



Annual Review 2021/2022





Contents

About us	3	Working in partnership as One Institute	26
		- Communications and engagement	27
Welcome	4	- Patient and public involvement and engagement	28
Highlights of the year	6	HDR UK Learning - attracting and developing talent	29
Accelerating the pace and scale of health and biomedical data science	8	Championing diversity in data science	31
Response to COVID-19	11	Plans for the future	32
- Data and Connectivity National Core Study	12		
- ICODA – A global health data response to COVID-19	14		
- BHF Data Science Centre	16		
Assembly of a UK-wide data infrastructure and services for health research	18		
- UK Health Data Research Alliance	19		
- Health Data Research Innovation Gateway	21		
- Health Data Research Hubs	23		
- The DARE UK (Data and Analytics Research Environments UK) programme	25		



About us

Who we are

Health Data Research UK is the national institute for health data science.

What we do

Our mission is to unite the UK's health and care data to enable discoveries that improve people's lives.

Our vision

Our 20-year vision is for large scale data and advanced analytics to benefit every patient interaction, clinical trial, biomedical discovery and enhance public health.





Welcome

In the year that the word data hit international headlines, Health Data Research UK (HDR UK) strengthened its leadership and demonstrated impact as the national institute for health data science.

As this annual report shows, 2021/22 was a pivotal year for HDR UK with challenges, learning and progress in equal measure. As we approach the end of our first five years, we have reflected with colleagues, partners, funders and advisors on progress to create an ambitious new five-year strategy starting in 2023. This strategy will be defining for HDR UK and will pave the way for the UK to harness the value of data in the most trustworthy way to transform knowledge and understanding of disease and boost medical innovation to benefit us all.

In our 2021/22 report, we highlight the significant research achievements made across the Institute, with over 2,300 research papers published and 66,000 citations. Many have directly informed policy decisions and healthcare practice during the pandemic where our research has delivered discoveries about the virus, its prevention, causes and management and provided insights into

vaccine efficacy and safety. We have enabled data discovery, access and linkage, that have underpinned the UK's pandemic response, through our joint leadership with the Office of National Statistics of the Data and Connectivity National Core Study. Our leadership of the International COVID-19 Research Alliance, led by our new Global Research Director Trudie Lang and delivered with partners across the world, has enabled actionable insights on the pandemic for global ministries of health and populations.

Our partnerships have continued to grow and flourish during the year and demonstrated the value of combining expertise to deliver further and faster on joint ambitions. Strategic partnerships with the [British Heart Foundation](#) and [NHS Digital](#), under the leadership of Cathie Sudlow, has enabled – for the first time ever – secure access to the largest linked health data resource covering 54.4 million people in England.





This year we joined forces with [Diabetes UK](#) to create a [Diabetes Data Science Catalyst](#), which we hope will be a model for collaboration with other charities.

HDR UK's leadership of the [UK Health Data Research Alliance](#) has delivered some of the greatest collaborative efforts to drive forward best practice and inform policy in health data research. This included white papers on TREs in partnership with [NHS England](#), data standards, and recommendations on building transparency in data research through standards on data use registers.

The year marked the second anniversary of the [Health Data Innovation Gateway \(the Gateway\)](#), our platform for researchers to discover and request access to UK health datasets, which lists over 780 datasets from more than 60 data custodians.

A new Cohort Discovery tool was launched to enable researchers to search by specific population criteria across multiple datasets, widening the data discovery capabilities of the Gateway even further and facilitating a faster pathway to impactful research.

Our network of Health Data Research Hubs continued to grow with the addition of two new Hubs; [DATAMIND](#) for mental health research and [Alleviate](#) to advance research into chronic pain. As a network, the hubs are pioneering and informing the development of health data research in the UK, demonstrating how insights from data are improving people's lives. To date, they have secured over 400 contracts for health data science projects.

Our commitment to working in partnership with the public and patients remains a key part of our strategy with our leadership to ensure transparency and build trustworthiness in data initiatives. HDR UK's innovative approaches to bring public views in quickly and effectively has shaped research and improved outcomes with over 14,700 patients and public contributing to our work this year. We have widened our public engagement and reach through social media, press campaigns, and events, showcasing the benefits and impact of data.

To support the future generation of data scientists we launched HDR UK Futures, our new virtual learning environment, to provide bitesize training from UK's experts on areas including phenomics, data engineering and data access. In response to the significant underrepresentation of Black people in data science this year, we launched the [Black Internship Programme](#) with 48 interns matched to data roles in 23 of our Alliance partner organisations - many going on to secure roles in the field.



The achievements highlighted in this report are a result of the commitment and dedication of the many teams across HDR UK's federated national network, our partners and our patient and public advisors. It is a testament to their focus on our values of transparency, optimism, respect, courage and humility that we have made such progress.

As we transition to our future strategy, we look forward to working with our partners and community, strengthening our connections and focusing on the delivery of novel insights and innovative data research and services to enable discoveries that improve people's lives.



Highlights of the year

PRINCIPLE Trial, supported by HDR UK, finds budesonide – a cheap drug used to treat asthma – helps people with early COVID-19 symptoms recover more quickly at home.

