors inhibitors ('anti-VEGF') treatment has revolutionised the outlook for people with wet AMD. Timely initiation of treatment with ongoing surveillance is 'sight-saving' in these people.

Despite the development of these therapies, there remain important questions for NHS organisations, academia and industry. These include: understanding *at the patient-level* what factors are limiting outcome; evaluating the burden of treatment on patients and carers; exploring service redesign to improve patient outcome and achieve efficiency; and understanding *how* changes in the structure of the eye from disease impact its function.

The INSIGHT AMD dataset:

The INSIGHT AMD dataset is an exceptional longitudinal dataset of routinely collected imaging and clinical metadata from patients receiving treatment for AMD at the world-leading Moorfields Eye Hospital NHS Foundation Trust, from 2004 to the present day.

This dataset encompasses all patients at Moorfields who received at least one injection of either Lucentis (ranibizumab) or Eylea (afibercept). These therapies began at Moorfields in 2007, however, the dataset includes any imaging or clinical metadata that is available for patients prior to that time - for example, patients who were initially monitored for the early or intermediate forms of this disease prior to receiving treatment. This dataset additionally includes data from both eyes in each patient case - for example, it includes data from

fellow eyes not receiving injections. Consequently, the dataset is rich in data on 'dry AMD', i.e. prior to development of sight-threatening 'wet AMD' in that eye.

The clinical metadata includes demographic information, visual acuities (predominantly measured with Early Treatment Diabetic Retinopathy Study charts), diabetic retinopathy grading, and information regarding intravitreal therapies and ocular surgeries. The imaging metadata includes colour fundus photographs, optical coherence tomography (OCT) scans (Topcon, Heidelberg Engineering), fundus autofluorescence (Heidelberg Engineering), fluorescein and indocyanine green angiography (Heidelberg Engineering).

The AMD dataset is continuously refreshed. As of February 2021, it consisted of 13,562 eyes receiving Lucentis or Eylea, 173,112 injection episodes and 908,242 ophthalmic images. This comprises the largest single-centre database of patients with AMD anywhere in the world, covering more than a decade of follow-up for these patients.

The provenance and strength of the AMD dataset is supported by several publications in high-impact, peer-reviewed scientific journals, including *Nature Medicine*, *Ophthalmology*, *British Journal of Ophthalmology*, *JAMA Ophthalmology*, and *Ophthalmology Retina*.

Applications to INSIGHT for use of the AMD dataset:

Sector: NHS and academia

Project: *Assessing response to treatment for people with AMD affected by having Diabetes Mellitus*

AMD and Diabetes Mellitus are common conditions that both affect the retina, with profound impacts on vision, wider health and well-being. Diabetes causes changes in blood vessels throughout the whole body and, like AMD, can lead to the growth of abnormal blood vessels in the retina. To provide the best treatment to an individual we cannot think of a disease on its own, but should consider the disease in the context of the person's wider health. Understanding the impact that diabetes has on the development and response to treatment of AMD is a key step towards more personalised treatment for this blinding condition.

The application to use the INSIGHT AMD dataset compares the cohort of patients with AMD who (a) have diabetes and (b) do not have diabetes with regard to:

1. Demographics and metrics of disease at baseline;
2. Progression of disease in terms of function (visual acuity); and
3. Progression of disease in terms of structure (OCT imaging).

The study seeks to model the effect of AMD treatment on progression of the disease over time, and to test the impact that the presence or absence of diabetes has on outcome. By using the linked imaging and clinical data, the study is exploring whether the relationship between structure and function is different in people with diabetes, and assessing any impact on treatment response that may signal a need for alternative treatment regimens.

This project is about using the power of data to move beyond the current 'one size fits all' approach to treating AMD, and to explore how having another common disease (in this case diabetes) may affect how AMD behaves, and how it responds to standard treatment. The application for use was received in December 2020 and assessed by the INSIGHT Data Trust Advisory Board in January with a recommendation to the NHS Data Controller that access be granted without further conditions, reflecting that the proposed outcome of the study provides a clear benefit for the public.

How work on the AMD dataset has enabled Hub data quality improvements:

Quality of data acquisition: the dataset highlighted potential areas for improvement at acquisition to improve standardisation (e.g. type of acuity chart) and completeness. Rapid feedback to the NHS partners has led to consistent best practice being adopted across the NHS sites, improving data quality for both direct patient care and research purposes.

Quality of data curation, standardisation and harmonisation: working between NHS Trusts that use the two leading alternative EHR systems for ophthalmology, INSIGHT now has a common data dictionary and data catalogue for AMD that would be applicable to over 80 percent of UK hospital eye services.

Linkages and data reference sets: a unique aspect of the AMD dataset was INSIGHT's ability to respond to the COVID-19 pandemic. The ability to merge routinely collected clinical data was powerful to demonstrate what numbers of patients to expect following lifting of lockdown (wave 1). The quality of this data was further enhanced by use of the control data from Roche to predict the level of vision a person would have if they delayed treatment over the course of 12 months. This is an example of direct benefit back to the NHS, as it has helped inform operational planning of service restoration including Diagnostic Hubs and capacity requirements for virtual and in-person services.

How patient and public involvement has been central to this work:

Within its broader remit across eye health, INSIGHT is responding to patient and public priorities through the prioritisation and delivery of its datasets and related projects. Patient and public involvement and engagement (PPIE) has directly influenced this prioritisation in three ways:

1. National prioritisation: the selection of AMD as a project reflects its place as the number one priority within the James Lind/Fight for Sight Priority Setting Partnership for sight loss. Moorfields helped convene this national process which involved 2,220 people of whom over 60 percent were patients with sight loss or an eye condition. This directly informed dataset prioritisation and design of the INSIGHT platform.
2. INSIGHT's Patient and Public Advisory Group: provided direct input into how the Hub functions, including how it communicates, with the PPIE Lead being a member of the INSIGHT Leadership Team.

INSIGHT's Data Trust Advisory Board (DataTAB): ensures patients and public scrutinise all requests to access the AMD dataset, providing advice to the NHS Data Controllers for final approval. The DataTAB is chaired by Dr Wen Hwa Lee, the CEO for the charity Action Against AMD. He is also a member of the INSIGHT Leadership Team. The DataTAB have been extremely positive in supporting research outputs from routinely collected clinical data exemplified by INSIGHT.

1.6 NHS DigiTrials

Data Improvement Use Case 1 - *How are improvements to the data increasing opportunities for Hub users? Do users of the data span multiple sectors?*

NHS DigiTrials supports researchers through publication of clear, detailed metadata for a range of prioritised data sets on the [Health Data Research \(HDR\) Innovation Gateway](#).

During 2020 the programme sought to enhance this support by make quality improvements in several areas:

- Uplift metadata from the HDR metadata specification v1, to v2, including the provision of 'Observations' metadata to describe the coverage of each data set.
- Increase the number of DigiTrials prioritised datasets represented on the Innovation Gateway.
- Ensure the metadata for each dataset exhibits high levels of quality and completeness, as shown by the awarded medallion ratings.
- Ensure each data set has data utility information appended, making improvements to the underlying data as appropriate.

Eighteen datasets were specifically targeted for improvement as part of the DigiTrials work. Of these:

- The metadata for 15 datasets were uplifted from the HDR v1 metadata specification to the HDR v2 specification.
- A further three datasets were onboarded to the v2 specification.
- 14 datasets had observations metadata appended.
- 17 datasets were awarded either gold or platinum medallions for metadata quality and completeness. In all cases where the dataset already existed on the Innovation Gateway this was an improvement from either bronze or silver.
- In addition to v2 requirements, 16 datasets prioritised by DigiTrials have data utility framework metadata published to the Gateway.

This improvement has been possible due to the integration of DigiTrials priorities and HDR UK Gateway requirements with NHS Digital ways of working which embrace a philosophy of continuous improvement. A process was developed to identify, collate, improve and onboard Innovation Gateway metadata involving close liaison with the Information Asset Owner (IAO) and other subject matter experts for each dataset. IAOs have spent significant time taking part in interviews with the Data Architecture team ensuring that the metadata presented is both comprehensive and accurate. They have also supported in the provision of metadata from other parts of the organisation including analytical expertise in development of observations data. We are looking forward to embedding this new way of working across NHS Digital as we continue to improve our corporate metadata.

The DigiTrials programme has maintained ongoing engagement with MetaDataWorks during the onboarding process, taking advantage of the guidance service offered to effectively onboard v1 metadata and transition to v2. In addition, the programme has supported HDR UK in the implementation of the v2 Gateway, working through technical issues affecting presentation with HDR developers.

This robust approach to the development of processes and provision of metadata has resulted in comprehensive, quality information on the HDR UK Innovation Gateway, providing researchers with a great opportunity to identify and access the best possible data for their research.

Data Improvement Use Case 2 - *How has the data been used to enable Hub services, income generation opportunities and develop Hub sustainability?*

NHS DigiTrials is focussed on providing data services across the end-to-end journey of clinical trials. These services provide benefits across healthcare (patients, industry, trialists and NHS) and are a key component in the life sciences industrial strategy. NHS DigiTrials has a unique data offering through its access to a vast breadth of NHS datasets and the subject matter expertise that resides within NHS Digital.

NHS DigiTrials has taken a user-centred approach to deliver both standard and tailored solutions to meet the direct requirements of trialists. The business model contains four services (Feasibility, Identification, Communications and Outcomes) that can be mapped to the 'Find – Recruit – Follow' pathway. The initial plan for the hub, using the HDR UK grant, was to develop and take to market a fully developed Feasibility service.

However, due to the COVID-19 pandemic NHS DigiTrials was approached by several high-profile trials to urgently provide access to a wide range of outcomes data. To meet this demand an agile approach was adopted to rapidly develop a relevant and efficient Outcomes service.

An in-depth case study that demonstrated how NHS DigiTrials was able to pivot from a singular focus on the Feasibility service to the rapid development of the Outcomes service is the partnership established with the RECOVERY (Randomised Evaluation of COVID-19 thERapY) trial. The RECOVERY trial was established in March 2020, as a randomised controlled trial to test a range of potential treatments (including re-purposed and novel drugs) for COVID-19. Over 35,000 patients have been enrolled so far from 177 NHS hospitals in the UK.

NHS DigiTrials worked directly with trialists to identify requirements and tailored data solutions to support the following objectives of the RECOVERY trial:

Primary objective: To provide reliable estimates of the effect of study treatments on in-hospital death.

Secondary objectives: Assess the effects of study treatments on duration of hospital stay; the need for (and duration of) ventilation; and the need for renal replacement therapy.

The RECOVERY Trial required NHS DigiTrials to move at unprecedented speed to deliver both the Data Access Request Service (DARS) application and data production processes in a matter of weeks rather than months. To facilitate this a new service wrapper was developed so that the NHS DigiTrials team engaged much earlier with trialist than previously would have been expected. This allowed the team to get a better understanding of the requirements and to contribute to shaping decisions that could affect the path to live. In addition, the team took responsibility for working through and completing the DARS application on behalf of the trialists. Access was also provided to subject matter expertise within NHS Digital to support on coding definitions, data dictionaries, data quality, frequency for collection etc; to facilitate quick and effective decision making. The result of this was that the DARS application was completed and approved within 13 working days whilst maintaining a robust approach with independent advice. The approach, techniques and service wrapper developed with the RECOVERY trial have been used to achieved similar results with other high profile trials including PRINCIPAL, HOSTED PHE and Convalescent Plasma.

Dataset selection was focused on providing rapid access to timely data to support early decision making (SUS+ and fact of death). However, the datasets available to NHS DigiTrials has allowed this requirement to be balanced with the need for more complete datasets to ensure the quality of outcomes (HES and Civil Registration of Deaths). Early engagement of Data Production teams allowed automated and tailored solutions to be setup to provide more frequent disseminations (weekly/monthly) of data. This service is now being offered to a wide range of trials including daily/twice weekly dissemination of testing data to the PRINCIPAL and FLARE trials.

Using an agile and iterative approach has also allowed new and existing datasets that have been made available for COVID-19 research to be quickly added to the RECOVERY portfolio. These include:

- Second Generation Surveillance System (SGSS)
- CV-19 Hospitalisation in England Surveillance System - ICU critical care/ventilation (CHESS)
- GPES Data for Pandemic Planning and Research (GDPPR)
- Primary Care Medicines Prescribing data
- NHS DigiTrials are continuing to work to make more datasets available to clinical trials both within and outside of COVID-19 research (vaccination data).

The NHS DigiTrials team have also provided continued business support over the last year holding weekly meetings with the Recovery team to ensure the timeliness and high quality of the data was maintained. This approach to the relationship has also allowed changes to formats to be efficiently managed through the DARS and production processes in a timely manner, maintaining the relevance and efficacy of the trial. Other large

and complex trials have subsequently benefited from this approach including studies such as Convalescent Plasma.

By providing timely and comprehensive data on the impact of study treatments NHS DigiTrials has been critical to the following real-world achievements:

Reduce the burden on front line workers, already stretched via the pandemic, by minimising the time taken to collect data from patients when recruiting them into the study.

In June 2020, the RECOVERY trial demonstrated that the inexpensive and widely available steroid, dexamethasone, reduced the risk of death by one-third for patients on an invasive mechanical ventilator and by one-fifth for those requiring oxygen. This was the first treatment for COVID-19 shown to save lives and was rapidly adopted as part of standard hospital treatment around the world.

In Feb 2021 it was found that Tocilizumab, an intravenous drug used to treat rheumatoid arthritis, reduces the risk of death when given to hospitalised patients with severe coronavirus.

In addition to these successes the Recovery Trial has demonstrated results showing that hydroxychloroquine, lopinavir-ritonavir, azithromycin, and convalescent plasma have no benefits for patients hospitalised with COVID-19.

Being able to access this data from one place has reduced the time from closing the sub-group of the trial to doctors being able to prescribe the treatment.

The impact of the data provided by NHS DigiTrials was recognised by Prof. Sandercock, the chair of the RECOVERY data monitoring committee, when he took the unusual step to write to NHS and Research staff saying: 'We have been so impressed with the progress of the trial, and the quality of the data, we decided we should write a letter to thank you for your involvement in this national research effort'.

In addition to the Outcomes service, NHS DigiTrials has been able to use RECOVERY trial as a pilot for its Communications service. The RECOVERY trial had a requirement to send a series of newsletters to communicate outcomes information to participants. However, due to the speed of setup and the desire to reduce the burden on front line workers contact data for participants had not been captured at the point of recruitment. NHS DigiTrials were able provide the missing data by linking the RECOVERY cohort to the Personal Demographic Service.

In addition, NHS DigiTrials were able to explore options to send newsletters on behalf of the RECOVERY trial and establish a potential blueprint for a managed contact service. Using existing contractual relationships with a third party mailhouse the NHS DigiTrials team were able to send circa. 13k newsletters to RECOVERY participants during Jan 21. Three further mailings are planned during 2021. This model has the potential to reduce lead times and the volume of data that is required to be shared with trials. In addition to this NHS DigiTrials has been able to deliver a dynamic invitations service to six vaccines trials (Oxford / AZ, Novavax, Janssen, AZ PROVENT, Valneva and Oxford COM-COV) sending out over 140k invites to members of the public that had expressed an interest in taking part via the Vaccines Registry.

The success of NHS DigiTrials in supporting RECOVERY trial as well as other key COVID-19 trials has generated significant interest from the clinical trials community. We now have a sizeable pipeline of work that covers all services and market segments (academia, government bodies, NHS & large Pharma). As at the 05/03/21 we have 17 trials that have an active data sharing agreement, nine trials that we are taking through the DARS application process and 23 trials that are at a pre-application/enquiry stage.

1.7 PIONEER

Data Improvement Use Case 1 - *How are improvements to the data increasing opportunities for Hub users? Do users of the data span multiple sectors?*

PIONEER is pleased to highlight significant improvements to the data, informed by user need/requirements, increasing data uses across multiple sectors including academic, SMEs, NHS, Government (Public Health England, SAGE, ONS), international pharma, health technology and healthcare consultancy, forming a strong base for sustainability and growth.

Increase in scale of data

PIONEER offers 20+ years of longitudinal data with >1.2million records now readily searchable through HDR UK's Innovation Gateway, constantly expanded and refreshed in near-real-time. PIONEER has increased datasets on the HDR UK Gateway from a baseline of 2 to 16 since Milestone 1 review.

The range of topics/disease areas has been extended, now including deeply-phenotyped data on Sepsis, Stroke, Thromboembolism, Acuity, COPD, multi-morbidity, genomics, ventilation and COVID. The expanded content includes >70M drug administrations, >100M blood pressure readings, >200M blood test results. It is disease/episode agnostic; inclusive of diverse ethnicities (40% of Birmingham's population is non-white), reflects social deprivation and is ageless, including children/adolescents. We have a future pipeline of datasets, including acute presentations of rare diseases, cancers, ambulance and primary care.

Data modalities have increased, including images (computerised tomography (CT) scans/X-rays) to complement our rich, diverse structured data. Next, our AI team are working on Natural Language Processing of free text, enhancing our offer.

The PIONEER datasets are extracted, designed and staged to enable linkage to other datasets, as demonstrated by data linkage with UCLH as part of DECOVID (PIONEER's COVID collaboration with University College London (UCL), UCLH, and the Alan Turing Institute (ATI)), the BHF Cardiovascular HIC and more recently, with the MRC-funded ADMISSION consortium.

The experienced PIONEER technical team have enabled improvements in data quality in other settings, including identifying programming issues with EPIC, improving data QA/QC processes in one large NHS Trust, and helping another with data extraction processes; sharing our learning to benefit all.

Importantly, PIONEER holds all its data within a central repository, with a single Data Controller and overarching ethics, meaning we can respond agilely to data requests, cutting data to requirements, increasing the opportunities for end users. This is demonstrated through our 20 diverse projects listed in the KPI.

Improvements in the usefulness of the data

PIONEER undertook workshops with ABPI, Medicines Catapult, large pharma, data consultancy, academic researchers and SMEs, to understand multiple sectors' data requirements. E.g., SMEs expressed concern around IG requirements, so PIONEER created 24M synthetic records. We review frequently searched terms on the Innovation Gateway and build datasets based on emerging themes of interest. E.g., COVID and multi-morbidity datasets were rapidly staged responding to Gateway searches, and are being used in projects listed within our KPI.

We adopt a "you need, we will..." approach, with our agile responses noted as a real strength by academic and industry partners.

The metadata is more richly completed with significantly improved medallion scores; currently 16 platinum datasets scoring $\geq 90\%$.

Additionally, we have improved the technical data quality dimensions listed in the Data Utility Matrix:

- Automation and refinement of extract, transform and load process to reduce lag between the data being collected and added to the dataset from monthly to weekly (Gold) and others near-real-time (Platinum).
- Additional documentation, such as URLs, exemplar papers and infographics, with PIONEER support staff available to answer questions (Platinum).
- The infographics (Appendix 1) have been widely praised for presenting complex datasets in an accessible format.
- Data usages are fully consented for commercial uses (following IG approval) via CAG and Ethics approvals (Platinum).
- Any of the datasets can be used in a PIONEER bespoke Trusted Research Environment (TRE), with other data and tools securely brought in as required. Since MS1, PIONEER has developed a secure safe haven on MS Azure and can stage TREs as required (Platinum).
- The experienced Technical Team (>25 years in NHS and Private sector data industries), with industry and customer voices, enabled the extraction, transformation and loading process to be iterated and refined so automation can be achieved to reduce burden whilst maintaining quality.
- Provision of OMOP data models, conforming to international standards (platinum) for Academic Data Scientists after workshops highlighted potential opportunities.
- User feedback has been incorporated into our systems on a “*you said, we did*” basis. For example, after feedback, PIONEER benefited from a workshop with Microsoft International Azure “Ninjas” to build a solution to an end-use case which required a firewall modification without compromising our rigorous security.

Examples of Data improvements leading to opportunities for data use across multiple sectors, translated into meaningful patient benefit.

In March 2020, the UK faced an unprecedented challenge, urgently restructuring health services, developing novel diagnostic and clinical pathways for COVID-19, while identifying signals of risk for the UK population. PIONEER responded by rapidly curating data, leading to 12 projects as described within our KPI. Examples of opportunities provided for Hub users across multiple sectors are highlighted below.

NHS

- PIONEER facilitated data collection and sharing across 4 regional NHS hospitals by designing and implementing a fully electronic COVID-19 screening and management system and real-time COVID dashboard, live from March 4th 2020 (Appendix 2). This guided clinical/management processes, while enabling capture of deeply phenotyped, structured clinical/hospital process data, mapping patient needs to staffing levels.
- The dashboard provided a fast-track screening tool for nationally prioritised clinical trials, supporting and reporting >10,000 recruitments for the NIHR Clinical Research Network.
- Postcodes from cases were converted to local heat maps, shared with Public Health England, to highlight outbreaks prior to community testing.

