HDRUK
Health Data Research UK

GREEN SHOOTS:

HEALTH DATA RESEARCH UK'S ANNUAL REVIEW

2019/2020

Contents

Introduction

Health Data Research UK is the national institute for health data science for England, Scotland, Wales and Northern Ireland. Our mission is to improve people's lives.

Through uniting, improving and using health data at scale and enabling researchers and innovations across the UK – from the NHS, universities, and industry – we are enabling discoveries that are transforming the way we are able to detect, understand and treat diseases and keep people healthy.

We are still new, and we are learning, developing and evolving as this review sets out. We were set up in 2018 by our core funders who had the foresight to harness the power of the UK's health data to benefit patients, society and make the UK recognised as the place to do truly impactful health data science. The impact of the COVID-19 pandemic has shown how prescient this decision was and has put Health Data Research UK at the heart of realising the benefits of health data.

This year's annual review tells the story of the green shoots that are demonstrating how we are enabling research and innovation using the power of data and how this is improving lives.

Welcome

Welcome to Health Data Research UK's annual review of 2019/20. At the end of this year we were thrown into one of biggest challenges ever faced by the healthcare and research community in mobilising our expertise and resources to respond to the COVID-19 pandemic.

Our establishment as an institute federated across the UK has provided the foundation needed to rapidly co-ordinate and connect national data science efforts, accelerate access to data relevant to COVID-19 and leverage the best of the UK's health data science capability. The research response has required an enormous shift in effort, involving many colleagues on the front line of clinical and social care, and this has been made more manageable by the strong connections we have across the health data research community in the UK.

This year's annual report highlights the significant progress we have made in building these important connections across our growing community. Establishing an Alliance of the UK's leading health data custodians and a Gateway to discover data and request access to data has enabled these connections to flourish. We have been bold and during our journey we have started to solve some important, knotty, issues such as openly discussing the commercial aspects of health data access; ethics and information governance; and data quality standards.

Our work is starting to show tangible impact for patients, but this is only the start and we are seeing the huge potential of health data to address the most challenging health issues of our lifetime.

To do this, we have needed to overcome obstacles and our values of transparency, optimism, respect, courage and humility have guided us along the way.

It is a testament to the teams of people and the organisations that we work with in the UK and further afield that we have made such progress. We look forward to another year of demonstrating the impact and benefits of health data on people's lives.

Andrew Morris Director

Caroline Cake Chief Executive Officer

Our vision and strategy

Health Data Research UK is now established in 32 locations across the UK and brings together the sharpest scientific minds with safe and secure access to rich health data to better understand diseases.

Wellcome Trust
Great Ormond Street
DRIVE Unit

CENTRAL TEAM OFFICES

HDR UK Cambridge
HDR UK London
HDR UK Midlands
HDR UK North
HDR UK Oxford

HDR UK Scotland
HDR UK South-West
HDR UK Wales and Northern Ireland

RESEARCH LOCATIONS

BREATHE
DATA-CAN
Discover-NOW
Gut Reaction
INSIGHT
PIONEER
NHS DigiTrials
BHF Data Science Centre

HEALTH DATA RESEARCH HUBS

Belfast	Exeter
Birmingham	Leeds
Bristol	London
Cambridge	Manchester
Edinburgh	Oxford

TRAINING LOCATIONS (Masters and PhD)

Figure 1: HDR UK is a federated institute: 86 organisations, 56 offices at 32 locations

Our five-year One Institute strategy was published in April 2019 and sets out our 20-year vision for large scale data and advanced analytics to benefit every patient interaction, clinical trial and biomedical discovery and to enhance public health. It describes our mission to unite the UK's health data to enable discoveries that improve people's lives. To achieve this, our strategic delivery plan for 2020/21 focuses on our unique strengths and expertise of building a health data research infrastructure for the UK through:

- **Uniting health data** – providing national convenorship through the UK Health Data Research Alliance and Innovation Gateway with open standards, and in a way that earns the trust of patients and the public

- **Improving health data** – providing tools, methods, hubs, and national expertise in health data quality improvement for researchers and innovators

- **Using health data** – enabling research and innovation, demonstrating novel approaches to health data use, impact at scale, and establishing the next generation of research leaders in health data science

We are delivering this strategy through our inclusive, team-oriented One Institute ethos – bringing together NHS, universities, research institutes, industry and charities – built on our values of transparency, optimism, respect, courage and humility. We are enabling team science across the UK and, whilst still early on this journey, we are starting to see evidence of the the geographic reach of our national science priorities (see figure 3).

The COVID-19 pandemic has highlighted the power of this agile, federated model, allowing a rapid and galvanising response to the UK research effort at scale in partnership with NHS, academia and industry.
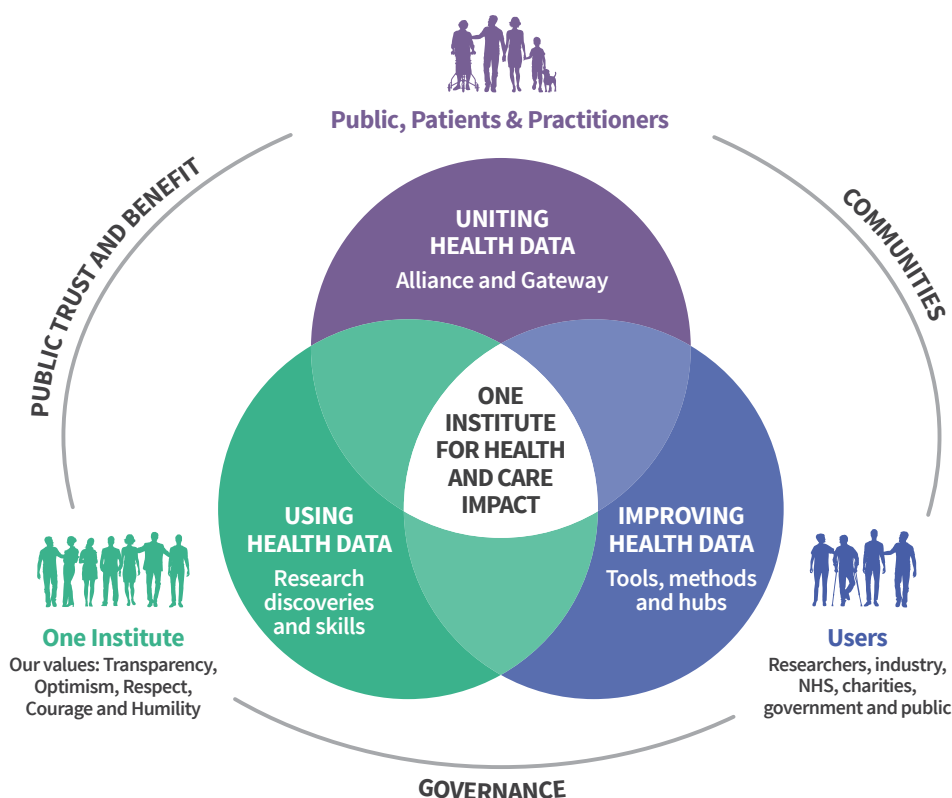

Figure 2: HDR UK One Institute Strategy schematic

Our five-year One Institute strategy was published in April 2019 and sets out our 20-year vision for large scale data and advanced analytics to benefit every patient interaction, clinical trial and biomedical discovery and to enhance public health.