April
2021

May
2021

June
2021

July
2021

August
2021

September
2021

HDR UK's innovative partnership with the Health Foundation delivers exemplar outputs for use of data-driven analytics in clinical practice, through three Better Care projects.

Through the Data and Connectivity National Core Study, HDR UK in partnership with Office for National Statistics makes COVID-19 datasets and data linkages available for research across the network of UK TREs.

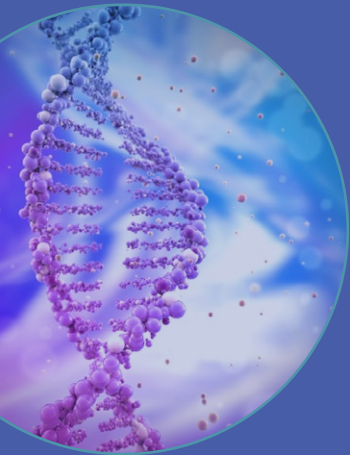
Two new Health Data Research Hubs are announced: DATAMIND to provide data for mental health research and Alleviate to support research to tackle chronic pain.

To tackle the underrepresentation of Black health data scientists, HDR UK launches its first Black Internship Programme with 48 interns starting placements with 23 Alliance partners.

The International COVID-19 Data Alliance (ICODA) and the Grand Challenges initiative announces ten global projects to accelerate knowledge of COVID-19 to address urgent research questions relating to the pandemic.

Research supported by HDR UK provides genomic risk prediction of coronary artery disease in women with breast cancer, enabling better-informed decisions about treatment.





HDR UK and The Alan Turing Institute award £2m to nine new studies to address priority research questions using large-scale linked data that will improve understanding of the pandemic and inform the continued policy response.



HDR UK welcomes seven new Public Advisory Board members to provide strategic advice on its direction of work, including its Public and Patient Involvement and Engagement (PPIE) strategy.

HDR UK researchers study large-scale health records to understand the impact of diabetes medication on outcomes in patients with ischaemic heart failure.

October
2021

OMICSPRED, a community resource supported by HDR UK, creates models that allow researchers to use DNA to predict the level of molecules - risk markers for certain diseases (such as cardiovascular disease or heart attacks) - present in blood.

November
2021

HDR UK and the Alliance publish two breakthrough papers to support standards in data research:

1. **Recommendations for data standards to encourage improvements in the utility and usability of data**
2. **Principles to improve public confidence in access and use of data for health research through TREs**

December
2021

January
2022

February
2022

Research, supported by HDR UK, finds that obesity dramatically increases the risk of death from COVID-19 but the extent of this risk across different body weights and ethnic groups still needs research.

HDR UK submits its Quinquennial Review report to core funders – setting out the impact the Institute has achieved since being established in 2018 and proposals for the next five years (from April 2023 to March 2028).

March
2022





Accelerating the pace and scale of health and biomedical data science

National Research Priorities

HDR UK's six National Research Priorities on Applied Analytics, Human Phenome, Understanding Causes of Disease, Clinical Trials, Improving Public Health and Better Care have produced UK-wide research that no single research organisation could achieve alone.

Highlights of HDR UK's research impact during the year are provided on the following pages.

Between April 2018 and March 2022,
HDR UK researchers published...

2,300+ papers

...which have received

66,000+ citations

...and delivered research that **directly informed policy decisions and clinical practice** during the COVID-19 pandemic*.



FURTHER EXAMPLES CAN BE VIEWED **ON OUR WEBSITE**

* Source: ResearchFish report received 28 March 2022. Includes outputs by a subset of HDR UK members published since April 2018 (HDR UK establishment) and attributed by submitters as HDR UK Institute outputs. Snapshot of number of citations from Dimensions AI 14 April 2022.



Better Care - the value of triage during periods of intense COVID-19 demand

Very little work has previously been done on the triage of patients who might need intensive care support. As COVID-19 cases mounted, scientific evidence was needed about which approaches for rational decision-making on the use of limited resources would yield the best outcomes. HDR UK created an open-source computer simulation model using records of 9,505 patients to assess the life years saved for triage strategies under different scenarios. This research can provide a basis for policymakers and care providers to start preparing protocols for future health emergencies.



Collaboration across the clinical trials landscape

Currently only a very small proportion of clinical trials access and use routinely-collected data – despite its potential to make trials more efficient across finding, recruiting and following up patients. HDR UK established multidisciplinary partnerships to address the technical, ethical and logistical requirements of providing clinical trial researchers with rapid access to large, linked datasets and integrated patient engagement. These novel collaborative approaches have helped researchers significantly enhance recruitment and subsequent delivery of trials – including RECOVERY, PRINCIPLE and TACKLE - for research into COVID-19.



HDR UK programmes responding to COVID-19

During the pandemic, the Welsh Government needed to understand COVID-19 in the population – from its spread to the effectiveness of control policies. A system was needed to track the pandemic and answer new research questions as they arose. The HDR UK Multimorbidity project was rapidly pivoted to create a TRE linking population-level data for all 3.1 million people in Wales to rapidly analyse COVID-19 and discover what works to prevent and treat the infection. The findings (for example, identifying vectors of transmission and assessing vaccine effectiveness by socioeconomic group) were used by the Welsh Government, the Government's Scientific Advisory Group for Emergencies (SAGE) and the New and Emerging Respiratory Virus Threats Advisory Group (NERVTAG) to inform national policy and save lives.