Academic partnerships

- PIONEER data fed into academic collaborations leading to COVID-related publications. These were shared with SAGE, including the first paper to highlight the risk of poorer outcomes but no health-seeking delays in those of South Asian ethnicity, NHS staff seroconversion and the increased incidence of falls, drug/alcohol and psychiatric ill-health in vulnerable adults during lockdown.
- PIONEER formed DECOVID, an alliance with UCL, UCLH and ATI, curating and housing matched COVID data from Birmingham (4 hospitals) and UCLH. These data were used to answer freely submitted questions (an “Open Science” approach) with analysts from across the country focusing on varying aspects of wave 1, 2 and 3 disease.
- Separately, PIONEER has shared Birmingham COVID data with other academic centres and organisations such as the West Midlands Applied Research Collaborative, ONS, and the BHF Cardiovascular HIC, contributing to national research priorities.

Industry partners

- PIONEER has secured contracted work with Microsoft as part of Project InnerEye. This develops machine learning techniques to combine insights from medical imaging and clinical data, enabling more accurate decision-making. The InnerEye project with PIONEER supports de-identified COVID positive, suspected and negative medical images/clinical data to be analysed using artificial intelligence techniques to form a rapid diagnostic and prognostic tool for COVID. Data and >6000 images are held within a bespoke PIONEER TRE, designed and built with Microsoft.
- PIONEER has also furnished non-COVID data requests including an SME for synthetic data to build an algorithm to identify a rare disease from routine data. Next, we will expand this into Deep Fake, synthetic image data being developed to assist a further SME and other customers.

Other examples of Non-COVID data sharing activity

- Safer prescribing: Building an algorithm which calculates and proposes prescriptions of N-acetylcysteine for paracetamol overdose, reducing drug errors and treatment delay significantly.
- Appropriate medical alerts: Assessing the impact (patient outcomes, resource) of automated NEWS2 alerting across the entire in-patient stay.
- Safer hospitals: Determining the impact of a systems wide, mandated and electronic health record enabled VTE risk assessment tool.

Our activity and future data pipeline are informed by end-user needs, leading to the following outputs:

- 14 x datasets shared with academic DLAs signed (UCL, University of Oxford, London School of Hygiene and Tropical Medicine, University of Birmingham, Leeds University and Cambridge University, Warwick, Leicester, Newcastle).
- Data accessed by over 185 analysts from across the UK.
- PIONEER has received 29 data sharing requests.
- 13 academic publications and 3 more in preparation.
- 32 projects supported across differing specialties and disease groups.
- Insights shared globally, with COVID infographics reaching all continents in 6 languages.
- Sustainable funding model:
- 5 grant applications (>£10M) funded from risk prediction models to secondary infection to thromboembolic disease and imaging, with significant outputs expected.

- Commercial funding secured from internationals and SMEs.

Security and Scalability

PIONEER can create environments to meet the needs of multiple stakeholders, with security reviewed by QinetiQ Security and Defence Contractors. Multiple consumers can be accommodated at one time, with separate and secure TRE's built for each consumer on demand, and then closed after use, saving cloud resource.

Impact of PIONEER to End Users

- Rapidly available real-world data and consultancy for impactful research
- Unique longitudinal diverse and representative health datasets
- Data and expertise to support industry marketing and grant applications
- Expert support and documentation codified using national, NHS and international standards.
- Industry standard cutting edge secure platform.
- Improving workforce skills through shared learning and training

Future and Sustainability

PIONEER is a model we own. We designed a scalable, interoperable system that can be federated to deliver high flexibility and agility both regionally, nationally and internationally. We are excited to move into the next phase of "business as usual", ensuring the sustainability of PIONEER.

Data Improvement Use Case 2 - *How has the data been used to enable Hub services, income generation opportunities and develop Hub sustainability?*

From inception, PIONEER has considered sustainability with a 4 part-model of:

- Product: Offering the highest quality product (data, safe haven and bespoke Trusted Research Environments (TRE))
- Process: Data access, curated to match end-user need and provided in a timely manner
- Impact: To build awareness of PIONEER and our offer
- Funding opportunities/partnerships: A mixed model informed by cross-sector feedback including NHS, academic, large companies and Small and Medium-sized Enterprises (SMEs)

As described in Data Utility (DU)¹, PIONEER has increased the number of datasets on the HDR UK gateway. The metadata richness, technical quality, increased specialty data and increased modalities has resulted in contracts and opportunities in 8 sectors, with 4 of these being industry (Sector breakdown is provided in our KPI).

Product

PIONEER is a model we designed and own. A scalable, interoperable system that can be federated to deliver high flexibility and agility regionally, nationally and internationally. The Microsoft Azure platform, built by Microsoft platinum partners and cybersecurity tested, offers a cutting-edge safe haven where our large data repository sits, with the ability to prepare data-cuts at will and on demand, through a single Data Controller and single ethical/governance approval framework. Data-cuts are placed in bespoke TREs which can house

multiple analysts in secure and separate environments. Feedback has informed the expanded analytic tooling suites available.

Process

PIONEER has established robust processes to ensure *timely, effective and ethical access to large-scale health datasets*, with public and patient oversight. We are building strong relationships with our customers, to better understand their priorities and provide the data and support they need. Due to the agile and responsive nature of PIONEER, we can accommodate urgent requests. For example, an urgent COVID request was furnished with data and initial analytics provided within 1 week.

Impact

PIONEER recognises the importance of exemplar outputs and networking events to increase awareness and gain insight as to end-user requirements. As well as the projects discussed in DU1 and our KPI, the following examples are highlighted from the 32 projects we are supporting currently.

NHS

PIONEER has enabled NHS partners to use their own data more effectively to improve patient services. This work increases awareness of PIONEER, with a further 2 hospitals currently in discussion to become data partners. To date, PIONEER has provided data for 18 NHS projects, ranging from pneumonia to maternal health to stroke, as well as the COVID-related impact highlighted in our KPI/DU1 report.

An impact project for the NHS is "*Same Day Emergency Care (SDEC)*". UK emergency hospital admissions are rising. A proportion are suitable for SDEC, where assessment/treatment are provided without admission. This is beneficial for patients and reduces pressures on hospitals. SDEC is highlighted in the NHS Long Term Plan, but rarely offered. This project identifies patients who might benefit from SDEC across 4 hospitals, informing new services. In phase 2, we will evaluate impact.

Industry

PIONEER has contracted commercial work, 4 quotations under review and 12 commercial enquiries in development, including 7 signed Non-Disclosure Agreements (NDA). Projects range from building tools for acute care pathways to identifying rare metabolic diseases to risk stratification for patients prior to unplanned surgery.

One specific contracted example is work with a large pharmaceutical company to combine SNOMED-CT codes and other clinical data to generate an algorithm to identify early signals of rare cardiomyopathies. These patients can then be specifically targeted for early clinical interventions (therapies/implantable devices) to prevent the devastating consequences of sudden cardiac death. A second is the Microsoft InnerEye project, described in detail in DU1. A third is sharing synthetic data to explore themes in cardiovascular disease, with data carefully constructed for diversity and clinical accuracy. Phase 2 will test the model on real-world data for further refinement.

Academia

Feedback from Academia has enabled us to adaptively respond, providing both data and tooling. PIONEER's large, granular data offering, promoted via the Gateway, has resulted in 13 academic publications and 3 more in preparation. Additionally, PIONEER has been embedded in 6 successful grant applications, with 4 more

submitted, in second stage applications. The team have staged datasets for 8 universities and 4 NHS Trusts, with data accessed by over 185 analysts from across the UK.

Three specific examples of impact include

- A UKRI-funded study developing tools to aid clinicians to risk-stratify COVID-19 in hospitals. This project includes data from all Waves (including new variant). The model appears more discriminatory than the 4C COVID score. Phase 2 will embed this into an EHR to validate prospectively.
- An academic partnership utilising PIONEER-built synthetic data to supplement clinical decision making for patients with hospitalised pneumonia. A model of machine learning artificial intelligence will be applied, called “Reinforcement learning”. This looks at scenarios where multiple decisions are required, altering future decisions and assesses the potential use of these technologies in this clinical setting.
- A partnership with academic microbiologists combining data with 4 hospitals to assess secondary bacterial infections in patients prescribed broad-spectrum antibiotics. Antibiotic stewardship is a World Health Organisation priority. This project is part of the “*Preserving Antibiotics through Safe Stewardship*” study, working with prescribers and patients to reduce unnecessary use.

Impact also includes marketing and communications. PIONEER have taken multiple approaches from our [website](#); speaking at national/international conferences; trade events; ABPI, AHSN, MD Catapult and HDR UK organised events; media interviews; LinkedIn and Twitter/YouTube. A West Midlands Combined Authority (WMCA) publication entitled “*Mobilising Research Excellence in the Midlands to Tackle COVID-19*” showcased PIONEER. We also have a brochure available for potential end users.

Funding opportunities and partnerships

Sustainability for PIONEER is a mixed model of academic and commercial income. It also includes partnership with the UK life-science sector, recognising the importance of SMEs.

Grant Income

To date, PIONEER has already helped HEE secure >£14.5million in funding awards, which are detailed in the ‘*Milestone 2 KPI spreadsheet*’.

PIONEER is included on grants under-submission, equating to a further £6million, including prestigious applications with national and international collaborators including Denmark, Sweden, Kings College London, UCL, St Andrews and Glasgow.

Commercial Income

Following input from industry experts, PIONEER offers a range of pre- and post-marketing services to businesses, scaling across customers from regional SMEs to multi-national pharma companies. PIONEER has developed a cost model including all facets of data curation, considering the complexity and the rarity of the data and the need for sustainability and growth. The commercial projects outlined above are contracted, and we have a pipeline of commercial work, ranging from early scoping to quotation to contractual negotiations.

SME Engagement

We are passionate about supporting SME businesses in, or looking to enter, the healthcare market. We have created a SME playbook, outlining how PIONEER can help. We have developed a flexible costing model which can include up-front costs, support in funding applications, feasibility testing or profit share/early adoption benefits. PIONEER works closely with other organisations across the West Midlands and UK in supporting

SMEs, including DEMAND Hub. DEMAND Hub provides study design, evaluation and regulatory support and PIONEER provides data, enabling product development.

The benefits of PIONEER's activities

- To patients and the NHS
 - Better healthcare informed by data which reflects the diversity of the UK.
 - Improved quality of NHS data, accessible to inform mandatory reports, benchmarking and improve/plan services.
 - Safety and privacy of data, with public awareness and oversight.
- To industry/ academia
 - Access to specialists in healthcare and data to ensure requestors select the appropriate data.
 - Rapid access to near-real-time data enables studies to start sooner, drugs/devices to get to market faster.
 - Provision of feasibility counts for all requests to ensure requirements are met.
 - Acceleration of clinical development to get therapies to patients faster.
 - Helping customers and researchers address both current and future opportunities.
 - Flexible cost models to suit requestor.
- To PIONEER
 - Reputational benefit from objective outputs.
 - 'Learn by doing' – growth in data and expertise.
 - Working to commercial timelines – agile responses.
 - Building strong partnerships – with repeat business and positive feedback.

The future

PIONEER is a central component of the new University of Birmingham Health Innovation Campus (HIC). This will provide >65,000m² state-of-the-art space acting as a catalyst for the growth of the Midlands' life sciences sector. The HIC will attract SMEs, large pharma and biotechnology firms, and PIONEER will link with these to provide health data and consultancy services.

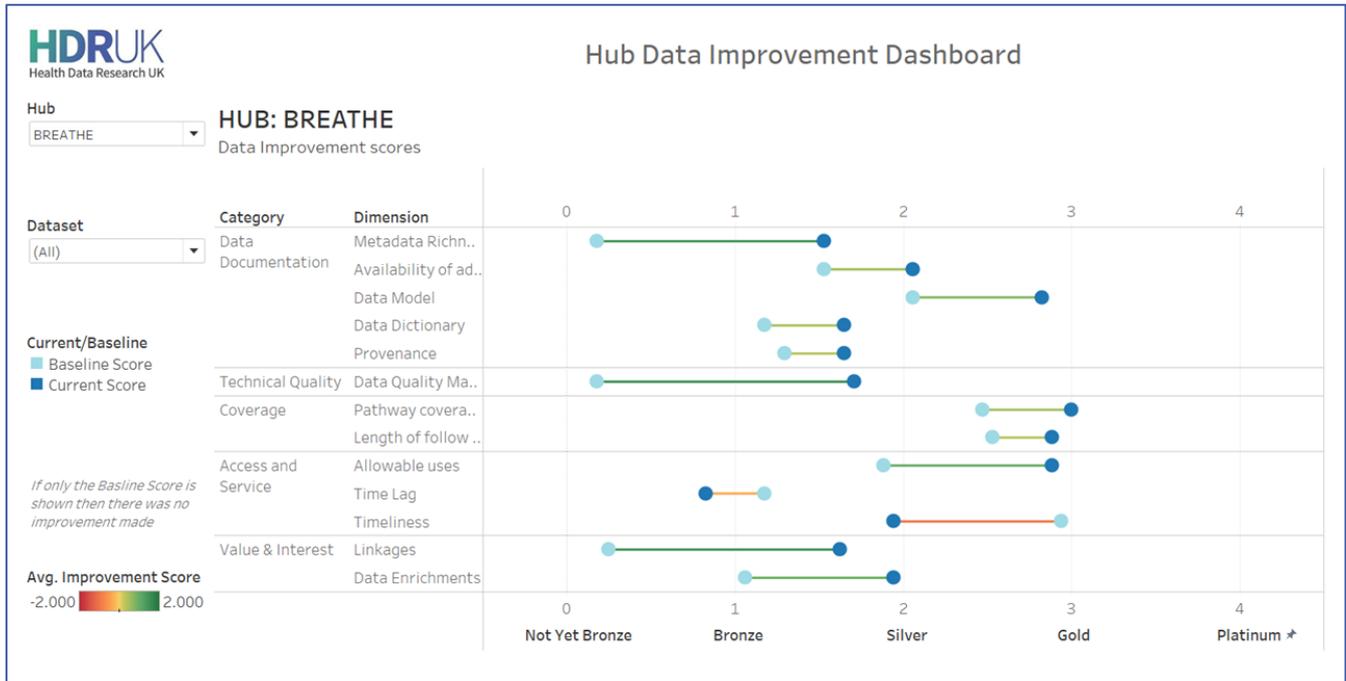
Health data is a core pillar of the Life Sciences Industrial Strategy, featuring heavily in the recent multi-stakeholder Life Sciences Recovery Roadmap, as well as the UK's new R&D Roadmap. In addition, '*Data-driven health and life sciences*' is central to the West Midlands Local Industrial Strategy, committing to "*pioneer innovative hospital networks, platforms and citizen engagement strategies to inform and enable needs-driven, real-world evaluation, validation and adoption of novel technologies, and new approaches to data sharing and implementation of AI*". These strategies set the agenda for major future investment for PIONEER.

We remain focussed on our goal to enable innovative healthcare companies to develop, test and deliver advances in clinical care with near-real-time information. We are looking forward to driving more innovation and impactful outcomes.

Appendix 2 – Data Utility breakdowns

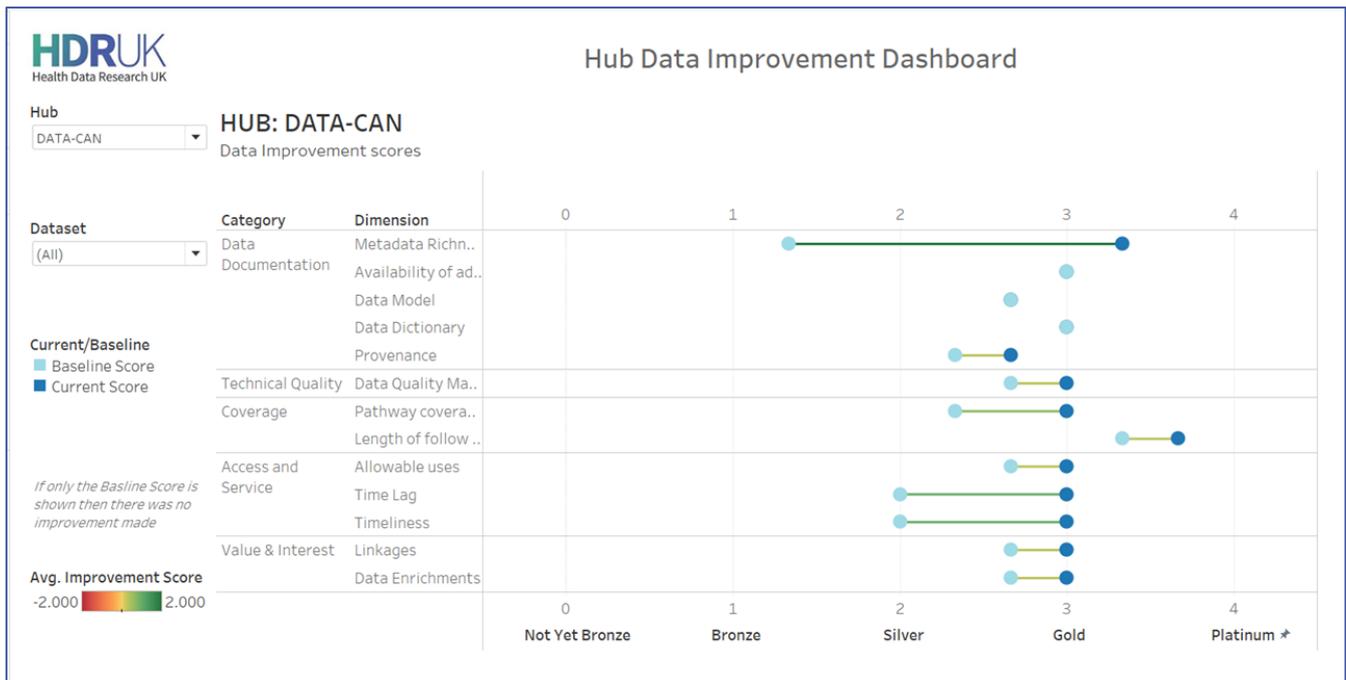
The following dashboards outline the Data Utility improvements on the Hub datasets provided at Milestone 1, using the [Data Utility Evaluation Framework](#).