Highlights of the year

OUR RESPONSE TO COVID-19

1,300 PEOPLE collaborating on Covid-19 research

62 patient and public volunteers in Covid-19 advisory groups

100 Covid-19 preprint publications

UNITING, IMPROVING AND USING HEALTH DATA

30 member organisations of the UK Alliance

649 peer reviewed publications

442 datasets discoverable on the Gateway

3,700 users searching datasets on the Gateway

56 influences on policy

5,000 searches for data on the gateway

15,000 views of the Gateway

OUR GROWING COMMUNITY

32 locations

8,300 monthly website visitors

5,000 followers on Twitter

3,000 contacts in our CRM

2,000 people on our mailing list

Over **450** institute members

Over **100** organisations involved in the Hubs

2019

APRIL 2019

One Institute Strategy published

Fellows Incubator launched in partnership with NIHR

MAY 2019

Competition launched for Health Data Research Hubs

JUNE 2019

National Implementation Projects underway

200 colleagues attended One Institute event

Masters Programme in health data science announced

SEPTEMBER 2019

Celebrated our 100th open access publication

Launched 7 Hubs, with over 100 partners, selected with patient and public involvement

AUGUST 2019

8 new member organisations joined the UK Alliance, taking the total to 18

Our first Summer School

JULY 2019

Partnership with NICE announced

OCTOBER 2019

Launched procurement for Gateway technology partner

BHF Data Science Centre launched

NOVEMBER 2019

Recruitment for HDR UK/ Turing Wellcome PhD Programme in Health Data Science

DECEMBER 2019

Hubs met their first milestone

Sprint exemplar projects completed

Joint meeting with Health Education England on building health data science capacity

Better Care Partnership and Catalyst Projects opportunity launched

2020

MARCH 2020

Mobilised resources to respond to COVID-19 with over 1,300 members of the community collaborating

5 national research projects up and running involving more than 18 research organisations across all 4 nations of the UK

FEBRUARY 2020

Held our first Alliance symposium with 140 attendees

Student scholarships funded with DataLab for 2 Scottish Universities

JANUARY 2020

The first stage of the Gateway with 400 datasets discoverable went live

Recruitment for Masters programmes opened

9 more member organisations joined the UK Alliance, taking total to 27

Joint event held in public (via live stream) to focus on the patient and NHS benefits of health and care data

Using Health Data

Our national research teams are demonstrating the use of health data for research and innovation, and this has played a vital role in the UK's response to COVID-19. Together, we are forging the discovery of previously unknown rare diseases, enabling more people to access the most innovative treatments by taking part in clinical trials, supporting the prevention of common diseases and creating more efficient health services.
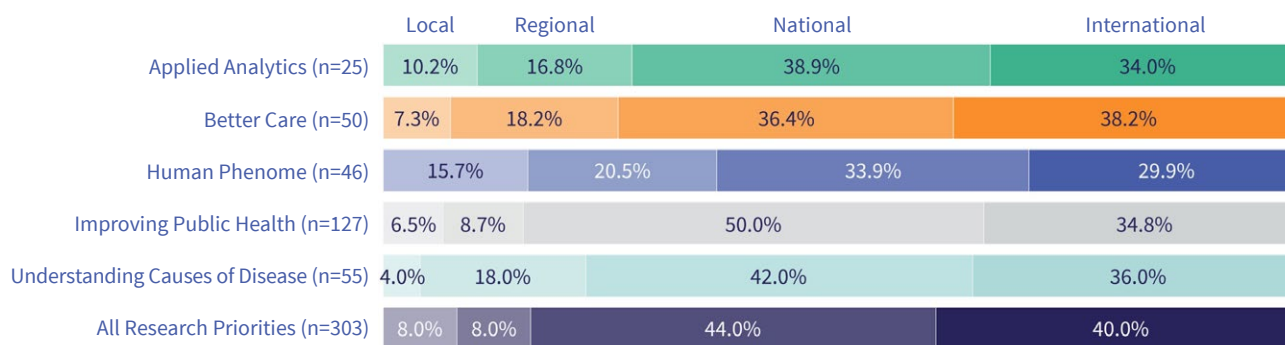
Delivering UK-wide research programmes

UNDERSTANDING CAUSES OF DISEASE
We are working towards generating major new insights into the molecular underpinnings of health conditions, through a "molecule to health record" approach, with an ambition to involve several million participants. The long-term goal is to contribute towards transforming scientific understanding of the causes of disease, its progression, and consequently disease prevention and treatment.

Genetics has transformed our understanding of how variation in DNA can influence risk of developing conditions, such as cancer and heart disease. Studies that can combine this genetic information with other blood-based factors – including proteins, metabolites and lipids - alongside the diverse information in health records, have the potential to provide more direct insight into the causes of disease. A key challenge is accessing this complex information at scale. The initiation of HDR UK's National Multi-Omics Cohorts Consortium earlier this year, aims to address this challenge by bringing together nine

Geographical Reach of engagements by Research Priority

	Local	Regional	National	International
Applied Analytics (n=25)	10.2%	16.8%	38.9%	34.0%
Better Care (n=50)	7.3%	18.2%	36.4%	38.2%
Human Phenome (n=46)	15.7%	20.5%	33.9%	29.9%
Improving Public Health (n=127)	6.5%	8.7%	50.0%	34.8%
Understanding Causes of Disease (n=55)	4.0%	18.0%	42.0%	36.0%
All Research Priorities (n=303)	8.0%	8.0%	44.0%	40.0%

n is the number of engagements

Clinical Trials recorded a national reach only

Figure 3

Source: ResearchFish (April 2020)

of the UK's leading research cohorts linked to electronic health records, involving more than 750,000 participants. This is a truly national endeavour that involves investigators from multiple institutions across the UK.

We also launched the Polygenic Score Catalog this year – an open database of polygenic scores, helping to predict the chances of developing a particular disease according to genetic make-up. This is an important tool allowing the sharing of complex health information, to better inform clinical insights.

CLINICAL TRIALS

We are using health data to ensure that every individual across the UK will have access to the latest treatments and technologies through access to clinical trials. We are driving the way we conduct clinical trials to catch up with today's technology, such as using electronic health records to identify patients who are candidates for trials, quickly, efficiently and at a national scale, instead of relying on local networks that not all doctors or researchers have access to.