The HDR UK Multimorbidity project was rapidly pivoted to create a TRE linking population-level data for all 3.1 million people in Wales



Polygenic Score Catalog

Polygenic risk scores (PGSs) have immense potential for helping predict people's likelihood of developing many serious illnesses. However, progress is hindered by a lack of common standards or best practice. HDR UK co-developed the new Polygenic Score Catalog – an open resource of published PGSs that includes full scoring information annotated with expertly curated metadata. It allows PGSs to be re-used and evaluated so their accuracy and clinical value can be fully assessed. The catalogue has transformative potential to help a nascent area of research become of real clinical value. It is already extensively used and holds >2,100 PGSs from >260 papers for >500 different traits, demonstrating its capacity to harmonise and unify them in one database.



Multi-omics consortium to inform disease aetiology and prediction

Studies that combine genetic information with other blood-based factors – including proteins, metabolites and lipids – and health records, have the potential to provide more direct insight into disease causes and prediction. A key challenge, however, has been accessing this information at sufficient scale. In response, HDR UK convened the Multi-omics Cohorts Consortium to bring together molecular research cohorts from across the UK nations, comprising over 800,000 participants, to test the feasibility of integrating the extensive molecular measurements across cohorts. The consortium has improved the scale and statistical power of research into the cause and prediction of disease, for example, the use of proteomic and transcriptomic data to identify causal factors related to COVID-19 outcomes – and laid foundations for infrastructural innovation to improve future research and outcomes for patients.

HDR UK convened the Multi-omics Cohorts Consortium to bring together molecular research cohorts from across the UK nations, comprising over 800,000 participants

A new method to link health records of mothers and babies

Targeted support for women before and during pregnancy has the potential to improve outcomes at and after birth, through promoting preconception health and reproductive choices and mitigating the adverse effects of maternal stress. Understanding which women are most vulnerable, and how to identify them in time for intervention, is the first step to the effective development of early health and social support programmes. However, existing studies evaluating the association between maternal risk factors and infant outcomes have mostly focused on single risk factors. HDR UK researchers developed a probabilistic linkage technique to analyse the association between multiple maternal psychosocial risk factors that can be identified routinely in early pregnancy. By identifying multiple risk factors associated with poor infant outcomes, this research clearly shows the need for risk-factor informed interventions before, during, and after pregnancy – and is expected to influence the current update of the [Healthy Child Programme \(HCP\)](#).



Response to COVID-19

During the year, health data research continued to provide enormous value to people across the UK in response to the COVID-19 pandemic.

HDR UK's strategy enabled us to mobilise expertise and partnerships across the UK to coordinate research efforts, accelerate safe and trustworthy access to priority data and leverage the best of the UK's health data science capability to address the wider impact of the COVID-19 pandemic. These efforts have delivered scientific discoveries about the virus, its prevention, causes and management.

The portfolio of research and studies facilitated by us have generated actionable insights about COVID-19 including:

- **Over 20 papers demonstrating the effectiveness and safety of COVID-19 vaccines** through the EAVEII studies led by [HDR UK BREATHE Hub](#)

Director Aziz Sheikh. These include research showing the [risk of heart complications](#) and [blood clots](#) from the COVID-19 vaccine and that the benefits of the vaccine still outweigh the risks of COVID-19, quantification of the vaccines' effectiveness of [preventing deaths from the delta variant of COVID-19](#), and the duration of protection it offers to COVID-19.

- **Identification of treatments for COVID-19** to [speed recovery](#) via the PRINCIPLE trial and [increase survival](#) via the RECOVERY trial.
- **Implications for other healthcare services** including [mental health](#) and cancer care.



Data and Connectivity National Core Study

HDR UK's joint leadership with the [Office for National Statistics \(ONS\)](#) of the [Data and Connectivity National Core Study](#) continued during the year.

One of [six National Core Studies Programmes](#), the study works with policymakers, data custodians, providers of TREs, researchers and the public to make COVID-19-relevant data available for research with greater speed, efficiency and scale.

Over the last 12 months, the programme has continued to underpin the [UK's COVID-19 response](#) and contribute to future pandemic preparedness. Notable outputs include deepening our understanding of the [Omicron variant](#), which became dominant throughout 2021 and 2022. Data and Connectivity-enabled research has demonstrated that Omicron infection is usually milder than the Delta variant, but more likely to break through immune protection due to changes in its spike protein and cell-entry mechanisms.

The programme has also provided valuable insights into [vaccine efficacy and safety](#) that are important for the UK's ongoing pandemic response, using large-scale data to show vaccines offer [high protection against infection](#) for six months, and that this persists for people of all body weights.

As well as enabling key scientific insights, an important objective of the programme is to support the development of data infrastructure and address unmet data linkage needs that will strengthen future health threat responses.

In the last year, the programme has continued to convene multi-organisation groups to drive progress in this area including continuing to develop a national UK TRE network across the four nations for the first time.