2.1 BREATHE



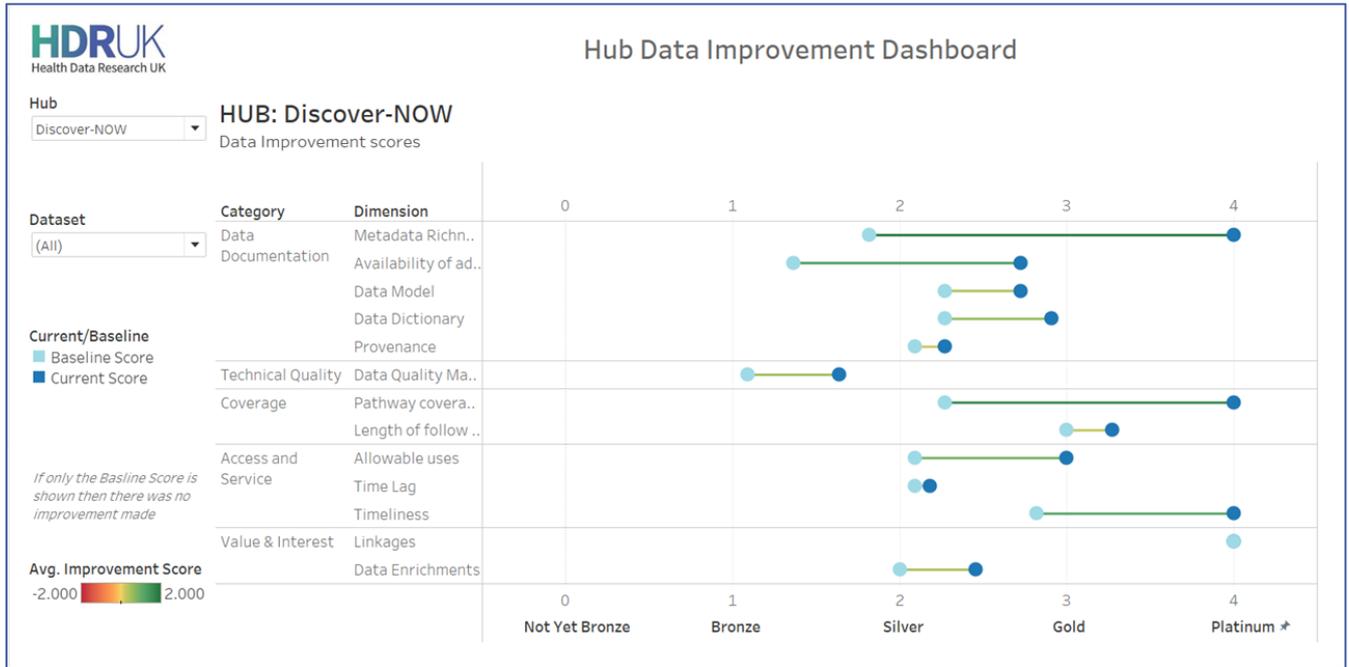
Datasets: 17

2.2 DATA-CAN



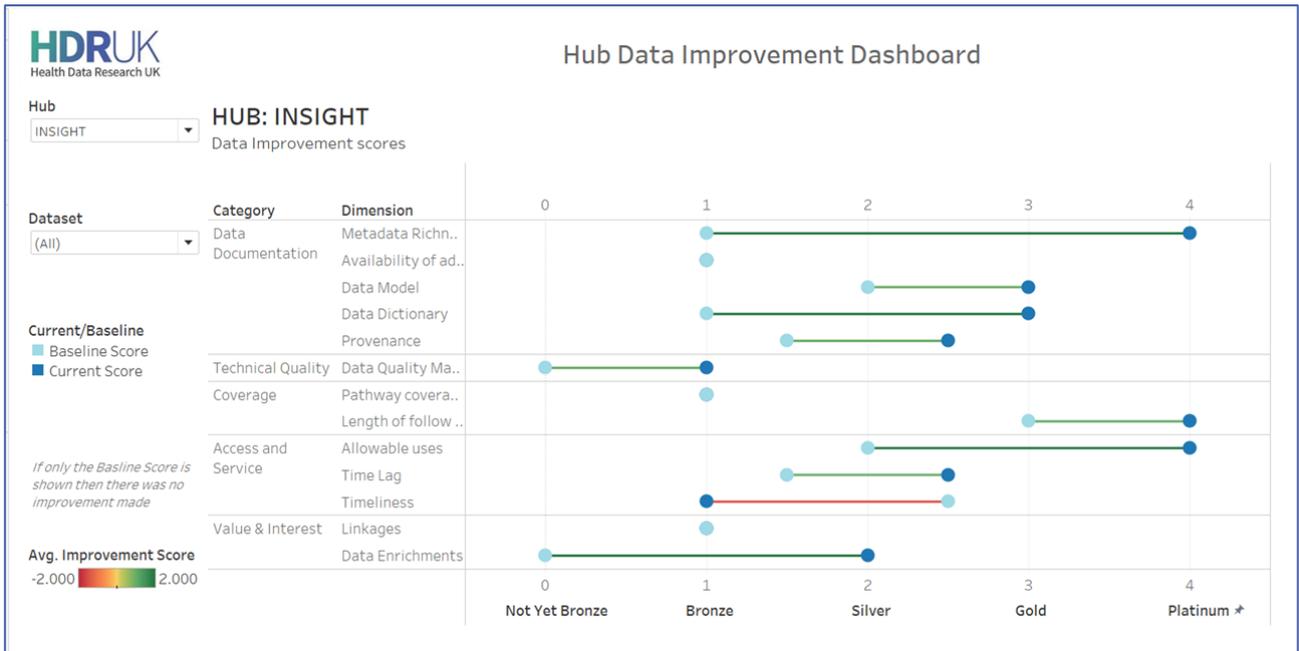
Datasets: 3

2.3 Discover – NOW



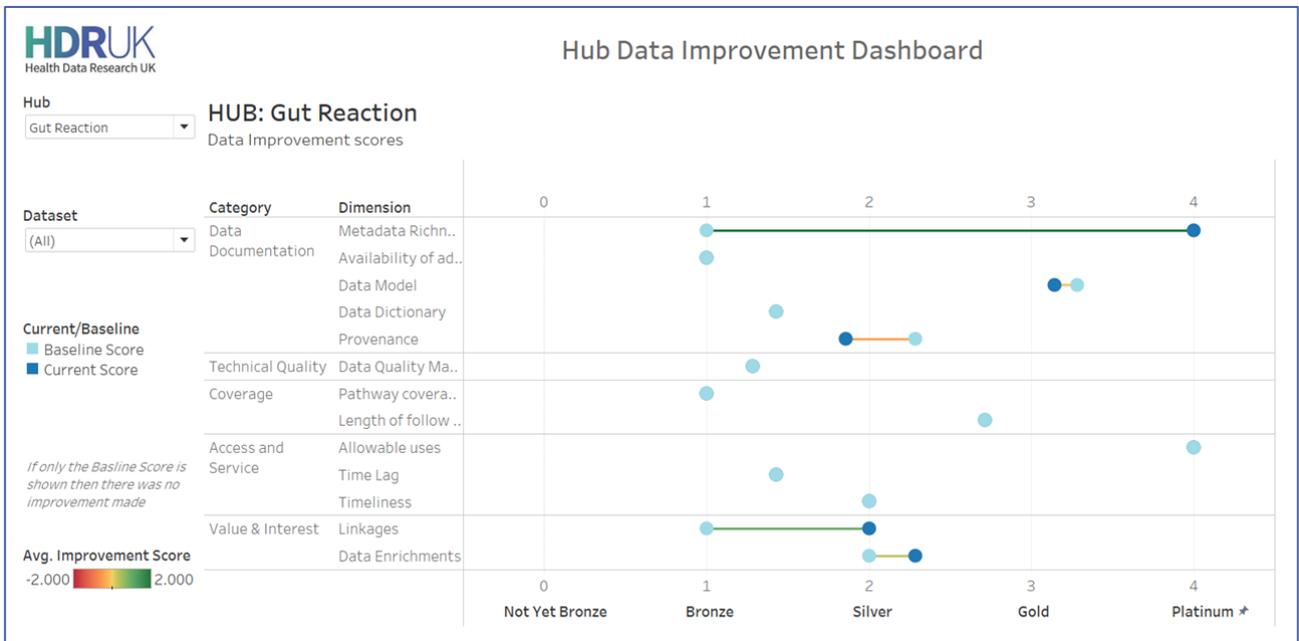
Datasets: 11

2.4 INSIGHT



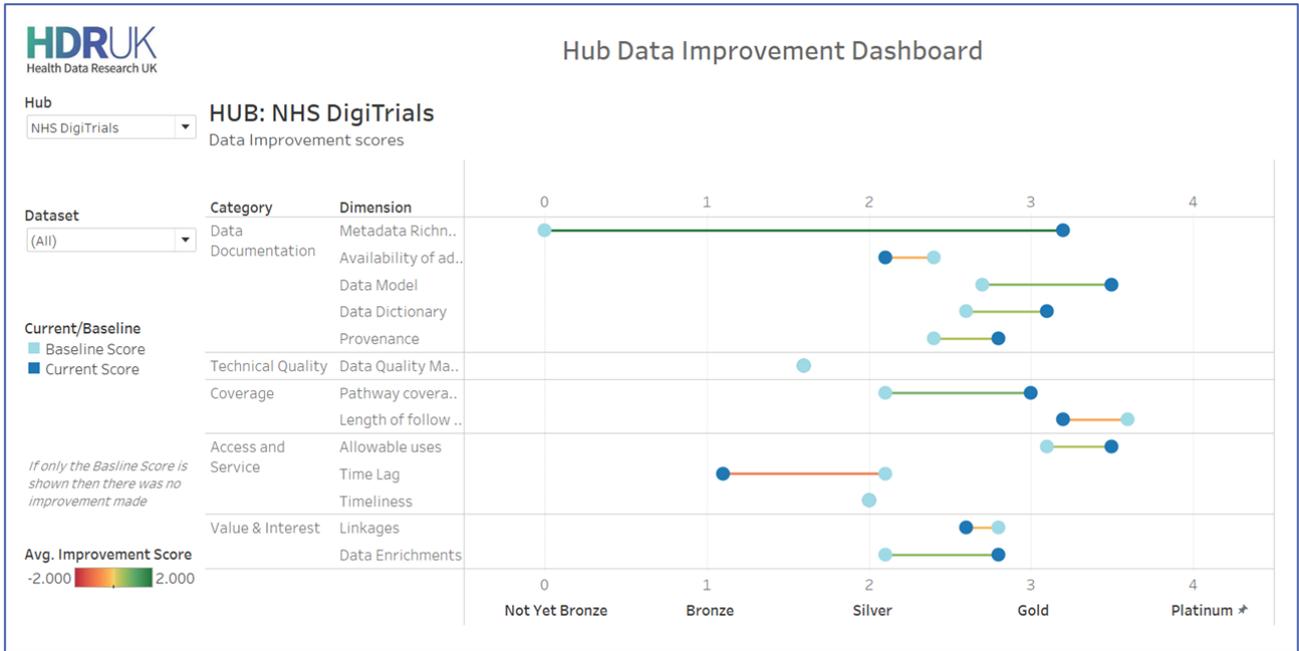
Datasets: 2

2.5 Gut Reaction



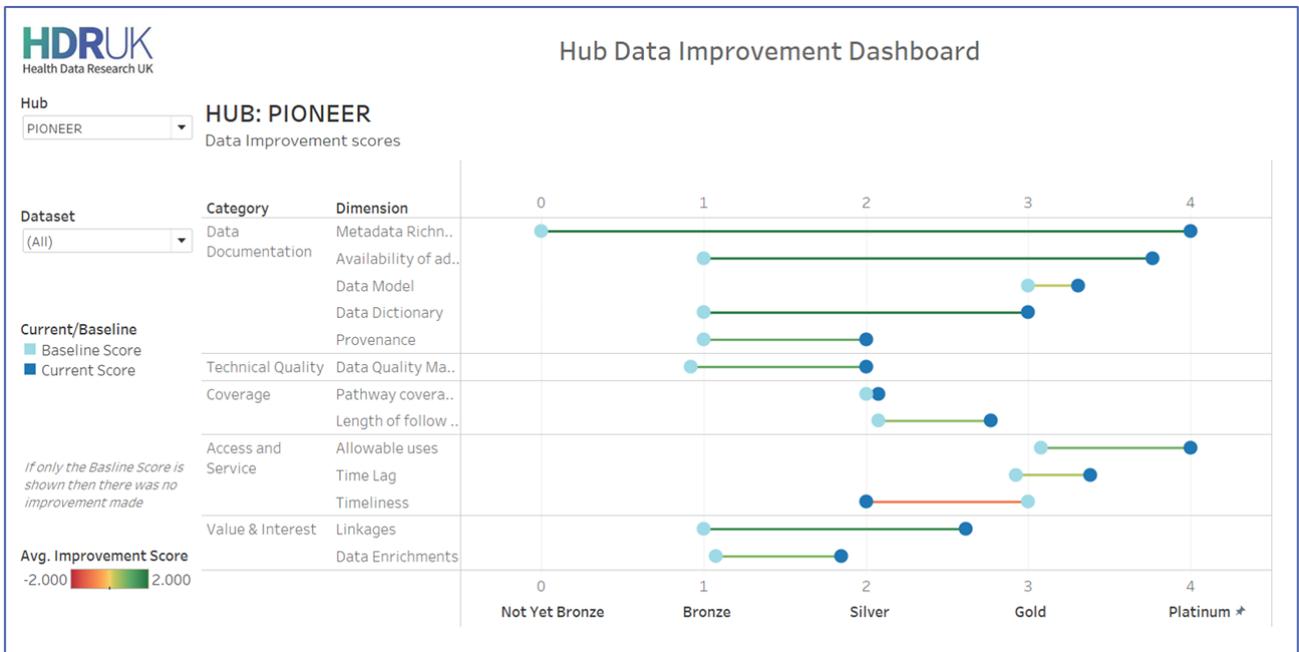
Datasets: 7

2.6 NHS DigiTrials



Datasets: 10

2.7 PIONEER



Appendix 3 – Patient and Public Involvement and Engagement case study highlights

Highlights from the Hubs showing how PPIE has been established and integrated across all Hub activities and the impact it has delivered.

3.1 BREATHE

Our BREATHE Hub benefits greatly from the input of patients and publics with lived experience of respiratory health issues and an interest in finding trustworthy approaches to accelerating high quality data-enabled research. Two co-applicant Lay Leads contribute to the Executive and Scientific Committees, where they help ensure that patient and publics views are included in all key decisions about the Hub. Contributing to the set-up of the Scientific Committee, Lay Leads have delivered real impact by helping shape the process for research projects seeking to become associated with BREATHE. This process now considers: whether projects have meaningful PPIE plans; if there is a lay summary; if steps are being taken to remove barriers for patients and publics to be involved; and if diversity has been considered.

Reviewing BREATHE-associated projects

For example, looking at how Artificial Intelligence (AI) is enabling public opinion to shape the pandemic response in Scotland, Lay Leads asked probing questions relating to the governance of the project, particularly pertinent since the privacy and data protection implications of AI are not widely understood by the public.

The project is examining attitudes towards interventions adopted during the pandemic through the AI-supported analysis of social media postings. The project team were encouraged to explain the legal basis of their use of social media posts, and how privacy-protecting techniques would be adopted to ensure the identity of individuals would not be apparent in the results of the analysis.

Following a demonstration of the dashboard, Lay Leads commented on the presentation of the data, including proposing changes to language used, as well as observing key limitations of this work including the likely gap of input from seldom-heard communities, who may not be active on social media. This interaction between the study team and BREATHE's Lay Leads delivered real impact in that the team were better prepared to address anticipated public concerns relating to analysing social media data using AI, when presenting to the Scottish Government.

Academic research projects

Where new grant applications are developed, many include BREATHE Lay Leads as co-applicants, often co-leading the involvement workstreams of these new studies. An example is the UK-wide Data and Connectivity: COVID-19 Vaccines Pharmacovigilance (DaC-VaP) study. Here Lay Leads co-developed a public involvement strategy within the first month of the project.

Industry connections

TinyMedicalApps, a Founding Partner of BREATHE, is working with our Lay Advisors. A new grant from Innovate UK's Sustainable Innovation Fund will enable this small/medium-sized enterprise to work on

developing a digital platform, the 'Patient Cloud', that aims to facilitate access to patient records across England. Supporting BREATHE public involvement activities will include a product scoping workshop and interviews, co-design workshops, and feedback on commercialisation.

3.2 DATA-CAN

DATA-CAN established a team of PPIE members who are embedded into all aspects of DATA-CAN. Not an 'add-on', they are an integral part of the core business of DATA-CAN, involved in all aspects including recruitment, media and communications, contract review, business proposals and partner discussions.

Patient and Public educational programmes

In 2020, a joint survey sent around patient-data advocates by DATA-CAN, the NCRI, use MY data and HDR UK established patients' experiences of any data-learning, and whether patient advocates saw a need for more. 90% of respondents said they would like more advice or guidance to help with their health data projects. Following the survey, DATA-CAN developed a programme of learning resources for patients and the public wishing to learn more about health data and its uses for analytics and research. Initially the programme comprised eight proposed modules subsequently extended to sixteen separate topics.

We use existing materials, e.g., patient data animations collaboratively developed with Understanding Patient Data, and materials produced by use MY data and the NCRI. We intend to scope out further opportunities to work with science educators and communicators outside of the usual PPIE networks/organisations to develop additional educational content. We have invested in the learning programmes available to our PPIE members, to equip them with a more detailed knowledge of 'all things patient data'

Industry partnerships

In July 2020, DATA-CAN was approached by Flatiron Health, a healthcare technology and services company focused on accelerating cancer research and improving patient care. Flatiron commissioned DATA-CAN to explore three key areas: patient, clinician, and commercial value. The patient panel established by DATA-CAN and Flatiron included members of use MY data and the DATA-CAN PPIE Group, with different voices and experiences, to achieve a mix of views and perspectives. An essential part was for Flatiron Health to explain their value proposition to a group of patients, seeking patient views on whether the process, controls, actions and benefits were clear. The patient group worked collaboratively with the Flatiron team to discuss, clarify and evaluate Flatiron's overall 'value proposition'.

A final DATA-CAN report, co-produced with the patient group, listed 16 areas where Flatiron could continue to refine its approach in the UK, together with specific recommendations for each area. Recommendations included:

- Creating a Flatiron UK PPIE Group and roles for patients in research oversight and data use
- Explaining to patients and the public the value to Flatiron of UK data
- Providing clear examples of how Flatiron could help the NHS and patients, bringing the value to life
- Working with patients to better define and communicate a patient view of 'fair value'
- Suggestions about communicating with patients at any NHS trust that chooses to partner with Flatiron.

Based on their engagement with the DATA-CAN patient group, Flatiron is adopting many of these recommendations. In direct response to our patient voices, Flatiron is working with a patient advocacy organisation to develop and refine their PPIE programme and are recruiting an internal PPIE group.

The framework employed to assess the value proposition in this project was designed to be reusable, and afterwards the patient members undertook further work to produce an “assessment grid” to help any similar approaches from other commercial companies.

3.3 Discover - NOW

Our purpose for PPIE is to build trust and confidence in how data is used and involve and engage more people in designing and delivering research, from the individual through to the wider population. This purpose has been jointly defined by our Hub Board which comprises senior representation from Hub partner organisations across multiple sectors as well as two ‘Citizen Partners’ – lay representatives from North West London who have been involved with Discover-NOW since application. Lay representatives are also members of the Data Access Committee - one as a co-chair - the group which reviews requests to access the Discover data. Our Citizen Partners have also been involved in the development of the Milestone-2 submission.

OneLondon Citizen Summit

Discover-NOW partnered with the [OneLondon Local Health Care and Record Exemplar programme](#) to undertake one of the most progressive large-scale public deliberations in the UK about the use of health and care data. Discover-NOW committed to receiving the recommendations from the OneLondon Citizens’ Summit and where possible aligning our policies with these to ensure the public would have trust in them.

The recommendations made by the 100 public participants as part of the four-day Summit [were published in June 2020](#). These set out a set of clear expectations in relation to how a fair and productive research partnership should work, as well as how the public should be involved in ongoing oversight and development of policy relating to the use of health and care data.

Discover-NOW has progressed at pace to align with the Citizen Summit recommendations, from rigorous adherence to the ‘five safes’ in the establishment of our Trusted Research Environment (TRE), to our commercial framework where we operate a fee for access model with tiered pricing based on organisation type. In line with the expectations set by the public in the Summit with regards to transparency, we [publish a summary](#) of all projects accessing the data and have more recently launched a [summary specific to COVID-19 related projects](#).

Finally, following the Summit recommendation regarding future involvement of the public in policy development and decision making, we have established a Citizens Advisory Group (CAG) that forms part of the Hub’s governance structure and directly informs decision making and policy development. The CAG brings together a diverse group of 40 citizens, reflective of our diverse population (see Figure 3), in a set of mini-deliberations, to explore, discuss and deliberate key dilemmas facing the Hub to inform decision making and policy development.

The CAG design and delivery is supported by Ipsos MORI and overseen by a multiple sector Steering Group comprising representatives from industry, HDR UK, the NHS, expertise in data law, information governance and technical delivery as well as two Citizen Partners. This group acts as a ‘critical friend’ to ensure the

process is balanced and authentic, whilst also ensuring the CAG discussions cover the requirements of multiple sectors that the Hub serves and is able to impact and inform key Hub decisions.

Through the OneLondon Citizens Summit and the Discover-NOW CAG, 140 members of the public have been involved in deliberative engagement events over the last year that have directly shaped trusted policy in data use and access.

Empowering patients to participate in research

We are equally committed to empowering public and patients to participate more in research. One way we are achieving this is through our NWL [Health Research Register](#) - a register of consented adults that links consent to be contacted for research to the Discover data set, thus increasing the opportunities for people to be invited to participate in research. The register also responds to a consistently identified need across multiple sectors of needing to speed up the time it can often take to recruit study participants that match research protocols and provides an important Service offer for the Hub users.

Through widespread community engagement activities, ranging from presence at community events and outpatient clinics, to outreach work with seldom heard groups, to more recently increased digital engagement, the register has c7,000 people that are linked to the depersonalised data set, with a further 8,000 in the process of being linked.

We have also established a number of PPIE activities to increase involvement in the design of research projects, for example working with expert patients to [co-design a diabetes risk prediction tool](#). We have also worked in collaboration with the Health Foundation's [NWL Network Data Lab](#) hosted by Discover-NOW to ask local communities key concerns in health since COVID-19 to prioritise topics for analysis. We worked with the [NWL Community Voice](#) programme to run an online workshop with 50 North West Londoners who were part of under-represented groups, e.g. Black, Asian and Minority Ethnic groups and people living in deprived areas.