We are using health data to ensure that every individual across the UK will have access to the latest treatments and technologies through access to clinical trials

A fantastic exemplar of this approach is the RECOVERY Trial, a national clinical trial aiming to identify treatments that may be beneficial for those hospitalised with confirmed COVID-19. A range of potential treatments have been suggested for COVID-19 but nobody knows if any of them will turn out to be more effective in helping people recover than the usual standard of hospital care which all patients receive. The RECOVERY Trial is testing some of these treatments. This is now the world's largest randomised clinical trial of potential coronavirus treatments and has recruited over 10,000 patients at an unprecedented pace. Successful treatments will given to NHS patients as quickly as possible once the scientific evidence for their efficacy and safety has been generated by this trial. NHS DigiTrials, one of the Health Data Research Hubs, is supplying national, routinely collected clinical outcomes data for the trial's cohort of patients to the RECOVERY Trial, which is reducing the need to collect data from busy clinical staff.

BETTER CARE

Our Better Care vision supports and enables a learning health system approach across the UK integrating clinical practice, large scale data and advanced analytics in a cycle of continuous improvement – equipping clinicians and patients with the best possible data-based information to make decisions about their care. To support this, we launched two new Better Care Partnerships in the North and South-West. These add to our existing network of digitally maturity health and care sites and will provide a key opportunity to develop solutions to scale and share learnings of Better Care loops across multiple health and care decisions and transfer Better Care innovations across settings. We have also established a partnership with the Health Foundation to deliver the Catalyst programme. Four innovative Catalyst Projects have been selected and form an integral part of this programme – these rapid projects will deliver rapid learnings about the Better Care Loop and generate insights and learnings to support the development of a learning health system approach.

Both the Partnerships and Catalyst projects include stakeholders essential to delivering Better Care across the patient pathway, bridging primary and secondary care organisations, local government and social care providers, systems providers, academic and research institutions, the pharmaceutical and technology industry and charities. This integrated approach is essential to delivering impact and improvement throughout the patient journey.

"There has never been a better time to demonstrate to the public how sharing their data and using it safely and wisely to build knowledge is the key to better health and care services."

Angela Coulter
Chair of HDR UK Public Advisory Board

PUBLIC HEALTH

Factors that cause poor health are complex, and we do not always have the right data to tell what kind of effect the environment, a behaviour, event or intervention has had – or could have – on a person's health. However in many cases, the data we need exists already – it is collected in GP surgeries, hospitals as well as schools and other administrative systems across the UK every day – the challenge is to harness it safely and securely in innovative ways and do the right analyses, to improve public health, address health inequalities and better understand the health of the whole nation.

This year, we initiated a major new project in this area, the National Multimorbidity Resource. It involves researchers across the UK, bringing together multiple datasets representing up to 10 million people, to explore multimorbidity - when someone is living with more than one long-term condition. We are discovering what diseases and conditions are found together, how they develop as people age and also which cause the most problems for people and the health service. This will enable better planning of delivering care to people with multimorbidities.

We have also established new national working groups in maternal and child health, mental health and environment and health who are forging ahead with new research ideas that could not be delivered by an individual research organisation alone.

Developing future leaders in health data science

Delivering our ambitious vision and world-leading health data science requires a community of scientists with new skills that will strengthen medical research and open up faster, smarter pathways to patient care. During the year, we have made progress in delivering our training strategy by providing opportunities for early career researchers.

Following an open competition to UK universities and a selection process that included an independent panel of academics, health professionals and a patient representative, we selected six universities and their 34 partner organisations to establish health data science masters programmes to help address the skills shortage in data science in the UK. With recruitment underway for students to join these programmes during 2020 and 2021, our aim is to stimulate the development of training that genuinely integrate statistics, informatics and health science, aimed either at medical students or life sciences graduates keen to develop their quantitative skills, or at core maths, physics, statistics and computing graduates wanting to move into the health sphere.

We continued to provide opportunities for our 46 HDR UK Fellows to build their skills, form a network with fellows from the NIHR incubator in health data science, and contribute to our early career researchers committee.

In August 2019, we ran our inaugural Summer School, hosted by the University of St Andrews. Over 60 HDR UK Fellows and other early career researchers attended to learn about new areas of research in health data science and to gain practical skills. On the final day, pupils from two local secondary schools joined the event to learn about career opportunities in health data research.

"I've learnt a lot about the applications of maths that I've never really thought about before and to see this applied in medicine is really interesting,"

Student attending schools engagement event at the 2019 Summer School

In November 2019, we opened recruitment for a Wellcome-funded PhD programme in health data science in partnership with the Alan Turing Institute. This four-year programme provides in-depth training for graduates interested in bringing their numerical and computational skills to health data science. We expect to announce the successful PhD students later in 2020.

Other progress includes 10 scholarship places awarded to two universities in Scotland, in partnership with DataLab, with students starting in 2020, and launching our oversubscribed 'Data Science for Doctors' training in partnership with the Software Sustainability Institute. Underpinning this we have continued to work in partnership with Health Education England to shape a 'system'-wide approach to establishing a talent pool for health informatics and data sciences.

CASE STUDY

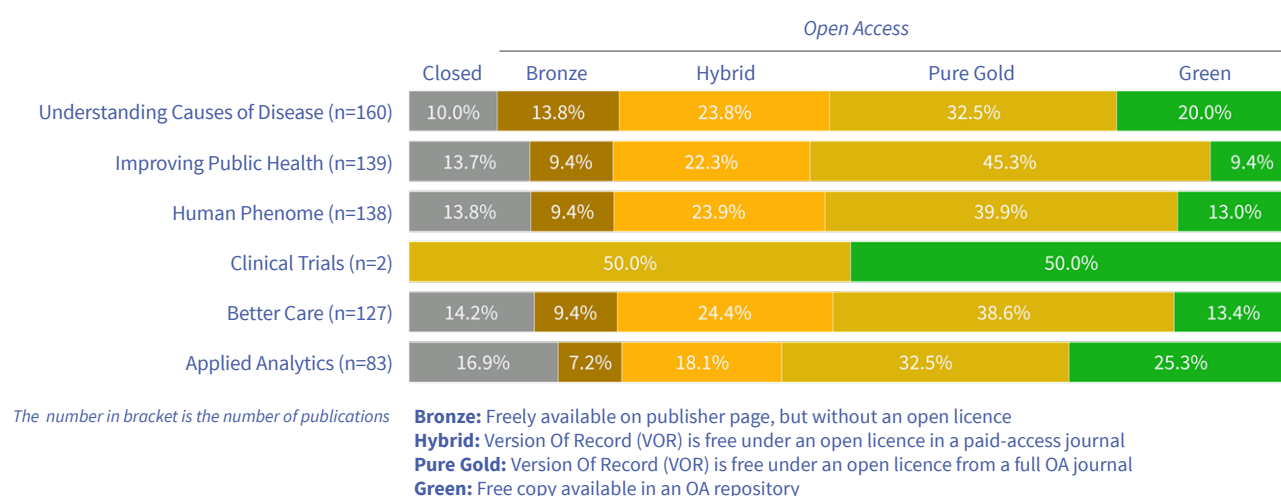# A snapshot of how health data is used to generate insights

During the year, the HDR UK community generated insights using health data at scale and published these in 649 peer-reviewed journal articles, as well as a substantial body of pre-prints. Figure 4 shows the accessibility of these articles by area of national research priority. In this case study we present a selection of impacts from across these priority areas.