[NHS Digital](#), [ONS](#), [SAIL Databank](#), [Northern Ireland Honest Broker Service](#) and [Public Health Scotland/Edinburgh Parallel Computing Centre](#) have now collaborated as a single UK-wide team to enable data linkage and research access to 91 UK priority datasets for urgent COVID-19 research.





The [Outbreak Data Analysis Platform \(ODAP\)](#) led by Kenny Baillie, University of Edinburgh and Sharon Peacock, [COG-UK](#), and supported by [Data and Connectivity](#) is providing an accessible data resource to enable research relevant to pandemics. It brings together outbreak-relevant data from clinical records, research studies and audit data into a streamlined data access process allowing a UK-wide, single panel of data custodians to provide transparent information governance adhering to the [Five Safes principle](#).

In September 2021, the COVID-19 Data and Connectivity National Core Study (NCS), partnered with [The Turing Institute](#) to fund nine projects focused on both using advanced analytics to contribute to our understanding of COVID-19, and leaving a legacy for future research studies by enhancing the value of data and following best practice in open science, sharing code and tools.

The selected projects are being delivered by 16 collaborating universities across the four nations using the health data research infrastructure and network of TREs established and supported by the Data and Connectivity NCS.

All datasets are being made discoverable and available (upon request) via the [Health Data Research Innovation Gateway](#).

COVID-19 National Core Studies
116 datasets, 1 data use, 422 papers, 8 tools, 12 people

At the start of the Covid-19 pandemic, experts and key funders identified a number of areas where the UK needed to increase its research scale and infrastructure to respond to near-term strategic, policy and operational needs regarding Covid-19. As a result, the National Core Studies (NCS) programme was established in summer 2020 to ensure critical questions can be answered quickly and to the best of our capability. The National Core Studies programme is enabling the UK to use health data and research to inform both our short and long-term responses to Covid-19, as well as accelerating progress to establish a world-leading health data and research infrastructure for the future.

Collection	Author(s)	Resources
COVID-19 Longitudinal Health & Wellbeing National Core Study	Aisha Agbaje	61 resources
PROTE A COVID-19 National Core Study on transmission and environment	Aisha Agbaje	41 resources
Data & Connectivity	Paola Quattroni, Aisha Agbaje, ...	599 resources
Immunity	Aisha Agbaje, Sophie Morris, C...	74 resources
Epidemiology and Surveillance	Aisha Agbaje, Sophie Morris, A...	1 resource
Clinical trials infrastructure & support - Therapeutics and Vaccines	David Seymour, Sophie Morris, ...	1 resource
Courses for accessing Trusted Research Environments	Sophie Morris, Lara Edwards, J...	4 resources

Over the last 12 months, the programme has continued to underpin the UK's COVID-19 response and contribute to future pandemic preparedness.



ICODA – A global health data response to COVID-19

Work undertaken by HDR UK to convene the International COVID-19 Data Alliance (ICODA), which was announced in June 2020 has come to fruition in the last 12 months.



The ICODA initiative, funded by the COVID-19 Therapeutics Accelerator, a large-scale initiative supported by the [Bill & Melinda Gates Foundation](#) and [Minderoo Foundation](#) and other donors, has twelve driver projects – specific research questions bringing together a range of datasets for analysis, all approaching completion. Ten of these were selected from a Grand Challenges ICODA pilot funding call announced at the end of 2020 which resulted in 418 applications from 69 countries.

Across the [twelve ICODA driver projects](#) we have an active cohort of 135 researchers from 19 countries working in our TRE, the ICODA Workbench. This provides them with a wide range of analytical tools, wrapped with rigorous [data governance processes](#) and a mechanism to share metadata for secondary data use through our ICODA Gateway. Our

researchers are working with a broad range of data from health systems, clinical trials, electronic health records, population archives, hospital admissions, and birth and vaccine records.

[Driver Project REHCORd](#) which is working on understanding the impact of COVID-19 on health service delivery and institutional mortality across multiple countries used an interrupted time series design to assess the immediate effect of the pandemic on 31 health services in 10 low to middle income settings. Despite efforts to maintain health services, disruptions of varying magnitude and duration were found in every country, with no clear patterns by country income group or pandemic intensity. They found that often the [disruptions in health services preceded COVID-19 waves](#).

“New data being generated from research into COVID-19 are hugely valuable. We have to unlock the full value of these assets by making them findable, accessible and securely usable by researchers around the world working in the public’s interest.”

Tariq Khokhar Wellcome



There is an active ICODA researcher community of practice with teams now contributing to the development of a wider global health data science community, hosting regular webinars and presentations. Working in partnership with [The Global Health Network](#), and pivoting to focus on community enablement, we have launched the [Global Health Data Science Hub](#), now with over 1,000 members. The Hub working group meets monthly, with members from across the globe, encouraging collaboration and knowledge exchange between health data scientists.

In September 2021, Professor Trudie Lang joined HDR UK as the Global Research Director. Professor Lang is responsible for leading a new programme, HDR Global. This builds on the ICODA initiative, with the objective of improving health data science practices globally, using learnings acquired from HDR UK and ICODA. To support the objective, a five-year grant has been awarded from the Bill & Melinda Gates Foundation to take forward our HDR Global work alongside The Global Health Network and partners in Africa, Asia and Latin America.