Lastly, we are building on previous successful collaborations with medical research charities to engage patients and the public in shaping specific research questions, and are in the process of formalising this arrangement with the Association of Medical Research Charities to ensure that the researchers we work with are able to easily access patient groups to engage in shaping their studies and supporting analysis.

Criteria for accessing Hub data

In designing PPIE activity our Hub teams, working in partnership with Citizen Partners as well as PPIE leads, start with impact in mind and success criteria is clearly defined (see PPIE strategy). Impact happens by design and we believe there are a number of conditions that need to be true for impact to be realised:

1. Clarity over what can be influenced, and what can't – PPIE is only authentic and impactful when participants can genuinely influence the outcome. It is important to be honest about what patients and the public can influence and what they can't, and why. This creates the environment for a more meaningful dialogue where outputs are more likely to have impact as they are developed in the context of reality.

Impact created: This condition has been at the heart of our deliberative engagement work and has led to significant impact with regards to trusted policy development (see CAG example below).

2. Senior stakeholder buy-in – Impactful PPIE is reliant on the support of senior stakeholders. First, it requires courage to be open about the dilemmas and difficult decisions stakeholders face and the constraints they work within. Further, meaningful and impactful PPIE requires investment – both in terms of the time and resource it takes to do this well, and the funding it requires.

Impact created: We have worked with stakeholders across our region to build the case for meaningful PPIE which has resulted in more impactful outputs (Figure 1). Previously sceptical stakeholders are now some of the strongest advocates for PPIE because of activities like our deliberation work.

3. Citizens involved in the decision-making process – Discover-NOW’s governance ensures patients and the public are able to genuinely inform and participate in Hub decision making.

Impact created: The Citizens Advisory Group (CAG) is part of the Hub’s governance and all CAG recommendations are received by the Hub’s Board which includes two Citizen partners; ensuring citizens are not only informing decision making, but are involved in it too.

4. Accountability - It is important that those participating in PPIE activities understand how their contributions have delivered impact, and importantly if recommendations they have made haven’t been taken forward, why this is. This ensures an ongoing and authentic relationship with our populations built on openness and trust.

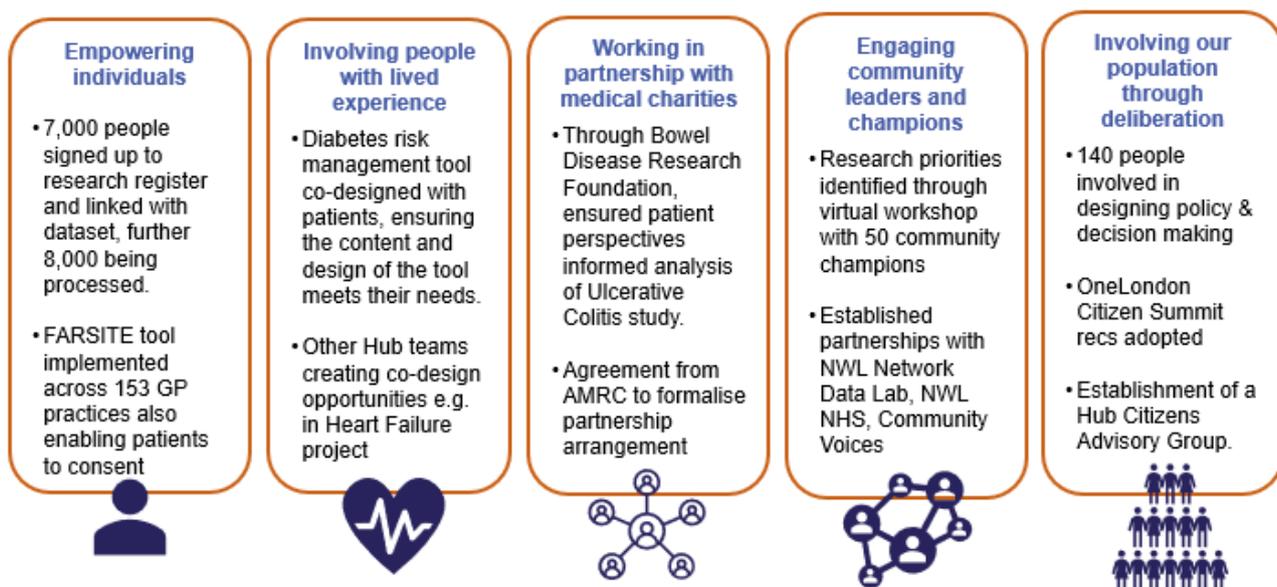
Impact created: Continuous feedback is provided to PPIE participants at all levels of the Hub, including between the Hub Board and the CAG. This has resulted in ongoing engagement from patients and the public e.g. OneLondon Citizen Summit members are now members of our CAG, and over 75% of CAG members have expressed an interest in getting involved in other PPIE opportunities.

5. Awareness – It is important to share the success and impact of PPIE activity to not only demonstrate the value that meaningful PPIE can have, but also to raise awareness amongst our broader population to encourage increased participation in the future.

Impact created: PPIE work is shared in a number of ways ranging from a [dedicated area on our website](#) connected to social media networks, to our Citizen Partners engaging their community and PPIE networks and [sharing blogs](#). This has not only led to more patients and the public participating in Hub PPIE opportunities but has also led to Discover-NOW PPIE activity having impact beyond the Hub.

Through creating these conditions, Discover-NOW has delivered many impactful PPIE activities over the last year; Figure 1 provides some examples.

Figure 1: Impactful PPIE activities from Discover-NOW



One prominent area of impact has been our deliberative engagement, starting with the OneLondon Citizens Summit and more recently with our Citizens Advisory Group (CAG). The first of the CAG’s mini-deliberations took place earlier this year. It explored ‘What conditions need to be in place to support trusted access to data for non-NHS organisations in the Trusted Research Environment?’.

Discover-NOW does not currently provide direct access to depersonalised data for these types of organisations. In developing the TRE we wanted to understand public expectations in relation to direct access to guide TRE development and shape trusted policies that underpin data access.

Over two 3-hour virtual workshops, the 40 public participants reflective of NWL's diverse population (see case study 1), supported by experts from multiple sectors, explored and deliberated criteria for accessing Hub data and the controls they believe should be in place to ensure safe access for non-NHS organisations. The workshops concluded in the CAG forming a set of recommendations (see Figures 2 and 3) that were presented to the Board on 4 March. [See full report here.](#)

As a result of the CAG's recommendations, the membership of the region's Data Access Committee, as well as the criteria it uses to review data access requests, will be reviewed and amended to ensure they are in line with public expectations. Furthermore, the Hub Board has committed to ensuring that the CAG's recommendations relating to controls are met as part of the TRE technical development. As a result of the CAG's recommendations, and the changes in practice it has informed, the Hub will now be able to confidently increase access to data for non-NHS organisations in a way that meets public expectations and supports Hub sustainability planning.

At the other end of the *person to population* PPIE continuum, our work to empower more people to participate in research through the research register has also delivered impact. As a result of the register, over 750 North West Londoners have participated in research over the last year, including a COPD Respiratory study and various COVID-19 studies.

3.4 Gut Reaction

At the outset Gut Reaction established a Patient Advisory Committee (PAC), a central mechanism to ensure the patient voice is driving the programme. By establishing a consistent cohort of patients, who provide ongoing insight, we have enabled a richer and deeper focus on ensuring the programme is of maximum benefit to all IBD patients across the UK. As we gradually understood the complexity of the programme and its interdependencies, this initial decision to focus on PAC as our key PPIE intervention proved invaluable.

PAC had its inaugural meeting - face to face - in January 2020 comprising 15 members. Members have a range of backgrounds and life experiences, bringing diverse views. Over the year, three further virtual meetings were held, and several exchanges made via email to understand the programme, review and make proposals for change.

Understanding where patient insight could make more impact

Understanding patient and public trust in the use of big data is at the heart of the programme, the first challenge for PAC was the complex landscape of governance arrangements associated with different data sources.

Central to the programme is the long-established, nationally important NIHR BioResource with its own governance structure, policies and procedures. The BioResource is the primary point for researchers to access Gut Reaction data.

From the outset, leaders across the programme, especially the IBD BioResource were actively engaged with PAC, presenting, and discussing their part in the programme. This ongoing dialogue identified current involvement of patients in the data access application (DAA) process, as an area to strengthen. Implications of change were significant, they would impact on all BioResource arrangements, not just DAAs for Gut

Reaction. The NIHR Bioresource has to date received ~75 DAAs (~28 in 2019, ~31 in 2020 and 8 in 2021, 13 for IBD).

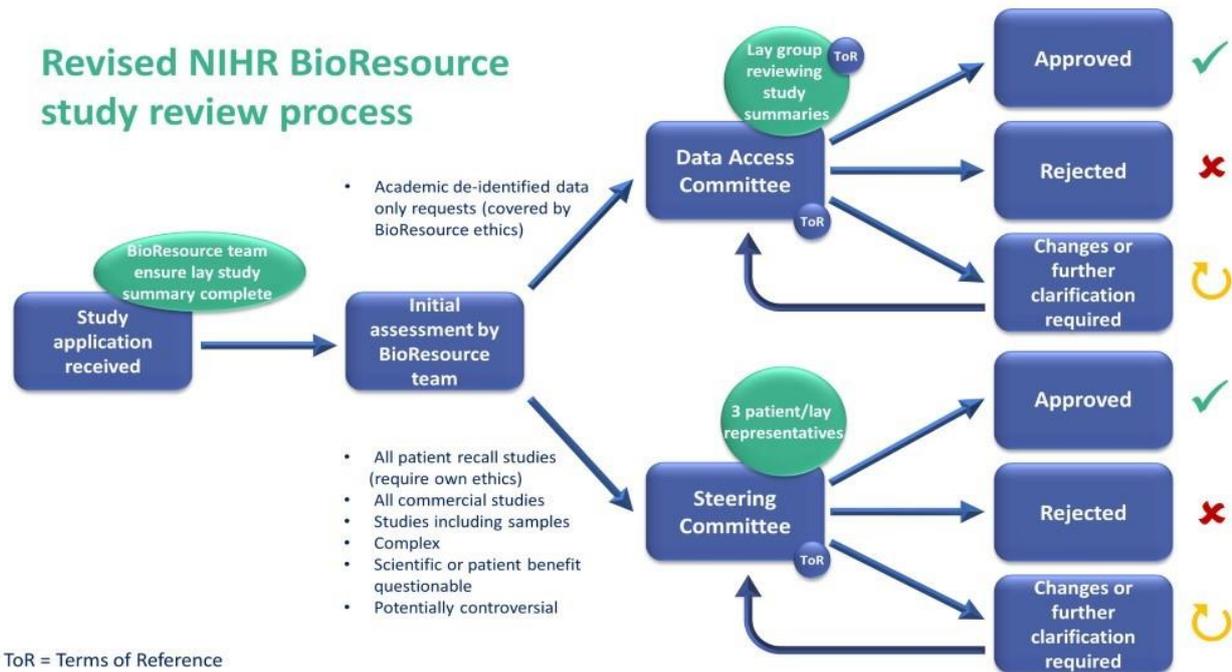
Current procedures for DAAs were made fully transparent for PAC to assess. PAC were then asked to make recommendations for more visible and robust patient involvement in decision making. Discussions were informed by wider public debate and intelligence shared by HDR UK.

What difference has it made?

PAC recommendations included:

1. Strengthen patient presence on the Steering Committee which considers the most complex/potentially controversial applications, and the Data Access Committee which considers simpler applications from academic institutions. Increase the Steering Committee from one patient to three, including one with IBD.
2. Undertake an audit of existing decisions made to send data access requests to either the Steering Committee or the Data Access Committee, which had no patient input. This would enable PAC to review decisions and provide a benchmark for the new system.

The recommendations were reviewed by NIHR BioResource in September 2020 resulting in broad agreement. Summary of the new processes below:



The involvement of PAC has resulted in a strengthened engagement with patients in the use of their data, not just for the Hub but across the BioResource - a major player in the UK's biodata infrastructure. PAC have played a vital role in changing the culture of health data research conducted through the BioResource by encouraging greater involvement in procedures in place for data access. We are now working on how the impact of these changes can be measured once implemented. These changes are already having an impact beyond the hub and will be a legacy benefit for years to come.

Initial impact and changes

It is expected appointments for lay representatives on the BioResource Steering Committee will be made in the next few months. A key area highlighted by PAC was the accessibility of proposed research lay summaries, which are often written in scientific language. All lay summaries will now be reviewed by lay representatives, and data access is not approved until the summary is accessible. Over the next few months, there will be a retrospective audit/review by PAC of all past DAAs to date.

Deepening patient insight into data access

The first studies looking to access data from Gut Reaction are now being formulated and shared with PAC for feedback. We will be working on training and support for PAC members to assess proposals robustly and consistently, potentially standardising the material and timetabling for such assessments.

One element of the training will be a modular programme looking at PPIE in health data research. It is intended to support people who undertake this essential work to achieve greater impact. We believe it will be the first open-access training of its kind and will be hosted on the HDR UK website available to all beyond the lifetime of the programme.

Additional PPIE work

Gut Reaction continues to develop its wider engagement with the creation of an animation making a complex programme accessible to all. This will be key to a renewed effort to broaden our engagement and will underpin a focus on increasing diversity in our work.

3.5 INSIGHT

INSIGHT has embedded Public and Patient Involvement and Engagement (PPIE) across its decision making and progress tracking groups, including empowering the public and patients with real, active roles in deciding who should get access to data and for what purpose. This is epitomised by the INSIGHT DataTAB, now widely recognised as an exemplar of best practice of PPIE in data governance, and is the focus of this in-depth case study. The purpose and role of the DataTAB and wider PPIE strategy has been jointly defined and key learnings are shared within the hub and further afield. The DataTAB is central to the function of INSIGHT. Within the hub the DataTAB Chair is part of the INSIGHT leadership team where he provides weekly updates, and the INSIGHT Director reciprocally attends all meetings of the DataTAB to answer any queries regarding INSIGHT.

A gated and safe process – the DataTAB in action

As implemented, the DataTAB's tasks include:

- providing recommendations on applications to access the data made available by INSIGHT;
- developing and iterating the criteria for assessing the applications;
- providing feedback on the overall procedure access requests.

When INSIGHT receives an application (Fig.1), it assesses whether it is legally compliant and contractually/technically feasible. Cleared applications are forwarded to the DataTAB to be assessed on DataTAB-

developed higher-level, unquantifiable criteria. The DataTAB will discuss and arrive on a consensus, advising whether the data request access should be:

- granted;
- granted with further conditions;
- deferred pending receipt of additional information or clarification;
- denied.

Capturing the voices of many through diversity and inclusion

We sought assembling a group that was diverse, with broad expertise whilst being small enough to facilitate cross-learning, and sufficiently agile to swiftly evaluate innovative data requests. After searches in PPIE and our networks, we invited ten individuals to join the inaugural DataTAB who attend as citizens, lending their personal experience instead of a particular institutional representation. We mix members from diverse communities, sectors, and dimensions (ethnic, age, gender, socio-economic, personal experience of sight loss, professional, academic, clinical, industry, policy, etc., Fig.2) ensuring a safe, inclusive environment and a spirit of shared-learning. The DataTAB members' bios are [publicly available](#).

Designing and delivering, together

We interviewed each member of the DataTAB and gathered their opinion on what constitutes responsible sharing of data into a draft list. These were closely aligned with the opinions of a separate workshop with a larger independent group of public and patients (case study 1).

The DataTAB first convened in September-2020 to deliberate and select on a set of access criteria which achieved appropriate scrutiny whilst being feasible, efficient and scalable - resulting in a focused set of eight criteria matching the members' collective opinion. The DataTAB successfully conducted its first full assessment in January-2021. Having reached a consensus (Fig.3), DataTAB's final recommendation, grading and comments were formally submitted back to INSIGHT's data controllers for a final decision. Subsequent feedback from the DataTAB is sought to continually improve this process.

The DataTAB is seen as a powerful exemplar of how patients and the public can effectively be involved in the design and delivery of governance on access to health data. This is underpinned by trustworthy, transparent processes and real-time communication: through regular ODI [blog series, podcasts, and reports](#) (Fig.4) we have shared our learning achievements and challenges directly with the public.

Our impact is evidenced by the DataTAB being showcased at events seeking to raise awareness and set a high standard for responsible sharing of patient data, including the [ODI 2020 Summit](#) (>1000 attendees; 72 nations, Fig.5), and the King's Fund '[Trust and Transparency in Health care](#)' (>1000 attendees).

Sense about Science

Sense about Science is an independent charity that 'champions the public interest in sound science and ensures evidence is recognised in public life and policy making'. They have a strong track-record in tackling contentious areas, an ideal partner as INSIGHT explores attitudes to the use of data, particularly in the context of industry involvement. *Sense about Science* brings independence and access to a more diverse group of stakeholders beyond those who usually engage with PPIE activities.

Key components:

Six deliberative workshops with diverse stakeholder groups - each of patient, public and practitioner (research/clinical) with participants chosen to reflect a broad and diverse population. We captured participants' understanding, questions and concerns about large-scale health data initiatives and also specifics on our messaging and Frequently Asked Questions (FAQs).

Co-creation and user testing workshops of prototype materials: Learnings from the first workshop were fed back to the wider INSIGHT team to improve our understanding and the participant feedback used to develop and refine our outward facing materials, e.g. the INSIGHT prototype website. New attendees representing our three different groups were recruited to provide a fresh, unbiased view with emphasis on: FAQs, 'INSIGHT Eye Data Journey' graphic, and 'Responsible Use of Data' (includes the 'five safes').

Feedback on changes to participants and assessing impact: Workshop feedback was incorporated into the website and then went back out to participants. The website is now live. We continue to assess its impact by measuring its hits and usage, as part of ongoing user testing and development.

A key theme raised at all our workshops has been around our industry partners. This programme provided a deep exploration of this issue, providing guidance on how we communicate these relationships more effectively – with direct relevance to INSIGHT, and to the wider UK policy around use of health data and engagement with industry. The value of the INSIGHT-*Sense about Science* programme has been that it has enabled public and patients (including those who do not usually engage with research) to openly explore these concerns, to challenge the INSIGHT team and for the INSIGHT team to learn iteratively as we have refined our understanding and communication materials accordingly with cycles of feedback. We are committed to doing all we can to hear, understand and address concerns and questions, and to share this learning. We will strive to address them in a transparent way across our communications where appropriate while humanising the partners' involvement where possible.

3.6 NHS DigiTrials

We have worked in partnership with people from diverse backgrounds and perspectives in a variety of ways. These include embedding public contributors on the NHS DigiTrials Board, ensuring that public/patient advice influences the development of NHS DigiTrials services and communications about them, and co-producing events and materials in partnership with the 11 members of our Co-Development Panel. We established and integrated PPIE across all Hub activities by recruiting two PPIE representatives to the NHS DigiTrials Board thereby putting public inclusion at the most senior levels of decision-making.

Activities and achievements

The panel were initially recruited to advise on the development of the feasibility service and communications about it. Their input included identifying the need for a short video which explained the concept of a clinical trial as a prerequisite to generating understanding of how the service would work. The [resulting video](#) has been used on the NHS DigiTrials website and at online public events.

Input into the outcomes service focused on discussing issues relating to release of data which might be associated with particular individuals. For example, questions were asked about the security of data and ensuring that data are only released to appropriate researchers.

Throughout October / November 2020, the panel formed dedicated work streams to design and deliver three online public engagement events, and to co-create materials including a [short animation](#). A field

research agency recruited 104 members of the public to ensure that those who attended were a representative sample of the adult population of England.

These events were designed to promote transparency and start the process of building public trust and confidence in the NHS DigiTrials hub. Our key aim was to identify public needs and concerns about the use of health data for clinical trials, by further interrogating some of the topics that were raised by the CDP in earlier sessions.

The events achieved a high level of engagement and generated significant data to inform service development and identify areas for further investigation, namely:

- key factors influencing public engagement with clinical trials, such as barriers faced by specific communities
- mechanisms for attitude change, including communicating via trusted sources
- issues around the use of patient data, consent and security, prompting clear explanation of data security processes
- levels of knowledge needed to improve propensity to participate.

Work with the CDP and the wider public highlighted the need to ensure that the voices of under-represented groups are heard, such as those from the BAME and LGBTQ+ communities, and people with learning disabilities.