A public health study looked at the long-term impact of common medications given to pregnant women, on children born by caesarean section. It was found that broad-spectrum antibiotics given to mothers to prevent infection following caesarean section had long-term impacts on childhood asthma and eczema. A further study in this area discovered that some antibiotics prescribed to pregnant women to

control bacterial infections are associated with birth defects in newborn babies. These insights will feed back into policy with the aim of improving medical practices and children's lives.

To achieve our mission, we aim to enable research to benefit the entire 66 million population of the UK. To achieve this, it is important to include people from all segments of the population, including diverse ethnic groups and those defined as hard-to-reach or vulnerable, such as homeless people. The ability to link datasets is critical to understand the health of people without a stable residence, as it enables researchers to have a view of their interactions with healthcare providers as and when they happen. By linking health data, a study found that the leading causes of death in

Accessibility of publications by National Research Priority

	Closed	Bronze	Hybrid	Pure Gold	Green
			Open Access		
Understanding Causes of Disease (n=160)	10.0%	13.8%	23.8%	32.5%	20.0%
Improving Public Health (n=139)	13.7%	9.4%	22.3%	45.3%	9.4%
Human Phenome (n=138)	13.8%	9.4%	23.9%	39.9%	13.0%
Clinical Trials (n=2)			50.0%		50.0%
Better Care (n=127)	14.2%	9.4%	24.4%	38.6%	13.4%
Applied Analytics (n=83)	16.9%	7.2%	18.1%	32.5%	25.3%

The number in bracket is the number of publications

Bronze: Freely available on publisher page, but without an open licence
Hybrid: Version Of Record (VOR) is free under an open licence in a paid-access journal
Pure Gold: Version Of Record (VOR) is free under an open licence from a full OA journal
Green: Free copy available in an OA repository

Source: Dimensions (April 2020)

Figure 4

homeless people – in contrary to the stereotype of drug and alcohol addiction or hypothermia – were more highly amenable to healthcare intervention, such as heart disease, cancer and stroke. This will have an important impact on the approach to protecting the health of some of the most vulnerable people in our communities.

Many colleagues working across HDR UK often form part of international collaborations. One example is a consortium that used a whole genome sequencing approach, to find novel genetic mutations in tuberculosis patients who show resistance to commonly used tuberculosis-associated drugs. This knowledge in understanding the causes of disease will contribute to improved management of tuberculosis patients through more patient-specific medicine regimens.

The work of our community to respond to the COVID-19 pandemic in the early part of 2020 resulted in insights captured in pre-print publications. These include the identification of airport screening as an effective way of detecting travellers infected with SARS-CoV-2, early studies on the dynamics of transmission and control, and feasibility of controlling 2019-nCoV outbreaks by isolation of

cases and contacts. These insights, and a multitude of others were reported to the government's Scientific Advisory Group for Emergencies to support UK policy developments on COVID-19.

The ability to use health data to generate insights such as those discussed above are underpinned by the development of robust and reliable statistical methodologies, algorithms, digital analyses and other approaches to make sense of diverse data at scale – our toolbox for health data research. An example is work that developed a way of using health data to identify people at immediate risk of heart attack, prior to it happening involving machine-learning, which could enable targeted and early intervention to prevent heart attacks from occurring in this vulnerable group of people.

This cross-section of discoveries from the last year provides a glimpse of the emerging impacts HDR UK is having to improve people's lives through better public health, understanding the causes of disease and clinical trials for better care, underpinned by the approaches developed through applied analytics and the human phenome project.

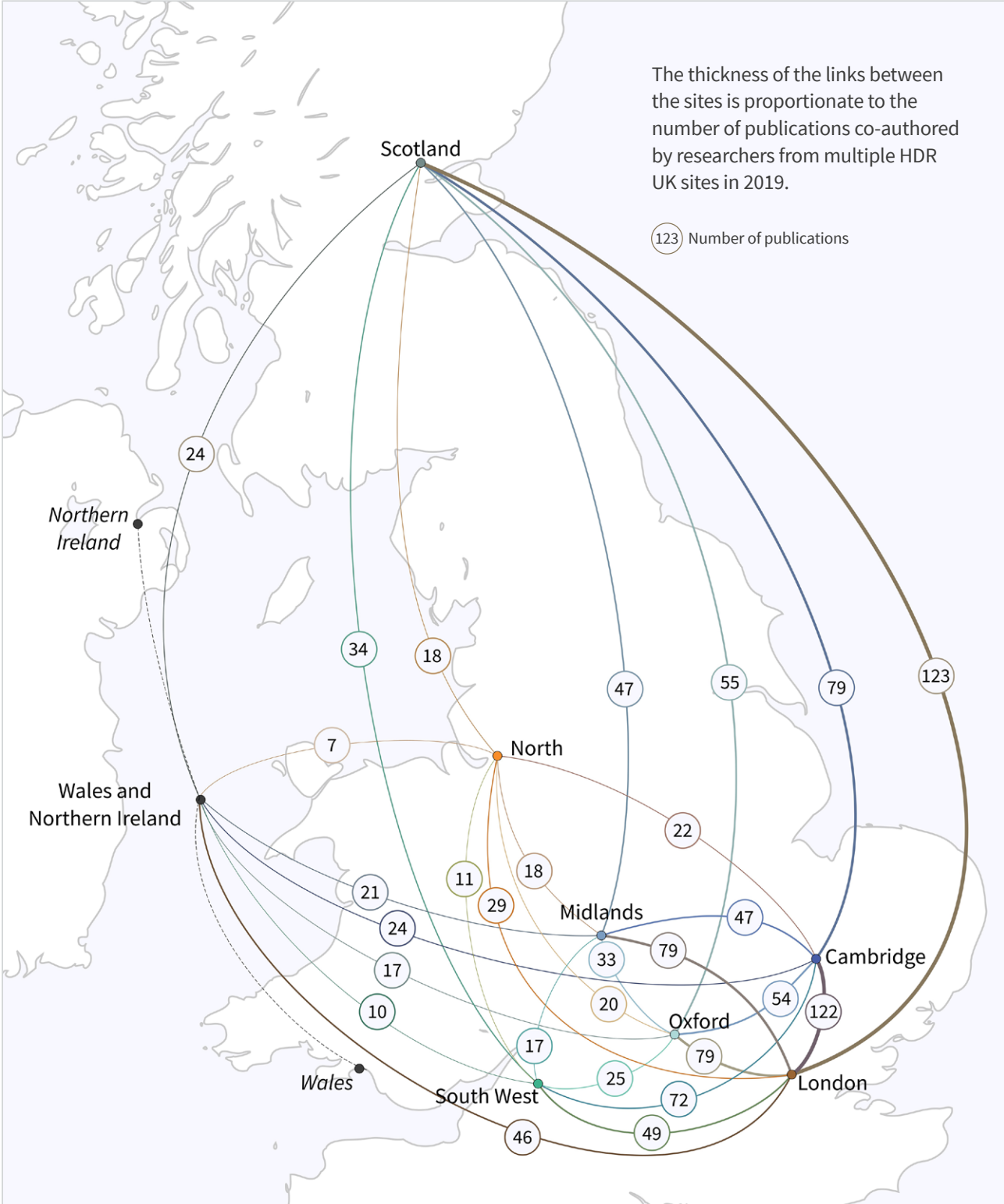## Shared publications between HDR UK sites