As the ICODA initiative comes to a close in the second half of 2022, we are sharing open data governance processes, tools and best practices and infrastructure lessons learned to ensure we are better positioned for future global health challenges.

Overall, the ICODA driver project cohort continue to have a strong impact, with 21 publications, articles and pre-prints to-date, in journals including BMJ Global Health, The Lancet and Nature Magazine, and dashboards informing ministries of health and populations, all of which are featured on the ICODA website.





BHF Data Science Centre

Launched in January 2020 as a partnership between HDR UK and the British Heart Foundation, the [BHF Data Science Centre](#) enables responsible, ethical research that combines the power of advanced analytic methods with the UK's large-scale and diverse cardiovascular data.

The work of the BHF Data Science Centre has grown substantially over the year. The impact of their work was presented at our [Research Showcase](#) event in March 2022. Over 500 people registered to join the event to hear key speakers including Professor Nilesh Samani, Medical Director of the BHF, and Kelvin Pitman, patient contributor to our Oversight Committee. The virtual event highlighted activities, such as enabling linked health data research across the UK, pioneering large-scale cardiovascular imaging, and advancing data-enabled clinical trials.

The CVD-COVID-UK consortium, a NIHR-BHF flagship project which was set up in 2020 to understand the relationship between COVID-19 and cardiovascular diseases, has enabled secure access to a linked health data resource covering 54.4 million people in England. The data is available to UK researchers to collaborate in NHS Digital's secure TRE and has

contributed to research including a [study](#) that suggested that blood-thinning medications in patients with cardiovascular disease is associated with lower risk of death from COVID-19.

Several workshops during the year have informed our plans for our thematic areas including [cardiovascular imaging](#), [wearables](#), and [data enabled clinical trials](#). These workshops have involved representatives from the NHS, academia, policy groups and the public.

In the summer of 2021, we joined forces with Diabetes UK and the British Heart Foundation to co-produce a strategic partnership that has evolved into the Diabetes Data Science Catalyst. This strategic partnership is a model for collaboration with other charities.





Research outputs from the centre during the year included:

- A [study](#) in PLoS Medicine led by the CVD-COVID-Consortium showing elevated risk of thrombosis following vaccination with the AstraZeneca COVID vaccine, although lower than that with infection. This study was widely cited in the mainstream press
- A [study](#) published in BMJ on establishment of a TRE enabling safe and secure whole population research on Covid-19 and cardiovascular disease (Wood and Sudlow).

Our webinar series showcasing our work, inspiring new ideas and generating insightful discussion attracts over 50 attendees each month.

Engagement with our public and patient representatives has grown over the year including participation at workshops, presentations at events and guest blogging such as the opinion piece [Data: What is it good for?](#)





Assembly of a UK-wide data infrastructure and services for health research

HDR UK is delivering not only technology but the underpinning governance, ethics, standards, public engagement and data curation to enable health data research.

HDR UK has created the [UK Health Data Research Alliance](#) and the [Health Data Research Innovation Gateway](#), facilitated the creation of the network of [Health Data Research Hubs](#) and convened a network of TREs across the UK.

This has enabled discovery and safe research access to over 750 datasets held by 60 data custodians (out of an estimated 5,000 UK health

datasets currently in research use) using, for the first time, linked UK-wide data on more than 65 million people.

Funded by UK Research and Innovation's Industrial Strategy Challenge Fund, our improvements have made access to data easier resulting in medical discoveries and breakthroughs for industry, the NHS and patients across the UK.

UK Health Data
Research Alliance

Health Data Research
Innovation Gateway



UK Health Data Research Alliance

The Alliance is a growing partnership of leading healthcare and research organisations who have united to establish best practice for the ethical use of large-scale health data for research.

In a year that has delivered some of the greatest collaborative efforts to respond to the pandemic, the Alliance has been a driving force for this approach, bringing together multiple health and research organisations to work in partnership to create a national, federated and coordinated approach to health data research infrastructure.

During 2021/22, membership of the Alliance grew to 67 members with new members announced in [May 2021](#), [November 2021](#) and [January 2022](#).

Together the Alliance members are making significant progress to establish best practice and inform policy in health data research.

In December 2021, the Alliance published a [White Paper on principles, approaches and guidelines for data sharing and linking TREs](#), as well as examples of best practice drawing on the experiences of Alliance members and their colleagues across the world. The guidelines are for UK data custodians and other organisations involved in data sharing and information governance both within and beyond the health sector and are structured around the 'Five Safes' framework for the access of health data. The paper also sets out principles for the federation of TREs describing how they can be linked to substantially enhance the scale of secure and trustworthy data linkage and research.

In January 2022, the [Alliance published recommendations](#) for a data use register standard to be adopted by organisations responsible for the safe sharing of data used for research and innovation, following consultations and workshops with health data research organisations, universities, data custodians, patients and the public.



A data use register – also known as a data release register or list of approved projects – is a public record of how data is being used for research, by whom and most importantly for what purpose, and it offers a clear demonstration of the benefit and value of using health data. The Alliance is now working with data custodians to support adoption of this standard and improve transparency in data use.