Universal design for greatest impact

Outreach events with specific audiences, including specific sessions with those identifying as LGBTQ+ and those with learning disabilities, have used materials developed with the CDP to build upon previous work and delve deeper into barriers and incentives to participation. We are improving our understanding of why potential trial participants make certain choices, how thoughts and ideas vary across different audiences and how well certain ideas are understood using a universal design framework – talking to the most disadvantaged and marginalised groups, so that NHS DigiTrials services are designed with the under-served in mind, thereby giving us the best chance of getting it right for everyone.

Successful recruitment for a clinical trial depends on the quality of the relationships between researchers and participants. If NHS DigiTrials can make the entire process run more effectively for researchers from multiple sectors and potential participants from all backgrounds, then we will have taken big strides towards building trust and participation. By setting and meeting expectations about use of data, focusing on excellent communication, and creating inclusive processes for participants, we are creating the greatest possible impact by making clinical trials truly inclusive.

The expertise of our CDP has been used to develop events for a larger group representative of the population of England, and to complement this work with investigation of the concerns and drivers of seldom heard communities. This is enabling us to ensure that barriers are not inadvertently built into the NHS DigiTrials services by providing practical recommendations for the Identification and Communications services, in particular. Members of the panel are working directly with the Identification and Communication Services development team and potential service users to scrutinise outputs from extensive user research conducted among the research communities.

Influencing communications

Members of the panel form an important component of an editorial board alongside NHS Digital's Head of Clinical Trials, members of the NHS Digital communications and web team, and researchers. They have developed content, reviewed platforms and materials, and contributed to a complete redesign of the hub website. The panel are developing communications standards for the programme and core narratives for use across all communications with the express aim of ensuring "all our content will be factual, informative and able to explain often difficult messages in ways that are understandable and engaging."

As well as informing our policies and service design, the PPIE programme has significantly impacted NHS DigiTrials' organisational culture. Staff have reported gaining a greater understanding of patients' experiences and an appreciation of the problems they face, hence reinforcing motivation to develop a truly inclusive service.

3.7 PIONEER

PIONEER has established and integrated PPIE across all Hub activities and enabled patients and public members to define the purpose and role of their involvement in the Hub.

The Data Trust Committee

The DTC is made up entirely of lay members through an application and interview process. As part of a Stonewall diversity champion, Disability confident employer and Athena Swan Silver awardee, PIONEER recognises the need for diversity, equality and inclusion. To ensure the DTC reflects the community it serves, we mapped local demographics and targeted recruitment to specific community groups. The DTC includes BAME members, patients with long term conditions, a carer, a person living with disability, younger members (aged 17+) and a member who self-defines as LGBTQ+. This targeted recruitment was supported by and led by patient representatives. Training resources have been co-built with DTC members. These include accessible materials to remove barriers for inclusion and meet member needs.

DTC members have worked with PIONEER to define their Terms of Reference and Standard Operating Procedures. A diagram of the DTC review pathway is shown in Appendix 2.

In brief, all data requests under-go due diligence and risk assessment, based on the 5 safes, which is compiled into a risk register. The risk register and full Data Request Form (DRF) is reviewed by the DTC. Specific and independent expert support is provided as needed but as non-voting roles. At fortnightly meetings, each DRF is discussed in full, and the DTC vote on data access. Unanimous decision-making is preferred, but 80% support is required for data sharing. A DTC-decision not to support data access is regarded as binding. For example, the DTC declined to support an American Health Insurance data request.

This pivotal role means that members of the public are involved in every access decision and each and every project reported in our milestone documents has their support. The DTC review lay summaries which appear on our website, to ensure the public can understand our work.

The DTC have co-written a protocol paper for PIONEER, to share our approach and learning (under review at BMJ Health and Care informatics), with all DTC members as authors.

Discussions with Small and Medium Sized Enterprises (SMEs) highlighted a lack of awareness of the benefits of PPIE interactions to their products or services, thereby giving them great interest in including PPIE.

To meet this commercial sector need (especially SME), PIONEER and the DTC have co-created a PPIE best practice document to be shared with requestors. This highlights the benefits of meaningful PPIE to the Data Requestor, and signposts mechanisms to interact. We are testing the first draft with end-users, to be developed further pending feedback.

To meet academic sector need, PIONEER is piloting monthly drop-in sessions, where researchers can discuss PPIE within data requests, with introductions to relevant patient and public groups.

This PPIE guidance and support is available to any requestor but can be tailored by sector and flexed to meet different sector requirements. For example, the DTC have agreed to work with an academic network, to discuss how to build a PPIE framework for a natural language processing project and PIONEER is working with a large technology company, to build patient and public feedback into an acute dashboard project.

Public events

Initial consultation with patients and public members suggested that knowledge of secondary use of health data and the National Data Opt-out (NDOO) was low. The PIONEER patient/public group (PPG) identified this as a priority, defining, co-creating and co-delivering a series of events to explore and enhance knowledge, measuring the impact of our PPIE activities across sectors.

Measured Impact:

1. >400 registered for the events. Feedback from the webinars has shown that:
 - i. 93% found the series very informative or informative
 - ii. 86% were very satisfied or satisfied with the level of audience interaction
 - iii. 93% felt the content was appropriate for a public audience.
2. Pre/post questionnaires demonstrated increased knowledge in attendees.
3. Many registered to receive the PIONEER newsletter.
4. All videos freely available on our website, with subtitles now added to remove barriers to inclusion; <https://www.pioneerdatahub.co.uk/videos/>
5. Colleagues from other HDR UK Hubs and similar research programmes have sought our advice in running similar events.
6. PIONEER is working with its public members to develop a new offering which can flex to different sector requirements.

Appendix 4 – Patient and Public Involvement and Engagement plans

The Hubs’ refreshed PPIE strategies, outlining their visions on how they will continue to deliver meaningful PPIE and strengthen public trust in health data research.

4.1 BREATHE

Our key PPIE strength is having two embedded Lay Leads on Executive/Scientific Committees and External Advisory Board. Lay Leads have been key to shaping many hub activities and a continual focus on public benefit and transparency. We have critically assessed our work with an evaluation workshop/survey and revised our PPIE strategy/plan.

Delays in expanding PPIE resource have led to limited input from more diverse groups. We are now able to address this, expand our lay advisory group, and engage with seldom heard groups.

Vision: Strengthened public trust and confidence in BREATHE and health data research

Aims:

- To keep patients and the publics at the heart of all we do, ensuring PPIE informs all aspects of BREATHE’s work.
- To role model, promote and support good PPIE.

BREATHE has adopted the Health Data Research UK Guiding Principles for Involvement and Engagement.

Aligned with those principles, BREATHE highlights the following key commitments to:

- Develop meaningful PPIE
- Give fair attribution and acknowledgement
- Evaluate impact of and improve PPIE activities
- Ensure Lay Advisors are compensated for their contributions, and have clear role descriptions
- Collaborate with other Hubs to increase opportunities for engagement with ‘seldom heard communities’.

Key involvement areas

In existing panels, we will:

- Continue to: work in partnership with Lay Leads in all key decision-making and data access reviews.
- Work with Third Sector partners, and network of academic-PPI groups.
- Continually evaluate and improve.

In expanded activities to reach new audiences we will:

- Offer opportunities to engage in a variety of formats to support as diverse a community as possible, via our 100+ ‘Curiosity Group’. Members can choose how they wish to engage – from being informed with newsletter updates, through offering advice via email to attending discussion groups.
- Develop guides to support projects to embed PPIE.

Key messages through our engagement programme

A webinar series (with PPI pre-meetings to inform content, illustrations from each webinar offering lay-friendly summaries) and supporting website 'Knowledge Hub' (information, resources, case studies) will share key messages on six priority topics (see Delivery Plan) aiming to improve openness and transparency, building public trust/understanding.

Governance: Lay Leads embedded within all key decision-making committees.

Impact: Impact log used to track involvement. Lay Leads, working with BREATHE's PPI Coordinator, are adopting PiiAF (Public Involvement Impact Assessment Framework).

Working Together: Further development of our diverse 'Curiosity Group' will bring wider perspectives. We will work together within this group to adapt roles to maximise meaningful involvement.

Inclusive Opportunities: Develop and monitor our diversity/inclusion, engaging with seldom heard communities. Every effort made to minimise barriers to inclusion.

Support/Learning: Tailored approach to involvement offers a wider range of supported opportunities.

Communications: BREATHE website showcasing PPIE successes, impact and lessons learned; also promoting opportunities and educational materials for patients, public and researchers.

We currently use a standard PPI activity log. An emphasis on end-to-end involvement means narratives are captured to illustrate impact (i.e., where voices are listened to, suggestions made, changes implemented, barriers challenged). Our Lay Leads are exploring a more formal method to capture the impact of PPIE, evaluating a number of tools (e.g., GRIPP2, PiiAF, EDGE).

Risks

Limited time/resources pose a threat to delivering robust PPIE.

Potential to damage the very trust that we are seeking to build if we don't do this right.

Mitigations

We will continue to work with partners and other Hubs to efficiently utilise PPIE resources.

We strive to build trusting relationships and to have open and transparent dialogue with publics and patients.

4.2 DATA-CAN

Introduction and Purpose

Our funding proposal included a clear commitment of PPIE involvement across all elements for DATA-CAN, built on national and international best practice.

Working with an initially small patient panel, we ratified our core PPIE principles within a month, produced a comprehensive PPIE strategy and recruited a diverse (gender, age, ethnicity, geography, gender and

experience) 13 member PPIE Group within 3 months with an appointed patient Chair. Our PPIE Coordinator and the PPIE Lead work closely with the PPIE Group Chair to support the Group.

After six months we surveyed our PPIE members to understand how our PPIE approach was working. The results were very positive, though it was clear that a few of the less-experienced PPIE members had struggled in some instances. We have put steps in place to address this and are keeping it under review.

As part of our bid, we committed to implementing the outcomes of the public review undertaken by OneLondon. It was positive to note that the key recommendations and practices in the report were already being adopted within our programme.

Vision and Aims

We already follow best practice described by HDR UK, and from the outset adopted the principles of transparency developed by the use MY data patient movement.

Our vision is to further develop our PPIE members as experts in their field, to become the voice of DATA-CAN and exploiting their documented 'reach', collectively linking to around a hundred other patient, data and research organisations.

Our core PPIE learning materials form the basis from which other Hubs may produce similar disease-specific modules, whilst collectively producing the common core modules, such as Information Governance, data access, commercial links and partnerships.

Approach

Involvement

Our PPIE voice is pervasive across DATA-CAN, with members working alongside our Commercial, Data and Communications functions. We will continue to operate in this way, building on what has already taken place:

- Production of the PPIE Strategy, and Terms of Reference for the Group
- Recruitment of staff, including Chief Operating Officer and Head of Communications roles
- Co-production of the DATA-CAN website content with communications team
- Authorship of blogs, articles, data case studies, plus patient quotations on press releases
- Review of all commercial contracts
- Active participation in Steering Board meetings, including presenting updates from the PPIE Group
- PPIE subgroups including for Data Services, Comms and Commercial teams
- Providing support for other Hubs – Better Care North, Surgery hub
- PPIE Training Programme
- Contributed to CRUK in their future TRE plans.

Engagement

Leveraging the 'reach' mentioned earlier and supporting the use MY data movement provides additional networks through which we can conduct our engagement activities and crucially brings external questions and a critical friend view to what we do. We have drafted a 'Lay briefing pack' to assist our members in their engagement activities.

Our links to use MY data and the National Data Guardian

We directly support the use MY data movement of patients, relatives and carers who share an interest in how data is used to improve care and outcomes.

- Our PPIE Lead is Expert Data Adviser to use MY data and two members of the DATA-CAN PPIE Advisory Group membership of use MY data and we also provide secretariat support.
- One DATA-CAN PPIE member sits on the National Data Guardian Panel.

UK Standards for Public Involvement

Our PPIE group ensures the [HDR UK Involvement and Engagement Guiding Principles](#) are embedded across the work of DATA-CAN. We believe that within this milestone report we have evidenced compliance with the UK Standards for Public Involvement:

1. Inclusive Opportunities – 12 of 13 members involved in additional DATA-CAN activities beyond PPIE Group membership
2. Working Together – through our sub-group structure and embedding within DATA-CAN workstreams
3. Support and Learning - comprehensive set of learning materials, plus 1:1 mentoring available with our PPIE lead, briefing opportunities and drop-in sessions used by members
4. Governance – two patients are members of the Steering Board, two patients on the Management Group, at least one patient on all programme subgroups
5. Communication – our PPIE team are included in design and sign-off for communications materials, including website content, press releases (which have included patient quotations) and reports
6. Impact – Still learning – we document activity which leads to influence via Impact Log.

Two quotes from our members in this regard:

“It is so good that in DATA-CAN we can identify with the National Standards as a reality and not made up for a report!”

PPIE Member1

“When I saw a slide showing the various levels in the DATA-CAN group and how PPIE members were involved at every level of this project, it gave me reassurance that my experience could be used to benefit others and my voice was indeed valuable to the group.”

Measures of Success

A key measure of success will be the ability of DATA-CAN to take innovative decisions, knowing that there is a strong PPIE guidance and support for these. Successful industry partnerships are critical and our PPIE work underpins all contracts, providing reassurance and transparency, helping to improve openness and trust in an area where the public have yet to feel entirely comfortable.

A joint paper between PPIE members, PPIE Lead and Scientific Lead is currently being written illustrating how we have achieved our current impact.

Additionally, we are developing an Impact Log which we will keep as a running document at each PPIE meeting.

Risks and Mitigation

- Our PPIE function must feel engaged, supported, listened to and trusted. Mitigation is to provide extensive support including training, mentoring and 1:1 support and budget to match.
- The overall risk is in the loss of public trust in how institutions use patient data. Mitigation is to ensure that what we do is clear, open, transparent, has value to patients and has their support.

Supplementary PPIE Strategy Update:

Publication of an academic article about practical PPIE and RWD

We will produce an article aimed for publication in a peer reviewed journal, with a focus on how we have taken best practice theory and embedded this as practice throughout our PPIE function. This will have a specific focus on the PPIE role to oversee and steer the use of real-world data and commercial partnerships.

Implementing Equality, Diversity and Inclusion (EDI) across DATA-CAN

Whilst we included an EDI approach into our PPIE recruitment, we will extend this further to ensure EDI best practice across DATA-CAN activities, including working with our partner organisations.

Refine, as necessary, our definition of ‘PPI impact’ with our PPIE Group and make appropriate changes to our impact recording mechanisms

Reflecting the need to demonstrate impact, we will continuously review and improve our mechanisms to measure impact.

Developing learning materials about commercial uses of data (with one/more commercial partner)

Building upon our existing work, we will work with commercial partners to co-produce learning materials to describe the role of commercial organisations in the use of healthcare data. This will examine the opportunities, risks, safeguards, benefits and realities of the uses of healthcare data.

Thinking about how we could connect with schools/young people

As a direct result of requests from within the PPIE group, we will be looking into partnerships or linkages which we can make to organisations which have existing footholds into communications with schools and younger people about the uses of healthcare data.

Areas of AI and data ethics

We will be exploring the areas of ethics and AI, to equip our PPIE members with information and reference materials to enable them to play an active part in discussions in this area.

Developing and support a wider network of networks

We already have strong links, and provide support to, the use MY data patient movement, and we will seek to build upon these linkages, together with other Hubs, to establish a “network of networks” of PPIE activity, specifically focused on the uses of healthcare data.

As part of the sustainability work, define, develop and deliver an “review/consultative” model for potential commercial users of data

Building on the work we did to review a commercial “fair value” offer from a commercial company (described in one of our case studies), we have already undertaken similar commissioned work for a charity and have several other similar commissioned reviews which have been requested by commercial organisations. This is potentially an opportunity which we will explore further.

4.3 Discover – NOW

This strategy sets out how we will continue to deliver our shared PPIE purpose through delivery of the aims set out in our *Person to Population* PPIE continuum (PPIE case study 1). There are several key learnings that inform this strategy to ensure our PPIE activity is impactful and inclusive:

- Ensuring inclusivity and being reflective of our diverse communities
Ensuring our PPIE activity is inclusive and reflects the rich diversity of our population is critical. We have achieved this in areas such as our CAG, however, there is more to do and this will be a key focus of our activity moving forward.
- Flexibility in approach
The pandemic has created a need to engage with our populations in very different ways. Whilst creating opportunities, this has also presented risk of exclusion of some communities.
- PPIE activity must be focused where there is genuine opportunity for influence
Ensuring meaningful engagement is being open and honest about what PPIE can genuinely influence. This leads to more impactful results and creates trusted relationships.
- Stakeholder engagement is as important as PPIE to ensure it has impact
Without stakeholder support, any outputs from PPIE activity are likely to have marginal impact in influencing decision making.

Our vision for PPIE is to build trust in how data is used and involve more people in designing and delivering research. We will achieve this through creating a range of PPIE opportunities across our *Person to Population* PPIE continuum (case study 1) which sets out five clear aims.

Guiding the design and delivery of our PPIE work are three core principles:

- Inclusive: In designing and delivering our PPIE strategy we must embrace the diversity of our population – ensuring activity is inclusive to reflect the range of diverse voices and views.
- Honest and authentic: Transparency, openness and being honest about what can be influenced and the constraints we operate within is essential.
- Working in partnership and collaboration: Partnership and collaboration is core to our PPIE approach, not only through working in partnership with our populations, but working with other Hubs and organisations to benefit from and share our collective learnings and gained insights.

Moving forward our focus will be to strengthen and further embed our *Person to Population* PPIE continuum to deliver against the five aims.

Empowering individuals to participate more in research – Building on the learnings of our community engagement, outreach work and digital engagement that have led to c7000 people joining the research register, we will continue to engage our communities to grow the register, recognising the need for

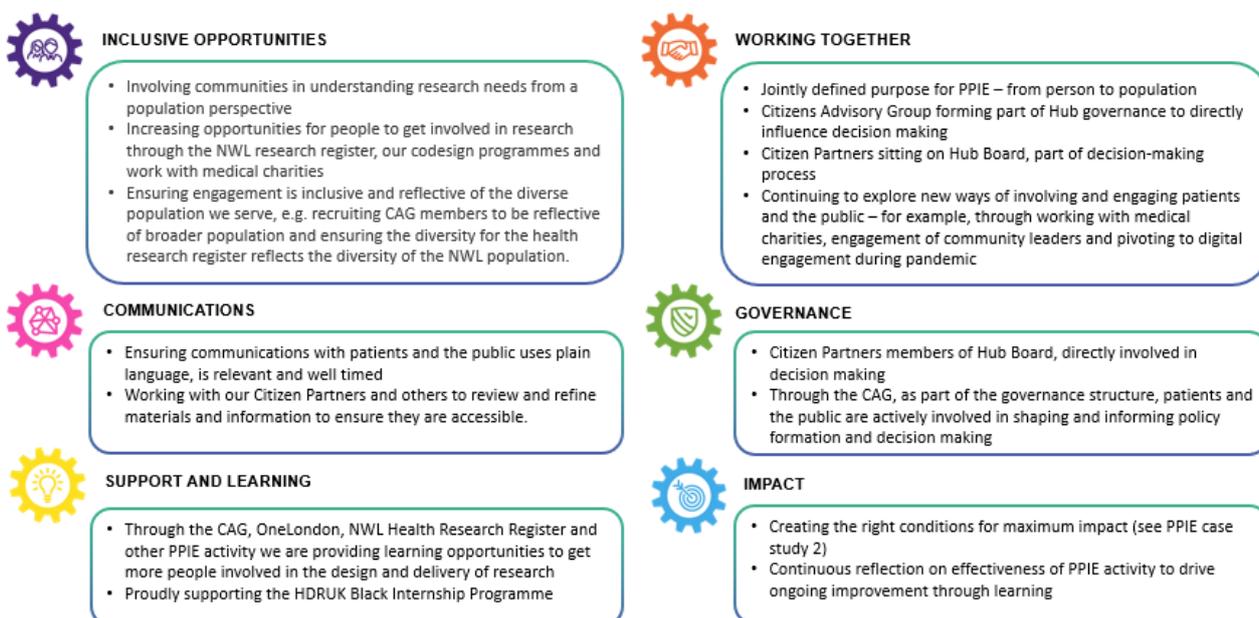
creativity and flexibility to ensure inclusivity. We will launch targeted engagement programmes to ensure the demographics of the register are reflective of the wider population. We will commence the HDR UK black internship programme and use learning from this to expand our own internship programme.

Involving people with lived experience – We will continue to involve people with lived experience as part of specific Hub workstreams e.g., developing co-design programmes to support our diabetes exemplar and Heart Failure workstreams.

Working in partnership with medical charities – We will formalise a partnership with the Association of Medical Research Charities to establish routes into charity partners for researchers trying to increase patient involvement in studies and project design.