The thickness of the links between the sites is proportionate to the number of publications co-authored by researchers from multiple HDR UK sites in 2019.

(123) Number of publications

Figure 5

Uniting Health Data

The UK has some of the richest health data of anywhere in the world. However, these datasets are fragmented making it difficult, sometimes impossible, to access for research purposes. This causes unnecessary delays in researchers and innovators being able to access joined up data and making important discoveries that improve people's lives. At HDR UK we have made significant strides in uniting health data, building a safe and robust infrastructure for data research and enabling the growing community of health data scientists to discover and request access to these datasets.

Delivering the Health Data Research Innovation Gateway

Improving access to data for research and innovation is central to our mission. The first phase, or 'Minimum Viable Product' (MVP), of the Gateway went live in January 2020 as a portal to enable researchers and innovators to discover which datasets are held by individual healthcare and research organisations across the UK. It aims be a one-stop-shop for researchers and innovators to share, manage and drive the ethical and safe use of data, exploiting cutting-edge analytics and big data technologies to improve health outcomes. The MVP was developed using teams across seven organisations working in partnership with HDR UK and involving both custodians and users of data throughout this process. The rapid development of the MVP took three months and at launch had over 400 UK datasets discoverable. By the end of March 2020, the Gateway had already been used by 3,700 people, with 5,000 searches for datasets showing its unique potential, even at this early stage.
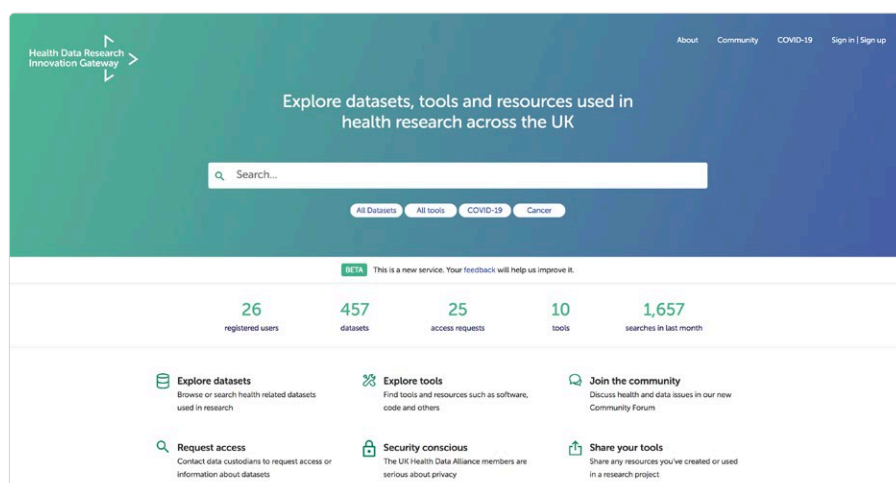

Figure 6: Homepage of the Health Data Research Innovation Gateway

Participants at the first UK Health Data Research UK Alliance symposium held in February 2020 at the Wellcome

In October 2019 we began an approach to identify a lead supplier to develop the full functionality of the Gateway. This involved multiple stages of increasing scrutiny, with colleagues from the Public Advisory Board, members of the UK Health Data Research Alliance including NHS Scotland, NHS Digital, and NHSX involved in the selection process. In April 2020 we appointed PA Consulting to work with us to co-develop the next phase of the Gateway.

Growing the UK Health Data Research Alliance

During the year we made great strides to increase the breadth and scale of data available for research and innovation, by building on the early foundations of the UK Health Data Research Alliance (the 'UK Alliance') – an independent group of leading healthcare and research organisations united to establish best practice for the ethical use of UK health data for research at scale. The UK Alliance experienced a rapid increase in members – growing from eight founding member organisations in February 2019 to 30 member organisations in April 2020,

covering all four nations of the UK, including some of the UK's largest NHS trusts, national bodies, charities, disease registries and research cohorts. With over 400 datasets discoverable via the Gateway, the expansion of the UK Alliance brings an exceptional opportunity to provide transparency and access to rich and diverse health data for research and innovation at a scale that is unprecedented in the UK.

Each member subscribes to the UK Health Data Research Alliance's Principles for Participation on public involvement, information governance, ethics, intellectual property rights and commercial models as a data controller, and to the set of services in the Gateway.

Through the development and sharing of best practice, and by earning trust in health data use, the UK Alliance aims to maximise research potential and ensure it benefits as many people across the UK as possible.

The UK Alliance has been pivotal in making data accessible for research in response to COVID-19.

Establishing standards for UK health data research

The UK Alliance is driving the delivery of five key priority areas that aim to transform the safe and responsible access of UK health data at scale. During the year we have focused on:

- Increasing the utility of data through standards and improving quality. The Data Officers Group consisting of more than 40 individuals across the UK Alliance, Heath Data Research Hubs and other bodies has helped to shape the strategy and set out the principles for data quality that are currently in development.

- Enhancing discoverability of data through the development of the Gateway and development of a metadata specification which resulted in over 400 datasets with comparable metadata being discoverable at the launch of the MVP (see case study on page 23)

- Developing a commercial framework for accessing health data through consultation with Alliance members, public advisory board and other stakeholders, with our report Realising patient and NHS benefits from health and care data – from policy to practice published in February 2020

- Developed a draft Green Paper for consultation on developing an aligned approach to using trusted research environments for health data research.