Health data research to improve people's lives relies on a set of common standards for healthcare data and metadata. The Alliance has carried out data user and custodian surveys with academic researchers, charities, data custodians, healthcare providers, life science companies and AI and technology companies to understand the current use of health data standards and opportunities for greater alignment.

The surveys identified:

- A higher level of data standards expertise in industry compared with academia
- Two thirds of health data users had basic or no data standards expertise
- Almost 90% of users were in support of defining a core set of data standards to enable health data research
- Data custodians and data users were discovered to be using a wide range of data standard and data models.

Based on this work, in November 2021 the [Alliance published a White Paper](#) making recommendations for data standards. The paper encourages improvements in data usefulness and usability and was developed with input from the health data community, patients and the public.





Health Data Research Innovation Gateway

The [Health Data Research Innovation Gateway](#) (the ‘Gateway’) was established in 2020 and is the UK’s only unified platform to search, discover and request access to health datasets for research and innovation.

Since then, over 750 datasets from more than 60 data custodians have been made available for researchers to search and request access to, in addition to over 3,000 other health data resources including publications, data uses, tools and educational courses.

The Gateway is not limited to being a metadata registry but forms an entire ecosystem for fostering the curation, quality improvement, access, and ethical use of health data. To date, the Gateway has over 2,000 registered users from across the health data community who are making more than 10,000 searches a month.

Throughout 2021, COVID-19 resources continued to be added to the Gateway which inevitably shifted towards vaccine-related datasets including [vaccination status](#) and [adverse reaction data](#) from NHS Digital. The [COVID-19 Symptom Tracker Dataset](#) remains the most viewed resource on the Gateway and continues to support scientific discovery, including this recent [publication](#) in Nature Communications.

In April 2021, the [Cohort Discovery Tool](#) was launched. Delivered in partnership with [CO-CONNECT](#) and funded by UK Research and Innovation, the tool enables researchers to search by specific population criteria across multiple datasets, widening the data discovery capabilities of the Gateway even further and facilitating a faster pathway to impactful research. 11 Gateway datasets covering over 70,000 subjects are now available for researchers to discover using Cohort Discovery – the latest being the [COVIDsortium study](#) from [University College London](#).



In October 2021, additional improvements were made on the Gateway to support data discovery with the release of the [Data Utility Wizard](#). The tool allows researchers to search for and discover datasets listed on the Gateway using specific criteria and filters. The simple, user-friendly interface helps evaluate how useful specific healthcare datasets will be via a series of simple questions. The development of the tool is based on HDR UK’s [Data Utility Framework](#), transforming it into a practical tool that supports the research experience.





More recently, the focus has been on the development of a [data use register](#) to publicly show how datasets published on the Gateway are being used, by whom and for what purpose. The functionality implements a [national standard for data use registers](#) created by the Alliance. It aims to improve transparency in the use of health data for research and provide a best practice example of how information about approved data uses is shared and managed. More than 800 data uses have been uploaded to the Gateway since its release in early 2022.

Two years after its creation, the Gateway is now a recognised platform for researchers to discover and request access to UK health datasets, offering a simplified and standardised approach to data access management and supporting transparency and trustworthiness in the use of health data for research across the wider community.

Health Data Research Innovation Gateway

Explore Help Usage Data About Us News Community

Search: I'm looking for... [Advanced Search](#)

datasets 783	tools 192	data uses 845	courses 199	papers 2,400	dataset requests 566	registered users 2,349	searches last month 10,218
-----------------	--------------	------------------	----------------	-----------------	-------------------------	---------------------------	-------------------------------

[View all usage data](#)

For Researchers
Find and request access to datasets and other useful resources for your research such as projects, courses, tools and publications.

For Data Custodians
Register your datasets to make them more easily discoverable for research and manage data access requests with automated workflows.

For Patients and Public
Discover what data is being used for research and the projects and publications that have resulted from this.

More than 800 data uses have been uploaded to the Gateway since the release of the data use register in early 2022.



Health Data Research Hubs

The [Health Data Research Hubs](#) provide a rich toolkit of healthcare datasets, infrastructure and capabilities that enable users to identify, access, understand and use data to improve people's lives.

Working in partnership with HDR UK, each Hub is the product of strong collaboration between the NHS, academia, industry and the public and forms an integral part of the vibrant UK research data ecosystem. Our pioneering network of Hubs is informing the development of health data research in the UK, and globally, demonstrating how insights from data are improving people's lives.

Since establishment, the Hubs have delivered remarkable work including:

- Making over 200 datasets discoverable via the [Gateway](#)
- Delivering over 500 projects involving hundreds of partners
- Carrying out over 30,000 meaningful patient and public interactions
- Delivering over 2,300 training activities
- Reducing the time to access data to 26 days (median time to get a decision on a data access request).

During the year we announced two new Hubs joining the network, [DATAMIND](#) and [Alleviate](#). The DATAMIND Hub aims to improve the use of big data for mental health research by providing expert data services, tools and expertise to a wide range of users.

Mental illness affects almost everyone at some point in their lives and, to varying degrees, is the biggest cause of ill-health for people living in the UK. Evidence shows us that people living with a mental illness are less likely to take part in or to remain in usual types of research studies and may also be actively excluded. This means that much of what we know about mental health is not based on the very people who are worst affected.