Engaging Community leaders and champions – We will continue to work in partnership with others across NWL when engaging community leaders and champions of seldom heard groups. This provides opportunity to explore common areas of interest and avoids risk of duplication and misalignment. Further, our Citizen Partners, with our support, will continue to engage with their own networks of leaders and champions on key areas of work, for example the outcomes of the Hub’s Citizen Advisory Group.

Involving our populations through deliberation – We will continue to work with our Citizens Advisory Group (CAG) in a series of mini deliberations on key dilemmas facing the Hub. The next will be in April and will focus on benefits realisation and distribution, particularly financial, to further inform the Hub’s commercial strategy. A review into the impact and effectiveness of this model of engagement will be undertaken in the Autumn and be received by the Hub Board.



In designing PPIE activity we start with impact in mind, defining success criteria and how impact will be measured. Below are some examples of the success criteria we have defined for each aim of our *Person to Population* PPIE Continuum.

Empowering Individuals to participate more in research

- Annual growth of 10,000 registrants
- Diversity of register is reflective of wider population across key demographics e.g., age, gender, ethnicity
- Establish internship programme

Involving people with lived experience

- Increased opportunities for people with lived experience to be involved in codesigning studies and projects

Working in partnership with medical charities

- Increased reach of opportunities for patients and carers to be involved in research design and delivery through building relationships with medical charities

Engaging community leaders and champions

- Strengthened relationships with community leaders and champions to generate more awareness of PPIE opportunities
- Increased diversity of voice through community leaders and champions who represent the views of seldom heard and under-represented communities.

Involving our population through deliberation to design trusted policy

- Generating ongoing opportunities for our diverse population to deliberate policy dilemmas facing the hub.
- Recommendations generated through deliberation translate into policy and inform decision making.
- Recommendations have impact beyond Discover-NOW Hub, e.g. in other Hubs.

No.	Risk	Mitigation
1	COVID-19 impacting on rates/ methods of engagement.	Pivoting to virtual engagement where possible but being mindful of risks of digital exclusion e.g., all CAG participants offered hardware (tablet) and technical support (wifi connection and training) to avoid risk of digital exclusion.
2	Attrition of CAG members in between deliberations.	Offer a range of opportunities to be involved with the Hub in-between deliberations. Regular communications on how their inputs have led to impact.
3	National policy direction misaligned with public expectations	Challenge / influence national position with evidence of considered inputs from public deliberation. Ensure continuous stakeholder engagement; full transparency over population expectations.

4.4 Gut Reaction

Key successes:

- The Gut Reaction Patient Advisory Committee (PAC) is active, engaged and meets regularly
- Involving PAC helped to identify areas to strengthen patient input in data governance notably in the re-design of the Data Access Application (DAA) process - impacting beyond the Gut Reaction programme
- PAC input increased the accessibility of our [Animation](#) as an easy-to-understand overview of Gut Reaction
- Patient representatives were closely involved in procuring training resources for PPIE

Things to improve:

- Create opportunities for more engagement across all hub workstreams
- Support PAC contribution with clearer objectives, a platform for sharing resources, and more advanced planning
- Focus on increasing diversity and inclusivity in our work

Vision and Aims

Our vision is to deliver meaningful and effective PPIE by working in partnership with patients across the hub, developing proactive and lasting collaborations to maintain public trust in the responsible use of data.

Our aims:

- Ensuring PPIE is an integral part of Gut Reaction, adding value to outcomes of its work
- Achieving a gold standard approach to PPIE and inclusivity, based on [NIHR Standards](#)
- Continually evaluate and improve our approach

Approach

Patient and Public Involvement

Patient representatives from the hub Core Team and our PAC have helped to review and revise our PPIE strategy. It will strengthen their role in embedding the value of the patient voice across the programme and enable them to raise issues, ideas and areas for focus, as equal partners.

Inclusivity

We will:

- Work with PAC, patient partners, other hubs, and relevant organisations to engage diverse communities, particularly those known to have lower trust in projects such as this
- Widen opportunities for engagement, creating a network anyone can join
- Develop an Equality, Diversity, and Inclusion (EDI) Policy that:
 - fosters new approaches and activities tailored to the needs and preferences of identified groups
 - generates and uses EDI data to inform recruitment to PAC and our new network from underrepresented groups
- Review our approach to reward and recognition for involvement

Integration

To ensure patient voices are part of decision making across the hub, we are planning a workshop for all workstreams, including PAC, on progress to date and plans for the next phase of the programme. The aim is to enable a dialogue on where PPIE can add value across the programme. We will build on this with scheduled updates from each workstream to PAC, to ensure ongoing opportunities to ask questions and challenge.

Training and support

We are developing new training materials to upskill and empower PPIE representatives which will be informed by learnings from PAC members. Training materials will be publicly available on the HDR UK

learning platform as a resource to support patient engagement with health data research anywhere. 1:1 calls with PAC members will seek to increase engagement and understand needs for further support.

Reporting and learning

We are developing a review process to evaluate and share good practice to build trust and confidence in the use of data for research. Key to this will be a central record for PPIE feedback and subsequent actions across the hub - helping to identify themes and track progress. PAC have also asked us to support opportunities to meet without the project team (facilitated by Crohn's & Colitis UK) to feedback and identify areas for involvement and improvement.

UK Standards for Public Involvement

Inclusive opportunities

Inclusivity is central to our updated strategy, speaking to communities not represented on our PAC and offering alternative options for involvement. We will use new communication channels, such as a webinar series, targeted digital advertising, and media activity to engage a more diverse audience.

Working together

Following feedback from PAC, our strategy works towards ensuring PAC are equal partners in the hub. Our Gut Reaction workshop will give PAC a better overview of the programme. Our new resource centre will enable PAC to work together, and with us, more effectively.

Support and learning

The new PPIE training, being developed by an external specialist, aims to empower PAC and we will be collecting data to evaluate its impact. 1:1 phone-calls with PAC will enable us to better support them, and our new feedback tracking will ensure a full cycle of improvement and review.

Governance

Our updated PPIE strategy includes involvement across the hub, PPIE training for hub colleagues, and opportunities identified by workstreams and PAC to ensure the PPIE is inclusive and following best practice. We will continue to support patient involvement in the data access application process for the BioResource.

Communications

We will use plain English and follow guidance from [Understanding Patient Data](#) on the best language to use when talking about data. We encourage communication between PAC and all workstreams to ensure learning and opportunities are shared across the hub and the wider HDR community. We will draw upon various communication channels to get people involved in research, including news items, e-newsletters, social media activity and media.

Impact

We will use blogs, case studies, social media, and a webinar series (all developed with patients) to highlight the impact of PPIE. We also record training, PPIE feedback, and use the national standards to gather broader patient feedback.

Measures of Success

Recording all PPIE feedback and activities - including objective (e.g. number of activities led by each workstream) and subjective measures (reflective reports from PAC)

- Case studies assessing impact of PPIE on participants and research
- Measures on reach for all engagement and involvement opportunities
- Crohn's & Colitis UK are developing a tool to measure PPIE impact
 - A key area to use this will be to measure the impact of increased patient involvement in BioResource data access applications, given the PAC's involvement in this process

Risks and Mitigation

- Recognition and reward for PPIE. Informed by best practice, we will develop options to support engagement and inclusivity whilst maintaining fairness and transparency mindful of the hub's context.
- Patients involved in health data research are not fully representative of the population affected by IBD. We will use existing data to identify where we could do more and develop an EDI strategy to target activities to increase diversity and inclusion.

4.5 INSIGHT

From the outset, INSIGHT has been characterised by its commitment to patient and public involvement and engagement (PPIE). Patients and public are actively involved in key decision-making processes such as data access, communication, and governance.

Key Learnings:

1. Working together: we developed our PPIE strategy jointly with patients and public participants, and with cross-sector representation – NHS, academia, Pharma, Tech, and the charity sector – based on *Intent, Involvement and Integrity*. This integrated approach has helped ensure our PPIE strategy covers multiple-sector requirements, noticeably in how we engage effectively and ethically with industry.
2. Trust and transparency: we demonstrate international leadership for trust and transparency in the use of patient data, notably by involving patients and public in the data access process through the INSIGHT Data Trust Advisory Board, DataTAB (Case Study 1), and our exploration of PPIE in challenging areas such as industry usage of data and the value return to the NHS (Case Study 2).
3. Never stop learning: we benefit from the collective experience and leadership of Hub partners, for example, the Moorfields NIHR BRC PPIE Team who engaged in the highly commended Moorfields-DeepMind (Google Health) Research Collaboration. Additionally, we have advanced through new partnerships, such as with the [Open Data Institute](#) (ODI) – innovating PPIE in data governance, [Sense about Science](#) (SAS) – hearing and responding effectively to diverse views on health data usage including from the 'seldom heard', and the [Bloomsbury Festival](#) – communicating dynamically to a diverse audience.

4. The power of diversity: we have gained from the wider insights and challenges that come from the diversity of our stakeholder groups, including strong representation of young people through our 'eyeYPAG' (Young Persons Advisory Group), and the DataTAB.

Going forward, we want to continue to innovate, learn and share in these areas. Specifically, we are extending the leadership roles of patients within the Hub, and we are seeking to engage and involve an ever-more diverse group of people while also ensuring that we equip them to be able to effectively engage in what can be a daunting and complex area.

INSIGHT's vision, as set out through Milestone 1, is to facilitate delivery of better diagnostics and treatments through the safe and responsible use of data.

INSIGHT's aims which are jointly developed and cover multiple-sector requirements are:

1. Align our work to the needs of patients and those at risk of eye disease. We involve patient representatives in our key work-streams and encourage co-development of new projects, such as with eyeYPAG.
2. Involve patients and the public in the governance and safe use of data for health research. Our DataTAB includes patient and public members to develop consent models and review requests for access to data.
3. Break down barriers to broaden participation in INSIGHT. We operate fair and transparent recruitment processes to roles. We continue our work for reaching seldom-heard communities to inform engagement and training activities developed through the Hub.
4. Encourage cross-fertilisation of learning with other HDR UK Hubs and infrastructure.
5. Develop novel communications and training tools informed by patient and public preferences.

Delivery Plan

Across our Hub programme we seek to involve a wide and diverse population, and we are always looking for new and exciting ways of doing this by building on the cross-sector strength of the existing Hub partnership and exploring new collaborations, such as with the Open Data Institute and Sense about Science. Our PPIE activities, communications plan and the sharing of learning through our networks are focussed on embedding citizens in the heart of what we do. INSIGHT's active projects under the PPIE strategy include:

- Data Trust Advisory Board (DataTAB): this is a core component of our PPIE strategy. It is a powerful demonstration of our commitment to ensure that patients and public are at the heart of key decision-making, and that they are fully supported to be effective in these roles. We have been delighted to see the DataTAB model being replicated in other HDR UK Hubs, and we continue to invest in sharing our learning and communicating the principles and practice of the DataTAB to the citizens in the UK and beyond.
- Bloomsbury Festival: continuing the work from autumn 2020 and its encore in January 2021, we will revisit planned face-to face activities in summer 2021, subject to the continued and safe easing of COVID-19 lockdown measures. These include proposals for an interactive street stand with an 'AI Machine' and an option for passers-by to be provided with information and then to decide whether they post it into the machine or not. We seek to use this opportunity to talk to people and gain additional insights. We also seek to expand the reach of this activity by creating

additional communications and promotional materials for wider dissemination, with the potential of replicating the same approach at other events and roadshows.

- Sense about Science (SAS) – ‘Ask For Evidence’ Campaign: our partnership with SAS enables engagement with different community groups that we ordinarily do not have access to and the ‘Ask for Evidence’ events are promoted widely. The campaign ambassadors are established, experienced and active, and they are skilled in social media (promoting talks on Twitter and Facebook) and in creating shareable content/blogs.
- EyeYPAG (Young Persons Advisory Group): building on our work with EyeYPAG in 2020, we plan to engage with the group to develop resources specifically suited to younger people. The group will guide the type of resources developed, but one area we are keen to explore is an INSIGHT animation. We will disseminate the work undertaken with EyeYPAG using a variety of communication approaches, including social media posts.
- Website development: we will work with the INSIGHT BRC Patient and Public Advisory Group through facilitated consultation to ensure the continued accessibility, language, understanding and usefulness of our website, thereby maintaining patients and their priorities at the heart of our Hub activity.
- Data journey animation development: while developing our website we created a data journey diagram that has been reviewed by patients, public and professionals in our SAS workshops. With continued patient involvement we plan to develop this two-dimensional diagram into an interactive animation.
- Podcast series: building on the success of our first podcast series, “Insights into INSIGHT”, we plan to widely disseminate the next three instalments. These again feature Dr Amit Patel (@BlindDad_Uk) and aim to ‘normalise’ all things data, and the associated risks and benefits, by looking at how big data, algorithms and AI are already embedded into our everyday lives.

4.6 NHS DigiTrials

Introduction and Purpose

We have embedded Patient and Public Involvement and Engagement (PPIE) in NHS DigiTrials through the establishment of our Co-Development Panel (CDP). The panel have ensured that NHS DigiTrials meets the needs of patients and the public through their involvement in the development of specific services, the overarching strategy, the commercial strategy, and communications materials.

The panel comprises 11 members of the public both with and without previous experience of PPIE. We have achieved a good balance of membership in terms of age, ethnicity, gender identity, and geographical location. PPIE is integrated at the highest levels of decision-making with two CDP members sitting on the Board.

Our public events programme yielded rich data on public confidence and understanding of clinical trials and has informed qualitative work with seldom-heard groups such as those with learning disabilities. Activities were co-designed and produced with the CDP.

We would like to recruit additional members of the CDP to expand the role and scope of the panel and establish more task-specific groups. We would also like to facilitate more opportunities for CDP members to be involved in other work streams. Panel members have stated that they would welcome more opportunities for personal development and this is something we would like to improve upon.

The 2021-22 activities described are currently unfunded and are dependent upon new funding.

Vision and Aims

Our vision for PPIE is that: NHS DigiTrials is a service that public and patients trust and the public are confident that researchers using NHS DigiTrials will use their data correctly.

Our aims are:

- We develop PPIE activities that are relevant, engaging and achieve their purpose because they have been developed in partnership;
- The NHS DigiTrials team understand people's needs and concerns regarding use of health data for clinical trials and this understanding is used to help shape future activities;
- Those who engage with outreach activities have greater understanding of the data the NHS holds, what these data are used for, and how NHS DigiTrials will share this data;
- Those who engage with outreach activities trust NHS DigiTrials partners and researchers to use their data correctly and are enthusiastic about the benefits of sharing data for research purposes.

We see PPIE as a dynamic process that will continue to evolve as NHS DigiTrials develops. We intend to refine this strategy and action plan with our CDP according to the UK Standards for Public Involvement.

Approach

We intend to build upon existing involvement activities by:

- Continuing to embed public/patient representatives in the governance of the project through membership of the NHS DigiTrials Board, and ensuring that PPIE remains a standing item at Board meetings
- Reviewing/ rotating membership of the service user group to ensure continued public/patient representation and opportunities for direct input into service development
- Continuing to involve public/patient representatives in the editorial panel to co-develop communications materials
- Expanding the existing CDP to allow broader representation and greater involvement
- Providing guidance on PPIE to researchers using the NHS DigiTrials service.

We intend to build upon previous engagement activities by:

- Developing a programme of public engagement events
- Continuing to identify and address barriers to engagement
- Communicating information about opportunities in ways that appeal to different communities.

We have identified the following key messages:

- Patients and other members of the public are partners in our research
- You can help researchers develop treatments for many diseases by sharing your data and participating in research
- Our processes ensure that your data are kept securely and are only shared with researchers who use them to improve health and care for everyone.

These messages will help to build trust and engagement by addressing known concerns, specifically concerns regarding access to and use of data by commercial entities, scepticism about potential patient benefits, and concerns about data security.

UK Standards for Public Involvement

We will meet the standards by:

- Building upon our work with seldom heard audiences by coordinating additional tailored activities (Standard 1)
- Continuing to provide reimbursement for time, support with travel arrangements (once permitted) and support to use technology (Standard 1)
- Continuing to offer a variety of opportunities for involvement and personal development (Standards 2 and 3)
- Conducting a skills needs analysis for CDP members and providing/ highlighting opportunities for learning and development (Standard 3)
- Continuing to use and champion the use of plain language and involving PPIE representatives in the review of materials (Standard 4)
- Sharing the difference that public involvement makes to our work by developing case studies and guidance for researchers that provides them with information on PPIE best practice (Standard 5)
- Continuing to monitor and celebrate changes that have been made to the service as a result of public feedback (Standard 5)
- Continuing to involve public/patient representatives at Board level (Standards 2 and 6).

Measures of Success

We use a mix of methods to measure impact and success:

- CDP input into service development and associated activities
- feedback from lay Board members at monthly meetings
- polls used before and after events to gauge changes in audience attitudes and understanding
- qualitative feedback from group discussions
- a log of changes made to the service on the basis of public feedback
- analytics to assess engagement with digital content
- quarterly reports which include quantitative and qualitative measures
- reports produced at key stages of the project
- interviews undertaken by an independent reviewer.

Risks and Mitigation.

Risks and mitigation measures:

- Public scepticism regarding security of their data and the involvement of commercial partners: This will be managed through clear communication of the roles of private sector partners and the benefits of working with them, and transparent communications about the steps that are taken to ensure data are provided to researchers who will use them appropriately for public benefit.
- A mismatch between service delivery and PPIE timeframes: This will be managed through detailed project planning and weekly project meetings.
- Overlap with other hubs/HDR UK projects/organisations: This will be managed by sharing PPIE plans with other hubs and discussion with HDR UK colleagues and other organisations working on health data research, such as the NIHR.

- Inadequate human resources: We intend to recruit additional resource to support this project with new funding.

4.7 PIONEER

From inception, PIONEER has involved patients and public members. This included understanding our community's hopes and fears around health data and the processes they would trust, with questionnaires and workshops with >350 public members whose diversity reflects our community.

Public views fed directly into our ethically approved protocol, focusing our scope and operational processes. The Data Trust Committee (DTC) was formed based on PPIE learning, providing public oversight for data sharing. Further recruitment has expanded the diversity of the DTC, ensuring we hear from different communities.

Our wider community engagement has focused on awareness of health data research and the National Data Opt-out (NDOO), a priority from our initial, published PPIE learning¹. Recent activity includes webinar series and patient and public workshops. Next, we want to engage with Birmingham residents with little spoken or written English. We are working with community groups to achieve this.

Vision and Aims

Our vision;

- To work alongside diverse patients and public members to ensure sharing data for research improves health, welfare and care.
- To embed patient and public voices in our structures and decisions. To increase public trust in data sharing through open conversations.

Our aims;

- To grow our collaborative PPIE strategy with patients and public members, ensuring its purpose and outputs reflect our community.
- To ensure the people we work with are representative of our community and that opportunities for PPIE are inclusive and accessible.
- To collaborate with and listen to patients and public members, so their voices shape the direction, processes and decisions of PIONEER.
- To provide patient and public oversight of data sharing decisions.
- To build an ever-increasing network of patient and public groups, building trust in PIONEER by talking openly about our work.
- To share best practice and lead by example, celebrating how working alongside public members enhances our work.

Approach

Involvement: Data Trust Committee (DTC) review all data requests

- Public views define public benefit and acceptable risks.
- A decision not to support data access is binding for PIONEER.
- Data access applications are improved through lay consultation.
- Sharing best practice across HDR UK.

Involvement: Build a wider PPIE network for more diverse involvement.

- Nuanced decision-making informed by lived experience.
- Benefit from diverse views in ongoing PPIE.
- Involved members of the public feel valued and listened to.
- Feedback into changes.

Involvement: Co-create processes and resources to transparently share PIONEER work.

- Improved public understanding of the data PIONEER holds, the research we support and why it's important.
- Increase opportunities for conversations, supporting responsible data sharing.
- Reassure that PIONEER works with lay people in decision making.

Engagement: Reflect on current engagement and broaden our reach and understanding.

- More people are aware of PIONEER.
- Increased public trust in data sharing.
- More opportunities to listen to public members.

Show the impact of our PPIE work.

- Ensures meaningful PPIE work.
- Members of the public know their contribution is valued.
- Share the value of involvement and engagement.

UK Standards for Public Involvement

Inclusive opportunities

- Early Involvement with PIONEER co-developed with public members.
- Removing barriers with accessible documents. Webinars subtitled. Leaflets in draft in local languages. Open events; recorded to "listen again".
- Equality and diversity targeted to ensure public voices reflect our community.
- Transparency with processes and outputs freely available at www.pioneerdatahub.co.uk, blogs, tweets and PPIE papers 1,2.