- Working with patient and public representatives to provide input on these projects, including the development of the Gateway and proposal for trusted research environments

Sprint exemplars

During the year, the 11 Sprint Exemplar Innovation Projects completed their activity which has informed the creation of the UK-wide health data research infrastructure. These projects covered a range of different areas, including adopting digital technology in care setting, building underlying databases to support research, and providing confidence in concept of specific technologies. Each project reported its results in December 2019, and contributed to the next stage of the journey in different ways: contributing to the design of the Gateway; seeding the Hubs by sharing early lessons, idea generation and information about the health data market; and growing our community across healthcare, academia, patient groups and industry.

Raising the bar on transparency

Patients and the public are at the heart of our One Institute strategy and, during the year, we have demonstrated our commitment to earning and building public confidence and trust in our work. Working in partnership with our Public Advisory Board, our community members and partners, we are building a process to ensure that we are transparent our work and specifically how and why data is accessed. To support this, we encourage involvement and input from patients and the public across all our work programmes.

During the year, members of our Public Advisory Board were involved in the development of the Gateway, providing input and advice on accessibility, and user journeys with the MVP through to selecting the technology partner to develop the full functionality. We also sought advice from other public groups, which influenced the language used to describe the Gateway more clearly.

"It has been great to play my part in the public and patient involvement in building the Gateway so far. This is crucial to establishing and maintaining public trust. As with all of HDR UK's work, it is vital that we keep the patient voice at the core of the vision for the Gateway going forward to ensure that it continues to put benefits to the public, patients and the NHS first."

Sarah Brooke
Deputy Chair of HDR UK's Public Advisory Board

Increasing the scale and diversity of data available for research

During the year, the UK Health Data Research Alliance expanded to include 30 leading health organisations, research institutes and charities – each bringing a shared commitment to increasing the diversity of health data available for research and innovation. This growth in accessible health data is making a tangible impact on answering important research questions, including those to respond to COVID-19.

Established in February 2019, the Alliance is an independent group of health data custodians united to establish best practice for the ethical use of UK health data for research at scale. It develops and co-ordinates the adoption of tools, techniques, conventions, technologies, and designs that help researchers to answer some of the most difficult questions and address the most important health challenges faced in the UK.

The new members joining the Alliance this year included the first NHS trusts and medical research charities, which offer an exceptional opportunity to enable greater access to rich and diverse health data for research and innovation. Most Alliance members have made their datasets discoverable via the Gateway, increasing the accessibility to these important research resources. These are now searchable alongside datasets held by national health data custodians and our partners in the Health Data Research Hubs.

Research based on data that reflects diversity of culture, healthcare conditions and aspects such as race, ethnicity, gender and age improves the ability to generalise results and enables new discoveries and understanding about disease. In turn this provides fairer and more equal access to the latest treatments and medical technologies, benefiting as many people across the UK as possible.

The Cystic Fibrosis Trust is one of the new members joining the Alliance this year. This national charity holds a registry that records health data on people with cystic fibrosis in England, Wales, Scotland and Northern Ireland, who have consented to have their data used for research.

"People with cystic fibrosis are remarkably research-motivated and data-savvy. They have proven time and again to enthusiastically participate in research, even when the results may not directly benefit them. This dedication to discovery, coupled with the dedication of clinical teams who diligently collect data that is 99% complete and 98% accurate, means that cystic fibrosis data is ideally placed to share experience and insights across the health data science sector. Completeness of coverage and depth of data also means that this complex and multifaceted condition presents opportunities to act as a microcosm of more populous disease areas."

Patients and the public are at the heart of our One Institute strategy and, during the year, we have demonstrated our commitment to earning and building public confidence and trust in our work.

Improving Health Data

We are harnessing expertise across the UK to improve the quality of health data for research and innovation. We are creating the infrastructure, tools and methods to provide access to better data that will help establish the UK as a world leader in health data research.

Establishing Health Data Research Hubs

High profile global partnerships

In May 2019, we ran a UK-wide competition to select and establish seven Hubs to demonstrate and drive the utility of expert health data science, based on an area of expertise. Initially funded through the UK Research and Innovation ISCF investment to kick-start the approach, the Hubs are formal collaborations of over 100 NHS, academic organisations, patients, charities and industry. They bring their collective expertise together to maximise the value of health data research, potentially benefiting millions of people across the country. They provide data, curate the data and offer services, such as supporting researchers with study design or refining research questions in light of the available data, for research and innovation.

Within the first three months, the Hubs celebrated their first milestone. In this time, they identified 39 datasets and made them discoverable through the Gateway, set up physical locations around the UK, and set out plans to engage with and involve patients and the public on use of patient data.

Also this year, the BHF Data Science Centre was established, led by HDR UK and funded by the British Heart Foundation (BHF). The Centre is enabling responsible, ethical research that combines the power of advanced analytic methods with the UK's large-scale and diverse cardiovascular data. High impact outputs will help to shape better cardiovascular health services, provide patients and health professionals with the tools to make better decisions, and bring the latest medical discoveries to patients across the UK faster than ever before.

"Many people support the use of patient data for research but are understandably nervous about the involvement of commercial organisations. It's great to see the Hubs taking public and patient engagement seriously as they meet this challenge of driving innovation while protecting peoples' rights and interests over data."

Natalie Banner
Lead for Understanding Patient Data

Hub launch

The Hubs have received international attention, being described as major priorities for several multinational firms, including Microsoft, AstraZeneca, IQVIA, Novartis and Roche. Responding to the COVID-19 pandemic, Hubs have played a central role in convening their respective communities to feed into a nationally reported research pipeline.

Using applied analytics

We are matching expert knowledge of analytical tools, such as machine learning and artificial intelligence, with large-scale health datasets, to demonstrate the power of these approaches to inform health and care delivery.

This year we initiated an exciting new activity, our National Reproducible Machine Learning Project, which brings together data science, with machine learning, health data from wearables (such as 'Fitbit-like' devices), and the important quality of reproducibility. The overall aim is to better enable machine learning to provide trustworthy clinical insights from the enormous amounts of health data available. The work is addressing a currently unmet need – to provide confidence that machine learning results are trustworthy – before they can be routinely used in public health and clinical practice. Ultimately it will help clinicians to better prevent and treat a wide range of human diseases. This collaborative project includes colleagues from the Alan Turing Institute, and several of HDR UK's research sites across the UK.

We have also honed in on the most important areas of development for analytical techniques, like machine learning, as well as the most helpful areas of deployment, such as replacing the most time-consuming tasks in healthcare and identifying features in medical images, such as cardiac MRI.