Through DATAMIND, researchers can access UK data regarded as some of the best data in the world, which can be used in research to understand how to prevent mental ill-health and maintain and improve people's lives. DATAMIND will continue to maximise the value of this data by safely and securely bringing together data from diverse sources, including health





records, schools and administrative data, charity data, research trials, genomics, longitudinal studies and cohort data.

The other hub to join the network this year is [Alleviate](#). In partnership with the [Medical Research Council \(MRC\)](#) and [Versus Arthritis](#), HDR UK is leading the Alleviate: APDP Pain Research Data Hub (Alleviate Hub) to advance necessary research into chronic pain. Chronic pain is a major unmet public health challenge, with little known about its exact cause. Many pain-related datasets from research studies already exist across the UK; however, until now, there has been no national approach to coordinating and managing this data. [The Alleviate Hub](#) will address this challenge by safely and ethically bringing together pain-related data from diverse UK research cohorts allowing researchers and innovators to access linked pain-related health data.

“

INDUSTRY

The Hubs' 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

“

INNOVATIVE SMEs

The facilitation of data-sharing agreements, establishment of a trusted research environment and relationships with key UK academic groups was significantly quicker and smoother via BREATHE Hub than it would have been if we had to negotiate separately with every institution involved.

Steve Gardner

Chief Executive, PrecisionLife

“

PUBLIC

Through the work of DATA-CAN - getting data to the people that need it, such as cancer researchers - we are able to get better diagnostics. And we know that the sooner cancers are diagnosed, the better patient outcomes tend to be.

Yvonne Adebola

Global programme director, Healthcare, AstraZeneca

“

CHARITIES AND PATIENT ORGANISATIONS

Our collaboration with HDR UK is enhancing a lot of the work that we are doing in many different ways. We're developing skill sets within our teams, which have been supported by the work that HDR UK does.

Dr Sarah Clarke

Interim Director of Data and Quality Improvement, Cystic Fibrosis Trust



The DARE UK (Data and Analytics Research Environments UK) programme

[DARE UK \(Data and Analytics Research Environments UK\)](#) is a [UK Research and Innovation \(UKRI\)](#) programme which aims to design and deliver a more joined-up, efficient and trustworthy national data research infrastructure. This is to support safe and secure cross-domain analysis of sensitive data at scale for public good.

DARE UK

Phase 1 of the DARE UK programme, which began in July 2021 and is jointly led by HDR UK and [ADR UK \(Administrative Data Research UK\)](#), has been an extensive listening exercise with stakeholders across the UK including researchers, technologists, the public and others. The goal has been to understand, through open dialogue, what is needed to enable more efficient, coordinated and trustworthy data research. The aim is to ensure that subsequent phases of DARE UK work to address the needs of the UK in making the best use of data at scale for public benefit.

Engagement so far during DARE UK Phase 1 has included several key complementary activities. An [initial landscape review](#), published in October 2021, engaged broadly across the UK research and innovation ecosystem to establish an initial, broad framing of the key focus areas for the landscape that could be taken forward into further investigation, analysis and engagement. The review involved 60 interviews with a total of 79 stakeholders selected

from across the spectrum of research disciplines, as well as two open workshops with approximately 100 researchers and technologists.

In January 2022, a portfolio of nine [Sprint Exemplar Projects](#) was initiated, having been competitively selected at the end of 2021, with a total award of over £2 million from UKRI. The projects are focused on uncovering and testing early thinking in the development of a coordinated national data research infrastructure to inform future directions of DARE UK. Delivery of project outputs is scheduled for the end of August 2022.

In January and February 2022, a [UK-wide public dialogue](#) was held via a series of deliberative workshops with 44 members of the public. The aim was to deepen public conversation around data research practices and capture tangible actions that could be taken forward to better address public views. The dialogue found people want trustworthy

researchers to have better access to their sensitive data for research in the public interest, but that more proactive transparency is needed around who is accessing sensitive data, how and why.

In March 2022, a [series of six virtual workshops](#) was held to seek further input and feedback on a set of emerging recommendations shaped around six thematic areas of focus in the context of a coordinated national data research infrastructure. The input from the workshops has informed further development of the recommendations.

The findings of the DARE UK Phase 1 activities have now been incorporated into a findings report with a comprehensive set of recommendations for the delivery of a coordinated national data research infrastructure.



Working in partnership as One Institute

HDR UK is a multi-disciplinary, geographically distributed, UK-wide Institute of over 1,500 researchers across >39 organisations. During the year, the Institute has engaged a broader UK and increasingly global community spanning the National Health Service (NHS), research institutes, industry, charities, government and regulators with >500 organisations collaborating on programmes of health data research initiated and enabled by the Institute.





Communications and engagement

During the year, HDR UK's work with the media (broadcast, radio, press journalists) around key activities, events and campaigns, has significantly increased HDR UK's reach. It has raised awareness and boosted engagement in research projects, training and public engagement.

In June 2021, HDR UK enhanced its reputation as the leading voice in the health research community in response to debate around the implementation of [General Practice Data for Planning and Research \(GPDPR\)](#). The communications team successfully repurposed an existing #DataSavesLives mini campaign to support the case for GP data use in research, centred around a [Joint Statement](#) that was signed by over 250 leading health researchers, as well as maintaining a drumbeat of increasing activity around key HDR UK and national moments. The campaign also included a briefing hosted by the [Science Media Centre](#), attended by 19 journalists, resulting in national broadcast (BBC, Channel 4) and print coverage with quotes from Cathie Sudlow and Martin Landray.