Working together

- Public and patients define the purpose of PPIE within PIONEER.
- Roles, responsibilities and working arrangements co-developed including terms of reference and standards.
- PPIE has influence with actions from feedback.

PPIE Support and learning

- Co-built a suite of Training and learning resources, shared widely.
- The PIONEER website signposts opportunities, lists resources, events, hospital notices and patient/public groups.

Governance

- Patient representation is embedded in all decision-making structures. The DTC review all data access requests.

- PPIE plans grow iteratively through public workshops.
- Infrastructure and resource supported by a professional PPIE team with designated budget to enable participation.

Communications

- Communicating our activities through multiple mediums.
- Inclusive and flexible methods for an accessible communication journey.
- Sharing best practice with other Hubs and organisations to share our PPIE journey, outputs and lessons learnt. PPIE papers to share learning^{1,2}.

Impact

PPIE groups defined processes to assess increasing knowledge of data research using PiiAF³.

Reflecting on PPI, DTC members decided data requests needed more PPI support.

Acting on feedback and Incorporating learning. Using a “you said, we did” approach, we are co-developing a PPIE toolkit, PPIE drop-in sessions and building links with diverse patient groups, built to meet different sector needs.

Measures of Success

We are implementing PiiAF³ to measure impact. This includes:

- Involving patient and public stakeholders in designing and conducting impact assessments.
- Considering how PPIE will lead to the impacts we want.
- Understanding that context may affect PPIE impacts.
- Designing the impact assessment considering feasibility and potential challenges.

Examples of successes include:

Change as a result of PPIE:

- DTC developed after PPIE highlighted public oversight to improve trust.
- Feedback from DTC to researchers has improved lay summaries and, through wider PPIE networks, PPIE in data research.
- Changes made to the website after user experience testing.
- Opt-out awareness campaign shaped by public focus group feedback.

Tracking people reached through engagement, especially in two-way conversations.

Risks and Mitigation

Risk: Lack of effective PPIE/ PPIE at the periphery

Mitigation: A coordinating professional PPIE team. A PPIE roadmap and evolving strategy, reviewed through public activities. Co-defining terms of reference and embedding PPIE in all PIONEER strategic and operational elements. Established network of PCIEP⁴ & PPI groups in place.

Risk: Lack of support or training in PPIE

Mitigation. Training packages co-created with PPIE members and frequently reviewed, with experts for specialist domains and lay summaries for all activities.

Risk: Lack of inclusive opportunities in PPIE

Mitigation: Active review of diversity in PPIE. Positive recruitment in under-represented groups. Ensure outputs are accessible through subtitles, large print and languages. Build stronger links with community groups to enhance engagement.

Risk: Missing impact or opportunities for shared learning

Mitigation: Co-define “meaningful PPIE impact” and review progress against agreed milestones. Outward looking communications plan which is transparent about success and mistakes.

References and glossary

1. [Perceptions of anonymised data use and awareness of the NHS data opt-out amongst patients, carers and healthcare staff](#) Catherine Atkin, Benjamin L Crosby, Kevin W Dunn, Gary Price*, Eliot

D Marston, Clark Crawford, Margaret O'Hara, Charlotte G.

S. Morgan*, Martin Levermore, Suzy Gallier, John Attwood, Stephen Perks, Alastair

K Denniston, Georgios Gkoutos, Richard Dormer*, Andy Rosser, Agnieszka Ignatowicz, Hilary Fanning, Elizabeth Sapey medRxiv 2020.09.12.20193276; doi: <https://doi.org/10.1101/2020.09.12.20193276>

2. The infrastructure and operating processes of PIONEER, the HDR UK Data Hub in Acute Care and the workings of the Data Trust Committee. A protocol paper.

Suzy Gallier, Gary Price*, Hina Pandya*, Gillian McCormack*, Chris James*, Bob Ruane*, Laura Forty, Benjamin L Crosby, Catherine Atkin, Ralph Evans, Kevin W Dunn, Eliot D Marston, Clark Crawford, Martin Levermore, Shekha Modhwadia, John Attwood, Stephen Perks, Rima Doal, Georgios Gkoutos, Richard Dormer*, Andy Rosser, Hilary Fanning, and Elizabeth Sapey BMJ Health & Care Informatics

2021;28:e100294. doi: <https://doi.org/10.1136/bmjhci-2020-100294> * = public member authors

3. PiiAF: The Public Involvement Impact Assessment Framework Guidance. See

<http://piiaf.org.uk/documents/piiaf-guidance-jan14.pdf>

4. PCIEP” Public and Community Involvement, Engagement and Participation

Appendix 5 – Hub projects

Overviews of the Hubs’ key projects with industry, academia, NHS and charity, and how they have impacted on their services and sustainability.

5.1 BREATHE

BREATHE has 12 projects, redacted for confidentiality.

5.2 DATA-CAN

No	Project Description	Sector(s) Project Covers + name of client if possible	Short rationale of any impacts to Hub services the project has generated	Short rationale if the project provided any impact to Hub sustainability
1	Building a Real-Time Data Network for responsive monitoring of cancer-related activities across the country	Health - NHS, HSCNI / Academia / Industry Large - Roche, Janssen	Improving the speed of cancer data availability for service planning and development purposes and ultimately NHS cancer patient outcomes, by sharing real-time data about UK cancer services during the COVID-19 pandemic and afterwards.	The network is aiming for sustainability either by becoming a standard NHS data service, or by attracting external funding to cover costs. Any funds raised will be split, to fund both the central operations and payments to the sites to enable their participation
2	Fair value consultancy	Industry Large	Reinforced DATA-CAN's position as a source of expertise in UK cancer data and PPIE.	Income generating project which enriched the environment for data access and paved a pathway for inward investment for the UK
3	Consultancy project providing advice on real-time cancer data landscape	Industry SME - Carnall Farrar & Pharma - Roche	Reinforced DATA-CAN's position as a source of expertise in UK cancer data and PPIE.	Income generating project which improved the knowledge about the cancer data landscape and built on DATA-CAN's relationship with an important industry partner (Roche)
4	Project to extract, curate and analyse data routinely collected during the treatment of triple-negative breast cancer patients	Academia / Industry Large - Roche	Project has enabled a deeper understanding of patient outcomes and health service utilisation, which is part of DATA-CAN's core remit. The work has generated 5 subsequent requests for similar projects so has the potential to be a key service offering.	Curating and linking data across hospital sites to answer service evaluation questions could lead to patient benefit. DATA-CAN's investment in the curation of Leeds data will build a valuable dataset, enable the project and start a longer-term partnership with Roche. Project has attracted further commercial interest with income-generating potential.
5	Creating a Cancer Data Network a network to reveal insights that can help our oncology community improve cancer care and patient outcomes, accelerate clinical research and enable innovative contracting	Industry Large - IQVIA / Health - NHS	Using data to drive improvements in the quality of care and to increase treatment options for cancer patients, through on-site cancer analytics, national benchmarking and reporting, trial matching and research services.	Automated trial matching solution enables patients to be matched to clinical trials in near real time, and the CDN enables more research through the use of privacy-protecting technology and real-world data.

6	Primary customer of NHS Digital cancer TRE for recovery of cancer services	Health - NHS Digital	Collaborating with NHS Digital to develop a cancer TRE which facilitates data access and links routinely collected data to drive research is part of DATA-CAN's core remit.	Project is building on DATA-CAN's relationship with NHS Digital, and the TRE will create extensive opportunities for income generation through research commissions from a variety of sectors.
7	Health Data for London Strategy	Health - NHS/Government - OneLondon	Proposed accelerator project on targeted screening for bowel cancer is an important opportunity for high quality monitoring and improved knowledge on effectiveness of COVID recovery plans on future outcomes. Alignment at a pan-London level could provide mutual benefit to the cancer alliances' efforts in this area.	Project will build on DATA-CAN's relationship with cancer alliances and relationships across alliances. Because it is part of a broader data strategy, start-up funding and the development of a sustainable business model is built into the delivery plan.

5.3 Discover-NOW

No	Project Description	Sector(s) Project Covers + name of client if possible	Short rationale of any impacts to Hub services the project has generated	Short rationale if the project provided any impact to Hub sustainability
1	Does North West London Chronic heart failure care in adults vary by geography and social group? A retrospective cohort study to map the clinical intervention points on a selected cohort of Heart Failure patients from diagnosis to community care.	Industry Large - Big Pharma, Academic and NHS	Generating peer reviewed publications, this analysis shows and describes comparative flows and outcomes for heart failure patients by pathway and population. Baseline and foundation study that will build out to more advanced digital pathway transformation work as described by the innovation test bed service approach outlined in Case Study 2.	Largest revenue generating project to date for the Hub. Contributing to growing the sustainability and portfolio research for the Hub and building upon our strategic relationship with Hub partner, AstraZeneca. Builds expertise in Cardio Vascular Disease in collaboration with Imperial College London.
2	CovidLife - the national survey put together by Generation Scotland at University of Edinburgh in collaboration with the Wellcome Trust to explore which of 'lockdown' measures have raised major concerns about future employment and the economy	Charity and Academic	Demonstrated the value of the NWL Research Register that provides access to a consented population to be contacted for research opportunities. We were able to respond quickly to the survey request and gain input from the consented NWL population to an important national survey that informed SAGE.	This project was the first of type to highlight the value of consented volunteers on the NWL Health research register to participate at a national survey. We demonstrated impact of having a pre-consented population with response rates being significantly higher than equivalent general public appeals for survey responses.
3	To identify NWL patients with Hypertension that wish to participate in a study that involves a basic procedure to reduce your blood pressure without medication in collaboration with industry and the Royal Brompton Hospital NHS Trust.	Industry and NHS	The NWL Health Research Register methodology enabled the identification of an eligible cohort of 128 patients and confirmed interest for 45 patients to participate in this study that had proved hard to recruit to data using traditional recruitment methods in hospital settings.	This study has highlighted the high conversion rate for the pre-consented register onto trials. It also shows the impact the register has on the ability to support and streamline patient recruitment for studies with strict exclusion criteria.
4	To engage NWL patients with chronic obstructive pulmonary disease (COPD)	Academia and NHS	The NWL Health Research Register methodology enabled the identification of an eligible cohort of 136 patients	This study highlighted the high conversion rate for the pre-consented register and the impact a register can have

	in a study of the prevalence of Respiratory Syncytial Virus (RSV), a recognized serious pathogen, in people with chronic cardiopulmonary conditions in collaboration with the National Heart and Lung Institute at ICL and the Royal Brompton NHS Trust		that met the exclusion/inclusion criteria of the protocol with 48 patients agreeing to participate.	to streamline patient recruitment. It was also first of type service for the Hub and enabled the team to develop capabilities and ways of working to deliver this service offer.
5	<p>The Networked Data Lab is a new collaborative network of advanced analytical teams working together on the use of analytics in improving health and social care.</p> <p>The lab partners are Imperial College Health Partners, Institute of Global Health Innovation, Imperial College London and North West London CCGs.</p>	Charity funded to support a collaboration between ICHP, Imperial College London, NHS	The Lab includes analysis of pertinent health topics, beginning with: Covid Shielded Patient / Clinically Extremely Vulnerable and Mental health. The analysis has helped contribute to the national prestige of the Hub and has led to insights into the shielded patient / CEV cohort for the NWL sector.	This project provides a significant investment to the Hub and an opportunity to showcase the value of the health care data with the Health Foundation and other NDLS. Through the collaboration across the Lab Network is also enables sharing and development of analytical techniques and methods.
6	With British Red Cross, this research comprised a retrospective analysis of examine Discover data relating to people who are high intensity users (HIU) to identify common triggers, circumstances, and pathways. The health and are research aimed to uncover the varied success of features of interventions, especially relating to health inequalities	Charity and NHS	The study outputs have been shared with an inaugural pan-London HIU clinical meeting and helped establish focus areas for action. The analytical approach and methods taken can be replicated for other patient cohort and studies. This study highlights an innovative collaboration between a leading charity, NHS and a data Hub.	The research will support the British Red Cross in the development of policy and advocacy positions around support for HIUs and provides an exciting collaboration for Discover-NOW with a leading charity whilst also providing the NHS in NWL with insights to inform local service development. Publication planned for Spring 2021.
7	Working in partnership with AstraZeneca, Imperial College London, NHS. This project aims to improve the delivery of T2D care by identifying T2D patients most at risk of complications, facilitating proactive interventions, enabling the development of tailored digital health solutions and allowing for real-time management solutions.	Industry Large - Big Pharma, Academic and NHS	This rapid evaluation, will provide insightful evidence on the impact of this new model of care approach on clinical outcomes and adoption challenges. In addition, the Hub will lay tracks for new data flows, enabling us to carry out advanced analytics on cost effectiveness of the remote monitoring solution.	This project provides a timely example of the positive benefits of remote monitoring such as reduced service demand and improved patient activation. This provides a unique opportunity to gather patient reported data to supplement clinically coded data and provide a blueprint for replication in other future studies.
8	To investigate the current treatment pathway for patients with hypercholesterolaemia and to use machine learning to find populations which are at the highest risk and are most amenable to interventions in cholesterol pathways. This study also aims to determine the factors associated with	Industry Large - Big Pharma and NHS	The study highlights the number of patients who are currently receiving treatment in-line with National Guidance for Lipid Management for Primary and Secondary Prevention of CVD and the population who would benefit from treatment optimization. The gap analysis will seek to identify patients who would benefit from earlier intervention.	The project aims to highlight the value of real-world evidence in understanding this clinical condition using an AI platform DataRobot, that has been implemented in the Discover-NOW data environment. DataRobot aims to automate and accelerate data analysis and insights and this study provides an opportunity to test this AI offer.

	good management of these patients.			
9	To evaluate the predictive performance of an AF risk prediction model for the identification of AF, the health economic impact associated with the detection of undiagnosed AF based on the performance of the AF risk prediction model and to produce a manuscript highlighting the findings from the health economic impact.	Industry -Large, Big Pharma and NHS	This project generated peer-reviewed publications and supported a validation of an algorithm in Discover, that was originally developed in another dataset. Demonstrating the validity and utility of an AF risk prediction model relating patient-level factors and how an outcome (AF diagnosis) may support appropriate diagnosis and management of AF.	This project led to the client commissioning additional work and provided opportunities for collaboration with various NWL partners planning to implement the algorithm into their current care. It also demonstrated the value of the Hub for testing AI in real world data and has led to future collaborations and opportunities.
10	Cholesterol High Impact Intervention Tool - aiming to create an interactive, national cholesterol management data-driven dashboard for HCPs that will provide a landscape view of their chosen organisation.	Industry Large - Big Pharma and NHS	By partnering with industry and the use of modelling we are able to quantify the current care gaps for patients with high cholesterol levels. The project will estimate the impact of high impact interventions that address gaps in care on healthcare costs, CVD events and mortality.	The project will produce a real-world data decision support tool that facilitates planning, delivering and commissioning care. The project took centre stage in The Guardian focused upon the use of data to combat CVDs. The study is also planned be published in a peer-reviewed article.
11	Chronic Venous Disease (CVD) - In order to plan and improve access to services in the local region, we need to estimate incidence, prevalence, and pathways of care for people with CVD. This study aims to understand the treatment and care pathway for patients across primary, community and secondary care.	Academic and NHS	The study quantified the prevalence of Chronic Venous Disease (difficult to diagnose) and established the treatment pathway across primary, community and secondary care. This was a very complex study in terms of code mapping and our analysts were able to develop the methodology to extract the required data.	The project will generate additional studies with academic and commercial sponsors, hence bringing new additional income streams. The sponsor plans to publish three scientific papers that will promote the versatility of the Discover dataset in this challenging area of research
12	RADAR is a collaboration between the ICS in NWL, AstraZeneca, MyWay Digital Health, Institute of Global Health Innovation, Imperial College NHS Trust and Discover-NOW. It aims to revalidate models for risk prediction in Type 2 Diabetes and link them to interfaces for patients and clinicians.	NHS, AZ, Imperial College London, SME (MyWay)	This project involves a collaboration across multiple sectors, the development of further use cases for the dataset including the application of machine learning and data science. Using tools available in the secure environment such as Data Robot, this project demonstrates the value of using the dataset for AI model revalidation.	The project showcases the value of the hub assets; the dataset being used to revalidate the risk prediction models, and the NWL Health Research Register from which participants for the evaluation component of this project are being recruited. This will lead to recommendations on deployment of these tools in NWL.

13	To estimate the health care costs associated with active disease for Inflammatory bowel disease (IBD) patients and to understand if medicine optimisation leads to reduced costs of active disease.	Industry Large - Big Pharma and NHS	Retrospective analysis showcasing our capability to apply real world data platforms in elucidating care gaps and the associate cost in treatment of Inflammatory Bowel Disease, a relatively rare but severe indication. Has led to collaboration with leading clinicians and institution in the field and generated a publication (manuscript in preparation).	This six-figure project has led to commissioning of two other significant studies by this client. The income generated can be used towards Hub sustainability and the collaborations will endorse the utility of the Hub as a valuable tool in research and public health investigations.
14	To elucidate the current treatment pathways for Treatment Resistant Depression (TRD) in NWL and comparison of the care delivered in NWL with the national predefined standards for quality of care in this patient population to inform subsequent action plans to improve care.	Industry Large - Big Pharma and NHS	Generating insights into TRD and showing the value of real-world data to support understanding of pathways for mental health conditions. Study is being delivered in collaboration with West London Trust and plans to publish findings in a leading journal, supporting continued research in mental health.	As a six-figure project this will provide sustainability to the Hub. As a second project of this size with the same industry client, this long-standing partnership with the Hub advocates the strength of application of this platform to public health research.
15	To map the pathway of care for patients with Fabry disease, establish the burden of health care utilization by this population and shorten the time to diagnosis through implementation of an algorithm to profile and identify Fabry disease patients.	Industry SME and NHS	A new collaboration with small pharma which will enable validation the utility of this platform in stablishing gaps in provision of care and burden of a rare disorder. This will demonstrate the versatility of the Discover dataset in real world evidence studies of not only common but also rare indications.	This study enables us to target this specific area of rare disease, opening potential revenue generating opportunities with companies interested in this therapeutic area. This opportunity is further enhanced by the collaborations established with national clinical experts in the field.
16	Examine the extent to which guidelines for testosterone deficiency in adult males (BSSM standard) are currently followed by clinicians.	Industry SME and NHS	Expanding our portfolio of research topics, this study has established collaboration with a well-respected endocrinologist at a centre of clinical excellence and generated a publication (manuscript in preparation) demonstrating the relevance of retrospective real world data analysis to inform development of guidelines for treatment of hypogonadism.	A use case for utility of this platform in fostering collaboration across different disciplines to generate clinically important data which can lead to improvement in national guidelines and quality of life for patients. The sponsor has expressed an interest to initiate a second study.
17	To describe the burden of Chronic Cough in NW London and develop a dashboard for identification of Chronic Cough patients with the aim of improving clinical care for this population, reducing suffering from this often-unrecognised illness, and to facilitate appropriate and timely referral to specialists for treatment.	Industry Large - Big Pharma and NHS	This study showcases the utility of retrospectives real world data analysis in generation of a tool for identification of at-risk patients, prevention of delay in treatment and improving quality of life. This study has generated a publication in a peer reviewed journal (manuscript in preparation).	Revenue was generated in utilising analytics as well as tool generating capabilities of this platform. Strong collaborations were established with clinical specialists in the field and internationally recognised institution (Royal Brompton). This combined achievement will endorse the application of this platform for future investment in research in this field.

18	To investigate the current prevalence in primary care of known risk factors for poor asthma outcome in children and young people and guide optimisation of care and prevention of avoidable fatalities.	Academia and Big Pharma and NHS	This study showcased how real-world data can be analysed in a retrospective context to facilitate recognition of clinical risk and optimisation of care for patients in order to prevent fatalities in children and young adults. The findings were published in Primary Care Respiratory Medicine - Nature: 2021 Jan 29;31(1):3.	This study is timely with view to the current air pollution risks to health of children and young individuals and prompted discussions with the sponsor to address this threat to quality of life and guidelines on treatment pathway for this indication in children and young adults.
19	To evaluate achievement of guideline directed targets for cholesterol levels in patients with a history of stable coronary artery disease or acute coronary syndromes. The study will assess the quality of secondary prevention in North West London as a case study for the rest of the country.	Academia and NHS	This study is a use case for utilisation of real word data analysis in determining adherence to established guidelines and the need for development of new guidelines for patient care improvement. The findings from this study are being prepared for publication as two manuscripts.	The data generated in this study has led to further interest from industry to expand the work from this initial study. The collaboration with Professor Lüscher, an internationally respected cardiologist, will endorse the validity of use of real word data analysis through the Hub to address fundamental public health issues.
20	Assessing the clinical effectiveness of the Ibex Artificial Intelligence Histology cancer diagnosis system for prostate biopsies using the Discover data	NIHR funded to support a collaboration across Imperial College London and the NHS	AI-based pathology system can facilitate earlier diagnosis to meet cancer waiting time targets. The NHS will benefit through reduced pathology burdens and compensation for the workforce shortages over the next 5-10 years. The expectation is for AI IBEX to be rolled-out to every UK pathology laboratory.	The study is a winner of the AI in Health and Care Award, announced by the Secretary of State and led by NIHR. It provides the Hub with an opportunity to build the AI service offer for the Hub and builds our profile with NIHR, a significant funder of research.