Developing the 'human phenome' to understand characteristics of disease using data

We are driving forward the Human Phenome Project – redefining how we describe all human diseases in a data-driven way. We have more ways of measuring disease than ever before, beyond symptoms, including images (e.g. x-rays), continuous measurements (e.g. physical activity via wearable devices), and a person's genetic make-up (e.g. through genome sequencing). Much of this is captured in NHS electronic health records, which make them a valuable resource for studying human disease. We are developing new, more detailed and consistent descriptions of all human diseases, which will facilitate a deeper understanding of disease and how to treat patients.

This year we initiated two major national projects in this area: the National Phenomics Resource is developing new tools to harness the power of electronic health records to characterise disease and the efficacy of treatments. It brings together the UK's community of researchers working in this area, who are now learning from each other and gleaning insights from the mass of health data in the NHS, towards improving health and healthcare. The National Text Analytics Resource is drawing out anonymised health insights from doctors' notes within electronic health records. They are developing computerised tools to process these words to build a full picture of all patient symptoms, experiences and diagnoses to use in research for patient benefit.

CASE STUDY

How the Health Data Research Hubs are improving data to respond to COVID-19

The response to the global COVID-19 pandemic has taught us how vital high-quality data is in gaining insights to this devastating disease. Having access to this data quickly and safely enables researchers to answer key questions about the course of disease, how it impacts people and which drugs might be effective.

In the early stages of the pandemic, the Health Data Research Hubs have supported researchers in asking and answering the key questions across a range of areas, including those about the impact on people with cancer, respiratory disease and patients requiring acute care. Where our existing health data has not been sufficiently developed to answer these questions, the Hubs have had to rapidly improve it. This has required reformatting, mapping and linking data to provide a greater scale and depth of information.

An example is the work of DATA-CAN, the Health Data Research Hub for Cancer, which has brought together data from several NHS trusts in England and Northern Ireland to demonstrate the impact the pandemic is having on routine cancer services. This real-time analysis showed significant impact on cancer diagnostic and treatment services, reporting a potential increase of 20% more deaths over a 12-month period in people who had been newly diagnosed with cancer.

PIONEER, the Health Data Hub for Acute Care, has worked to safely refine and combine extracts of data from NHS trusts to provide an integrated, interoperable dataset to support the national DECOVID project. Using detailed, frequently updated health data in a secure database, the project provides up-to-date information about patient care during the pandemic. The data is being analysed to answer the most pressing clinical questions to support the emergency response and to improve the quality of patient care for the future.

BREATHE, the Health Data Research Hub for Respiratory Health, has increased the scale of datasets relevant to COVID-19 and making these discoverable and accessible for research. This includes data from the COVID-19 Symptom Study, led by King's College London and health science company ZOE, which has over 3.5m users who are regularly reporting on their health and symptoms. This data has influenced health policy, evidenced by the addition by the UK Government of loss of sense of smell and taste as a recognised symptom of COVID-19.

Discover-NOW, the Health Data Research Hub for Real World Evidence, has co-ordinated system wide approvals to allow for linkage of priority datasets. This builds on the data already present in the Discover dataset by bringing in direct data feeds from pathology labs and near-real time data feeds from five NHS trusts.

"Our study highlights the value of bringing together data from multiple sources to enable researchers, health systems and policy makers to improve cancer management for our patients, both during and after this pandemic."

Dr Charlie Davie
Hub Director for DATA-CAN

One Institute

Health Data Research UK has grown in the past year, in terms of the scale of communities involved, programmes of work and impact on the health of patients and public through research and innovation.

Growing the community

With the addition of the North and South West Research Sites, training sites, Hubs, the Cardiovascular Centre and Alliance members, and Better Care catalysts our community has grown substantially during 2019/20. Through this growth we have added new expertise, new geographies and welcomed members from the charity and industry sectors. We now have over 450 members of the HDR UK community across all four nations, enabling us to lead and represent all aspects of UK health data science.

Embedding the involvement of patients and the public

During the year we have demonstrated the positive impact on our work from working with, learning from, and listening to patients and members of the public.

Our Public Advisory Board have guided and influenced our work and been involved in joint decision making, including the selection of the Health Data Research Hubs, Better Care partnerships and projects, and national implementation projects.

In March 2020, we set up a virtual group of 60 patients and members of the public to guide our work on COVID-19 and, in just the first week of establishing the group, 37 people provided comments on the development of an app to estimate the risk of dying during the pandemic based on previous health factors. This feedback and input has significantly altered the future focus for this piece of work.

Patient and public representatives are involved our weekly process to prioritise COVID-19 research questions. The aim is to ensure that questions the public see as important are being addressed.

We also aim to make all our key events accessible to the public, whether inviting people to join the event or ensuring it is live-streamed, recorded and published. An example of this was our event

Creating an inclusive, team-oriented culture

We are an organisation that lives our values and it is imperative that our policies and processes support us in doing that. Over the course of the year we have set out our principles and aspirations by which we aim to achieve this culture, in terms of: team science, open science, reproducibility, research assessment, diversity and inclusion, and attribution. These clearly set out the expectations of how we work in partnership to incentivise a culture of research and innovation that aligns with our Institute values of transparency, optimism, respect, courage and humility.

In June 2019, we held our One Institute event bringing together over 200 people from across our HDR UK community to focus on strategy, celebrate impact and progress and build towards our aim of the UK being recognised as the place to do health data science. We have further established our communication channels during the year with our website as the core channel. Whilst open to the public, is an important source of information for the HDR UK community, which many of our members contribute content to. It contains the latest news and events, project updates, HDR UK publications, and biographies of our 340 members. The website receives is now receiving an average of 8,300 unique visits a month.
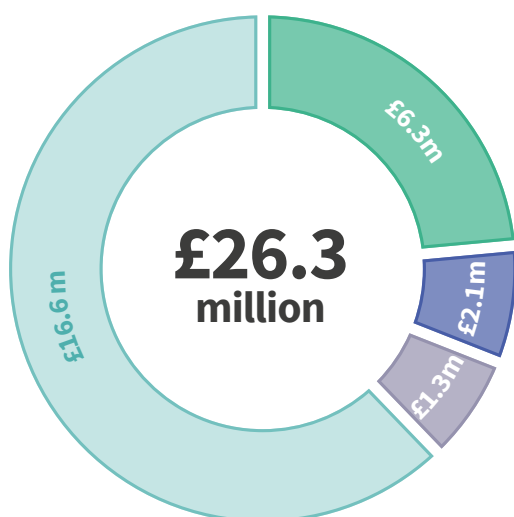
in January 2020 on Realising patient and NHS benefits from health and care data – from policy to practice, in partnership with the Academy of Medical Sciences and the Collaboration for the Advancement of Sustainable Medical Innovation (CASMI). This event was about building trustworthy commercial partners, for which we provided a live, open broadcast so that members of the public could participate.