5.4 Gut Reaction

No	Project Description	Sector(s) Project Covers + name of client if possible	Short rationale of any impacts to Hub services the project has generated	Short rationale if the project provided any impact to Hub sustainability
1	METRIC-EF: Magnetic Resonance Enterography and Small Bowel Ultrasound as predictors of disabling disease in newly-diagnosed Crohn's disease	Academia - UCL		Evidence of utility in use of digital imaging data and developing statistical models
2	Recruitment of genetically stratified participants to commercial phase 2 trial	Pharma		Experience of setting up recall by genotype trial for stratified therapy
3	IBD Registry - pharmacovigilance (5 year study)	Commercial		Demonstrates industry interest and demand for post authorisation safety surveillance capabilities
4	Evaluation of anxiety, depression and suicidal thoughts in the UK UBD population using IBD BioResource	Academia	Importance of coordination of research across Hub, and PPI to ensure appropriate support for patients involved in studies	

			(e.g. when surveying about suicidal thoughts).	
5	IBD drug therapies and COVID-19	Academia - University of Exeter	Demonstrates utility of data modelling to provide risk stratification	Evidence of responsiveness of Hub
6	Correlation between risk genotypes and biomarkers in Ulcerative Colitis or Crohn's Disease	Academia / Industry		
7	Impact of therapy on SARS-CoV-2 infection and immunity in inflammatory bowel disease	Academia - University of Exeter	Demonstrates flexibility / adaptability in building datasets	Utility of linkage to other datasets (PHE)
8	Genomic basis for Inflammatory Bowel Disease	Commercial	Although the study was declined by the Steering Committee it helped shape the strategy for working with industry and highlighted the importance of patient involvement in decision making	Informed procedures for release of large data sets
9	A precision medicine test for risk stratification of immunofrail patients	Commercial	Evidence of utility for biomarker research	
10	SARS-CoV2 vaccine immunogenicity in immunosuppressed IBD patients	Academia - Imperial College	Demonstrates ability to rapid access to data / patients	Linkage to PHE data
11	Genetic determinants of cytokine response in IBD	Commercial/pharma	Exemplar industry project on IBD immunogenetics	Pilot study for commercial company; success will lead to further studies
12	Diagnosis and management of elderly IBD patients	Academia - St Mark's Academic Institute	Exemplifies research in a patient group often underrepresented in research (elderly patients)	Demonstrates inclusivity of the cohort and data
13	IBD phenotype and response to treatment in the elderly and different ethnic groups	Academia - Imperial College	Exemplifies research in patient groups often underrepresented in research	Demonstrates inclusivity of the cohort and data
14	Biomarkers for non- invasive diagnosis of IBD	Academic - Chinese Academy of Sciences	Application shows international reach of the Hub	Enhanced international reputation
15	Association differences across multiple sub-phenotypes of IBD	Commercial/pharma	Association data	Pilot from commercial company that will inform future studies
16	Real world evidence of different treatments	Commercial/pharma	Comparator of treatment regimes	Feasibility assessment to carry out 2 separate RWE studies
17	BioResource contract	Commercial/pharma	Demonstrates confidence by industry	Long term support for the hub
18	BioResource contract	Commercial/pharma	Demonstrates confidence by industry	Long term support for the Hub
19	Registry educational grant (2019)	Commercial/pharma	Hub project invited speaker at our Annual Industry Forum	Insight into stratification
20	Registry educational grant x9 (2020)	Commercial/pharma	Data capture of IBD COVID-19 dataset (40,000 people) that will be used in Hub; Hub project	Pilot use in the new methodology to link datasets brought to the Hub

		attendees at our Annual Industry Forum	
--	--	--	--

5.5 INSIGHT

No	Project Description	Sector(s) Project Covers + name of client if possible	Short rationale of any impacts to Hub services the project has generated	Short rationale if the project provided any impact to Hub sustainability
1	Informing NHS service recovery: analysis of how the COVID-19 lockdown and associated interruption of ophthalmic services impacted on a selected population with incident neovascular Age-related Macular Degeneration at Moorfields Eye Hospital and University Hospitals Birmingham NHS Foundation Trusts.	NHS, with Roche-Genentech providing Real World data science and analytical expertise.	As a first 'use-case exemplar' this project enabled end-to-end testing of Hub data collation, ingestion and curation systems. It also provided the basis for establishing a consistent data dictionary and data treatment protocol to assure data quality.	Roche-Genentech analysed anonymised patient data, which yielded specific and valuable feedback about pharma industry expectations for data quality and curation standards. A case study has been submitted for publication in JAMA Ophthalmology, which will promote this project as an exemplar of Hub capabilities for engagement with other prospective industry users.
2	Assessing response to treatment for people with Age-related Macular Degeneration affected by having Diabetes Mellitus.	NHS & academia, University Hospitals Birmingham & University of Birmingham	This project was the first data use application to INSIGHT and it provided an end-to-end test of the data access request process for the Hub, beginning with the application form via the HDR UK Innovation Gateway through to consideration by, and recommendations from, the INSIGHT Data Trust Advisory Board.	This project underscores the Hub's capability to support research that informs the design of better clinical outcomes for patients, contributing to reducing the burden on the NHS. It demonstrates the Hub's capacity to link routinely collected anonymised data from major ocular diseases impacting blindness to wider systemic health.
3	Confidential			
4	Confidential			

5.6 NHS DigiTrials

NHS DigiTrials has 11 projects, redacted for confidentiality.

5.7 PIONEER

No	Project Description	Sector(s) Project Covers + name of client if possible	Short rationale of any impacts to Hub services the project has generated	Short rationale if the project provided any impact to Hub sustainability
1	A project to enable rapid, near-real-time, data collection in Hospital Electronic Patient Records while providing clinical support tools to redeployed clinical staff during the pandemic	NHS Clinical / Management/ IT/ Informatics/ Academia	This project enabled the development of rapid data curation protocols. The proforma was deployed within 5 days to meet critical need and remains in use. It developed cross-sector working (NHS medical, management, IT, Informatics, academic), informed our agile and flexible working approach based on customer need, contributing to 5 papers.	This project reinforced the benefits of PIONEER controlling its own data/data pipeline through a single data controller and sign off process. The uniquely granular COVID dataset is expanding daily, the significant ethnic and socioeconomic diversity led to projects (some outlined below) with government, charitable, academic, NHS and commercial sectors.
2	A Birmingham COVID-19 Hospital Dashboard, including demography, multi-morbidity, severity, patient flow and staffing. Modelling of community outbreaks shared with PH WM	NHS Clinical/ Management/ HR / Public Health WM	Our real-time, granular COVID datasets were expanded to include hospital flow and NHS staffing to create a tool for immediate deployment in one and then four hospitals. Modelled community outbreaks were shared with PH West Midlands prior to community testing, increasing our experience working with HR departments and public bodies.	This project increased awareness and reputation of PIONEER with Government bodies and led to an ONS project (Project 5). Building the dashboard developed a skill which has now been utilised for a commercial project (Project 15). Working with an HR department increased our experience with different kinds of sensitive data.
3	A study to assess antimicrobial stewardship and the presence and consequences of secondary bacterial infections in the era of COVID-19	Academia: UCL	Promoting PIONEER COVID Datasets/Infographics led directly to this project. Rapid assimilation of microbiology data highlighted the benefits of controlling our own data. PIONEER clinical/pharmacy expertise focused data curation and enabled meaningful data interpretation. We sent data to an external TRE, working with their development team, to ensure data interoperability.	This project contributed to our external reputation to deliver a high-quality product to time and target. The rigorous QA/QC processes and clinical input to ensure appropriate data curation/interpretation were praised by the end user and adopted as standard processes within PIONEER, including consultancy as a service

<p>4 A project to improve acuity and risk prediction models for COVID-19 - a funded UKRI, academic-led project.</p>	<p>Academia (multi-centre) funded by UKRI</p>	<p>Awareness of our COVID-19 data led to an approach to be part of a UKRI COVID application successfully awarded. This study was a collaboration between frontline clinicians, PIONEER and academia. It has generated novel prognostic models predicting mortality and ITU admission. This was the first grant application to which PIONEER contributed.</p>	<p>The prediction tool from this project will be shared via HDR UK Gateway and PIONEER website, for use by the community. This activity increased awareness of PIONEER, whilst providing income. It allowed us to develop text to describe our offer succinctly to grant boards, and to build an academic costing framework.</p>
<p>5 A project with the Office of National Statistics to describe the national impact of hospitalised COVID-19 considering ethnicity and social deprivation.</p>	<p>Government/ ONS</p>	<p>Frontline clinical experience led to a published, in-house analysis of the impact of ethnicity on COVID hospital admission, intensive care assessment and outcome. When shared with SAGE, ONS requested a larger dataset, focusing on granular ethnicity and social deprivation. To ensure data completeness, PIONEER developed processes combining primary/secondary care data.</p>	<p>The ability to use shared care records and publicise data findings through “initial insights” publications has become standard operating practice to raise awareness. Links to PHE and ONS led to approaches to form DECOVID which provided grant funding to support further focussed COVID-19 work.</p>
<p>6 DECOVID - a highly granular, near real time clinical database and research environment to answer critical questions and improve patient care during the COVID pandemic. A consortium project with UCL, UCLH and the ATI</p>	<p>Academia - multi centre collaboration. UCL/ UCLH/ATI/ Edinburgh/Leeds/Cambridge/ LSHTM/ Oxford</p>	<p>PIONEER normalised data from UCLH and UHB, using the OMOP common data model. Data includes 148,267 patients and 9.7M drug administrations. The PIONEER team noted a programming error with EPIC. Curated data was shared with >100 analysts across the UK from 7 different centres, with outputs expected shortly.</p>	<p>PIONEER learnt how to work with large-scale external data sets, where different QA/QC processes had been applied. OMOP was a new offer for PIONEER, which we can now apply as an additional service if required. The project provided funding for data curation and analyst time, securing the COVID-19 data pipeline.</p>
<p>7 NIHR Health Informatics Collaborative (HIC) COVID-19 database: Curating COVID and cardiovascular data to contribute to national learning.</p>	<p>Government/Charitable (NIHR/ BHF/ Hub)</p>	<p>PIONEER’s profile through the Innovation Gateway lead to data collation and sharing for this national flagship project. Interoperability and QA/QC were key. Learnings from previous projects supported refreshed data near real time. These data have supported numerous publications, and more in future as more data is added.</p>	<p>This project provided financial resource, but also the opportunity to collaborate with international leaders in data science, enhancing our reputation and standing within the data community. We were able to seek feedback on data quality and completeness (which was excellent) and look forward to further impact from continuing analysis.</p>

<p>8 The potential to use inhibitors of C-Reactive Protein to improve outcomes in COVID-19: A data feasibility study.</p>	<p>Academia (UCL) / NIHR UCL BRC</p>	<p>PIONEER were approached after researchers saw COVID OMOP data available on the Gateway. The study included both modelling the association of CRP with COVID outcomes and assessing how inclusion/exclusion parameters might impact on potential recruitment: a feasibility/'fast screen' study. This 'fast screen' has been adopted as a PIONEER service.</p>	<p>PIONEER has developed a 'one-stop-shop' for clinical consultancy and data to inform the design of experimental/therapeutic studies. This enables an assessment of how representative and relevant the likely recruited cohort will be of the wider population, to inform study protocol. The initiative has led to commercial enquiries under development currently.</p>
<p>9 Venous Thromboembolic events (VTE) in hospitalised patients with COVID-19, a Rapid Analysis to change practice and improve adherence to risk assessment guidelines.</p>	<p>NHS/ Academic / NHSE</p>	<p>Rapid analysis of real-time data from PIONEER noted increases in COVID venous thromboembolic events (VTE). PIONEER shared data with academia to model risk-prediction; fed back to hospitals globally, advising changed VTE assessments; assessed prophylactic prescribing and changing patterns of adherence. Expanded granular COVID data available, with publications and continuing research.</p>	<p>With rapid adaptation, a dataset can pivot to service a number of use-cases; here for non-COVID NHS reporting to NHSE, COVID research and COVID clinical care. This is only possible through real-time access to data which can then be 'cut' to meet end-user demand. The funding also contributed to sustainability.</p>
<p>10 Microsoft InnerEye - Development of a deep learning model as a diagnostic adjunct to identify COVID19 in chest radiographs performed in a multi-racial population</p>	<p>Industry Large - Microsoft Research</p>	<p>This project uses <i>cutting-edge machine learning technology to build innovative tools for automatic quantitative analysis of 3D medical images to inform decision-making such as in COVID-19</i>. PIONEER worked with industry experts to build a Trusted Research Environment (TRE) with the functionality to use machine learning on unstructured image data.</p>	<p>This project benefited PIONEER in terms of funding, expert learning and reputation. The TRE has undergone cybersecurity testing and this design can be rebuilt for other imaging projects. It directly led to project 11. The PIONEER team have attended Microsoft lunch club meetings. Other affiliated commercial projects are under development.</p>
<p>11 A project with an SME to use acute CT images to assess biomarkers of response to surgery through automated machine learning</p>	<p>Industry SME</p>	<p>This SME-focused project stems directly from the learning gained from project 10. <i>We are building a TRE for unstructured image data, to deploy machine learning algorithms to assess risk prior to emergency surgery.</i> PIONEER developed new costing models and ways of working to accommodate the needs of SMEs.</p>	<p>Working closely with an SME increased our understanding of this sector's needs. <i>We have developed an SME brochure/ playbook to highlight our offer.</i> We have built cost models now used within other SME projects. We have increased consultancy services for SMEs, offering exposure to PPIE, clinical teams and purchasers</p>
<p>12 The impact of the COVID-19 pandemic on acute presentations associated with social isolation, including falls in the</p>	<p>NHS/ Academic/ Society of Acute Medicine</p>	<p>The Society of Acute Medicine and Royal Colleges raised concerns about the impact of social isolation on 'non-COVID' health. This</p>	<p>The collaborative partnership across learned societies helped raise awareness of PIONEER. This led to a PIONEER</p>

	elderly, and alcohol, drug and psychiatric conditions		project was built from the SAMBA datasets within PIONEER, leading to a collaborative project across Society, NHS and academic partners. It was published and shared with SAGE.	presentation at a SAM conference and further data enquiries from SAM members. The rapid curation, analysis and paper construction (2 weeks for all) was highly praised for enabling rapid insights to inform service planning.
13	DEMAND Hub (Data-Enabled Medical Technologies and Devices Hub): A project to enable West Midlands SMEs to benefit from health data	EDRF funded. Academic (University of Birmingham, Academic Health Science Network, West Midlands and Solihull LEPS)	PIONEER was included as a costed partner. Our increasing experience working with SMEs includes mapping data-needs and flexible commercial models. This 3-year programme supports SME businesses looking to enter the healthcare market. PIONEER provides both data and consultancy as a service to SMEs based within the West Midlands.	This work enables further linkages/exposure to SMEs. Working on this project has improved the content of our SME focussed marketing materials and an SME Playbook, through feedback and iteration. It has built links with teams with Regulatory expertise and product marketing skills, to add to the PIONEER offer.
14	A project to assess whether Same Day Emergency Care (SDEC) should be offered as an expanded service in secondary care.	NHS/Academia/Society Acute Medicine	This started with an enquiry of the SAMBA dataset, but was expanded to provide a bespoke dataset, focusing on clinical data but also care processes and hospital flow (knowledge gained from project 2). Relationships built from project 12 were expanded in this project, providing a returned customer.	This project required combined clinical/management expertise to curate a patient flow/clinical dataset. The interest gained from the Society of Acute Medicine has led to replications opportunities in different hospitals. The project identified the need for a clinical dashboard for service planning, and this led to a commercial project (15).
15	The development of an Acute Medical Dashboard to inform flow and assess the quality of clinical care in real time: a project with a large technology company	NHS/Industry Large	Having curated large acute datasets and with the experience of COVID Dashboards, PIONEER were approached to produce an Acute Medicine Dashboard. This is being built in collaboration with a large technology company and will lead to ability for hospitals (national and international) to monitor acute patients against relevant KPIs.	The project builds on our experience of clinical dashboards. The dashboard will be commercialised through the company with PIONEER benefiting through revenue generated. PIONEER will use its links with the Society of Acute Medicine to increase awareness of the tool. Market analysis indicates national and international interest.
16	ADMISSION UK Multimorbidity Research Collaborative on Multiple Long-Term Conditions in Hospital: from burden and inequalities to underlying mechanisms. An MRC/NIHR funded 4 year grant programme	MRC/NIHR funded - Multi centre collaboration (Newcastle/Birmingham/UCL/Dundee/Manchester)	Building on learnings from DECOVID and other grant successes, PIONEER leads on the technical aspects of data normalisation across sites as well as being a data and analytics site. We are sharing curation/QA/QC processes built on PIONEER SOPs and with data specifications. PIONEER Director was on the grant panel.	As well as providing funding, this grant led to PIONEER's inclusion in 3 other multi-centre programme level applications, with national and international partners. PIONEER was specifically sought due to our experience in real-time data curation, robust QA/QC processes and transparent data access processes; which

				has become part of our USP.
17	Building and using cardiology synthetic data to predict sudden death in patients with Hypertrophic Cardiomyopathy (HCM)	Large pharma/Academic	PIONEER Technical Team have developed >24M synthetic records. Promotional work led to an approach from a large pharma company working with a university. Synthetic data has been created to represent patients with a rare cardiomyopathy. Curating bespoke synthetic data for a rare disease provides a new offer for PIONEER.	As well as funding, this project increased our experience of generating high quality synthetic data. We are now starting to create Deep Fake synthetic images, to enable algorithm training without the governance concerns of using real patient data. This expertise is contributing to project 19.
18	Comparing the burden of hospitalised community acquired pneumonia (CAP) before and during the COVID-19 pandemic and using reinforcement machine learning to stratify patients and inform clinical decisions: 2 PhD studentships	Academia - single centre. Dunhill Trust Funding	PIONEER contributes to the health data workforce by supporting PhD studentships, building capacity and capability. These two projects utilise different approaches to the interrogation of health data, both of which PIONEER can support within bespoke TREs. These projects require mentorship, analytical oversight, and appropriate governance training for the students.	As PIONEER controls its own data flows and processes, we are able to accommodate students, either for masters/postdoctoral placements or to develop specific skills. This has enabled us to contribute to the recently submitted HDR UK Wellcome PhD studentships and Black Internship programme, providing another offer to support the data community.
19	Building a PIONEER Data Science TRE, loaded with synthetic data and used to train the Data Scientists of the future: a collaboration with the University of Birmingham School of Computer Science	Health Education England (HEE) funding	PIONEER are building a TRE, populated with synthetic data and used to provide training for Computer Scientists. The TRE will be used by the Black Internship and the Wellcome PhD studentship at UoB. It will also be available for training employees from commercial organisations or external academic/NHS partners.	PIONEER already offers a 'bespoke TRE for hire', where customers can analyse their own data. However, previous projects highlighted the opportunity to improve understanding of working with unconsented NHS health data. This project enables PIONEER to offer training bundles to a variety of potential customers across learning sectors.
20	An exploration of the number and reasons for pregnant and newly postnatal women's attendance at non-maternity emergency departments - the development of the Maternity Early Warning Score (MEWs)	NIHR Applied Research Collaboration (ARC)	The Society of Acute Medicine highlighted the poor outcomes experienced by pregnant and post-partum women attending general emergency departments. PIONEER joined with the West Midlands ARC to address maternal health inequalities	The maternity dataset has led to discussions with Birmingham Women and Children's Hospital (BWCH), to share data with PIONEER, expanding the data we offer in terms of condition, patient age and modality. We plan that BWCH will join PIONEER maintaining the single Data

		through access to high quality data and analytics. This forms a new partnership already leading to other projects.	Controller model that PIONEER benefits from.
--	--	--	--