£26.3 million

£6.3m
£2.1m
£1.3m
£16.6 m

Figure 7: Use of Funds to Support 2019/2020 expenditure

HDR UK Core Funders

Medical Research Council Capital grant

Industrial Strategy Challenge Fund

Complementary, leveraged, and matched funding, plus in-kind contribution

Implementing robust internal processes

In August, our quality management system was judged as compliant with the international standard, ISO 9001. We are among the very first research Institutes to achieve this certification and it demonstrates that our learning and improvement culture are built into how we work from the very start. 2019/20 also allowed us to test out some innovative procurement approaches, including those used to identify a lead supplier for the Gateway.

Funding sources and expenditure

HDR UK has built on its success in attracting major new funding streams in 2018/19, including the £37.5m Industrial Strategy Challenge Fund Digital Innovation Hub Programme. Our compelling programme of activity across uniting, improving and using data has generated significant interest, attracting a further £83.5m in 2019/2020 through new direct funds and partnership contributions. Our funds support long-term scientific and research studies, training and infrastructure that contribute to data science at scale, support our One Institute approach and deliver long-term impact on the health of patients and populations across the UK. We also raise thematic funding to support specific programmes of work.

Health Data Research UK's funders recognise the pivotal contribution of health data science to achieve transformative health benefits and the UK's ambition to remain a leader in life sciences. Our founding funders have jointly invested in Health Data Research UK over the next five years.
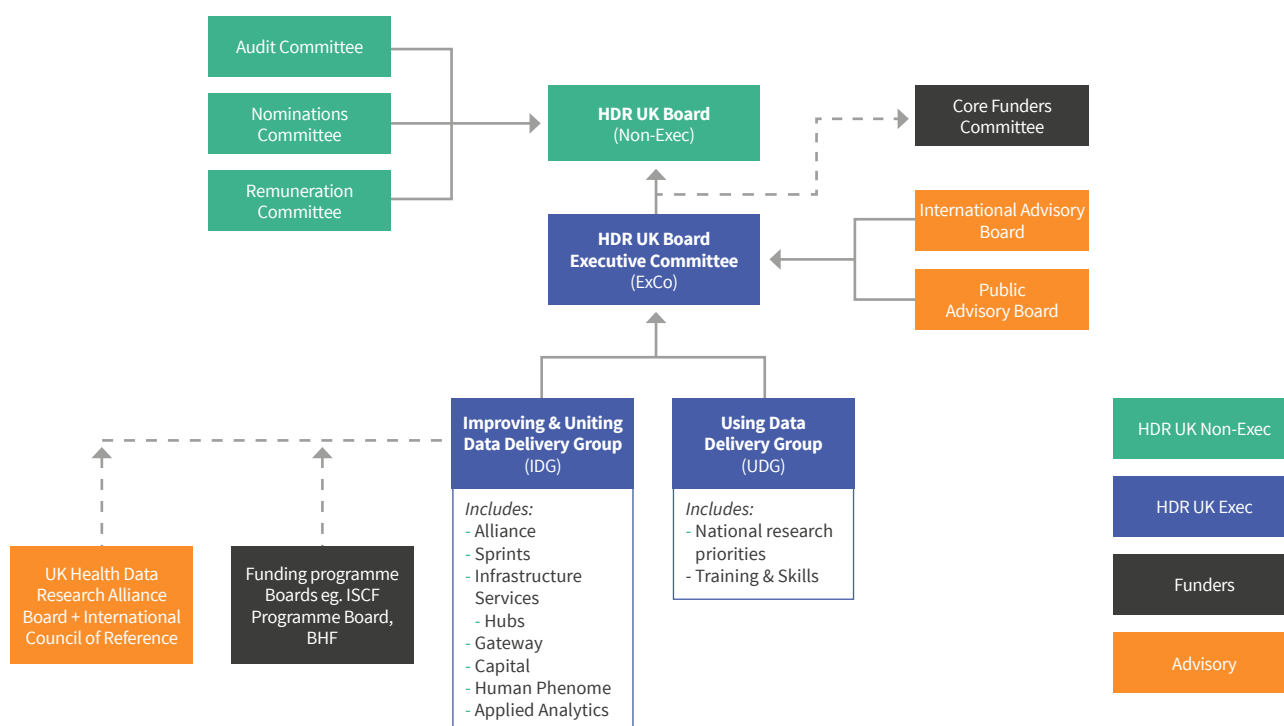

Figure 8: Summary of HDR UK governance structure

Governance

At the end of 2019/20 we reviewed and implemented changes to how we would organise ourselves in the future, to enable robust governance and clear accountability for the ambitious priorities set for the Institute. This included a new governance framework with an expanded executive committee, two new delivery groups and introducing a new Chief Executive role to run the Institute, provide strategic direction and leadership for our delivery programmes.

Our non-Executive Board

Health Data Research UK's Board members bring exceptional and diverse skills and expertise to support us in the delivery of our mission. The Board is responsible for the effective governance and development of the Institute, supports the Director in overseeing the delivery of our strategy, monitors key risks, and ensures resources are managed effectively.

The Board is supported in its role by three sub-committees: Audit and Risk, Nominations and Remuneration.
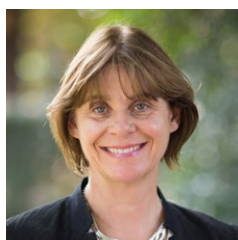
Sarah Harper   **Alex Markham**   **Jonathan Montgomery**   **Lucy Neville-Rolfe**   **James O'Shaughnessy**
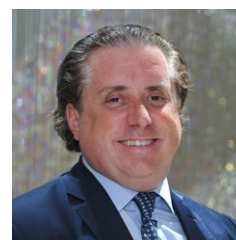
Frances Rawle   **Jim Smith**   **Graham Spittle Chair**   **Janet Thornton**   **David Zahn**

Find out more and get involved

Health Data Research UK is the national institute for health data that includes England, Wales, Scotland and Northern Ireland. Our mission is to unite the UK's health data to enable discoveries that improve people's lives.

By working in partnership with the NHS, industry, academia and patients, and providing safe and secure access to rich health data, our aim is to better understand diseases and discover new ways to prevent, treat and cure them.

Health Data Research UK is a not-for-profit public benefit company funded by UK Research and Innovation, the Medical Research Council, the British Heart Foundation, the National Institute for Health Research, the Economic and Social Research Council, the Engineering and Physical Sciences Research Council, Health and Care Research Wales, Health and Social Care Research and Development Division (Public Health Agency, Northern Ireland), Chief Scientist Office of the Scottish Government Health and Social Care Directorates, and Wellcome.

To find out more:
Visit our website **www.hdruk.ac.uk**
or follow us on Twitter **@HDR_UK**

HDRuk
Health Data Research UK

To find out more:

Visit our website **www.hdruk.ac.uk**
or follow us on Twitter **@HDR_UK**