The campaign received overwhelming positive feedback from members of the UK health research community, for helping to redress the balance of the debate in the national media and put forward the positive case for the secure, trustworthy use of health data for research in the face of negative headlines.

The HDR UK website receives an average of 28,000 users a month, a 106% increase on 2020 and a 349% increase on 2019. The HDR UK website links to the [Alliance website](#), the [HDR UK Futures](#) training platform and the [Gateway](#) to provide users with a clear route to discover more detailed, tailored information in response to their user need. The quality and quantity of social media output has increased throughout the year, improving since the start of 2021 with the recruitment of dedicated digital communications expertise into the team. At the end of January 2022, HDR UK had over 9,000 followers on Twitter, with an average of 326,000 impressions per month, and over 4,500 followers on LinkedIn. The main HDR UK newsletter, HIVE, which provides a concise update on key activities and developments across the organisation has over 4,900 subscribers.

HDR UK's annual scientific conference in June 2021 – [Data Insights in a Pandemic](#) – demonstrated the insights and value that a data-led approach has

delivered during the pandemic; and showcased the scientific discoveries that establish a strong foundation for the future. The conference took place alongside the [COVID-19 National Core Studies Symposium](#), which exhibited the progress on each of the six National Core Studies and their impact on UK COVID science, life sciences industry, and policy. Both events attracted over 2,300 registrations.





Patient and public involvement and engagement

HDR UK has shown leadership across the sector in demonstrating transparency and trustworthiness to patients and by embedding public and patient involvement and engagement (PPIE) in research, infrastructure developments and governance.

HDR UK's innovative approaches to bring public views in quickly and effectively has shaped research and improved outcomes, for example, enabling agreement on national standards for data use registers – which are now visible through the Gateway, and coordinated public and patient engagement across data custodians. This year, over 14,700 patients and public contributors were involved and engaged in health data research.

During the year, we held a campaign to recruit new members to our Public Advisory Board, which provides strategic guidance to HDR UK.



[Seven members were appointed](#), bringing new voices and expertise to the Board. The work of our Public Advisory Board, public advisors in our Governance structure, our COVID-19 Patient and Public Involvement and Engagement Group and groups and panels set up to support the Health Data Research Hubs and research priorities, have all made a significant impact and ensured the work we do has public benefit at its core.

Our PPIE has been influential, during this time, driving forward a range of activities including:

- Proactively getting involved to help shape our strategy for HDR UK's next five years, including having a public advisor as a co-lead for our PPIE approach
- Key members of funding and/or review panels, including the Health Data Research Hub Milestones, Data and Connectivity Rapid Funding Calls and the Grand Challenges ICODA COVID-19 Data Science pilot initiative

- Shaping the direction of COVID-19-specific research and communications, and providing their valuable insights to HDR UK's [COVID-19 response](#) through regular reports to SAGE
- Making [recommendations to the Goldacre Review](#) on the need for meaningful PPIE in the efficient and safe use of health data for research
- Leading on the development of [guidance for public involvement in data access processes](#) across the Alliance and wider health data organisations
- Co-developing [guidance, including templates, for public involvement in funding calls](#) which can be adapted and used beyond HDR UK and sector-wide organisations
- Running our [#AtoZHealthData twitter campaign](#) to raise awareness and understanding of health data science whilst engaging the wider public.



HDR UK Learning - attracting and developing talent

Over the past year, the HDR UK Talent and Training team has taken major steps towards our goal of building a vibrant and diverse health data science workforce and responding to the UK's health data science skills shortage.



HDR UK Futures

Launched in 2021, our virtual learning environment HDR UK Futures now has over 1,700 users, houses circa 100 bitesize videos, and over 12 learning pathways in topics such as Health Information Engineering, Phenomics and Accessing Health Data, giving users access to a range of online events. A subsequent refresh has added features to make playlists, track training records and earn badges. Success is monitored via user behaviour metrics and feedback, currently averaging at 5 stars. More developments are underway with a signposting feature launching later in 2022.

Workshops and webinars

During the year a successful series of HDR UK and partner workshops and webinars has been run. A fully-subscribed health information engineering workshop will be repeated in 2022 following extremely positive feedback. Online webinars have included topics such as reproducibility and working in a TRE.

Outreach

In February 2022, we signed the Technician Commitment which aims to ensure visibility, recognition, career development and sustainability for technicians working in higher education and research. To reinforce this commitment, we profiled and shared on social media health data scientists in bitesize videos, with a LinkedIn alumni group set up to nurture this community.

We maintain a strong online presence; our monthly training E-bulletin has over 1,000 subscribers and digital marketing attracts new Futures users, and the careers and training pages are among the most visited on the HDR UK website.

Partnerships

We have continued to develop our partnerships with [Health Education England \(HEE\)](#), the [European Bioinformatics Institute \(EBI\)](#), [NHS Digital Leadership Academy](#) and more, and in March 2022



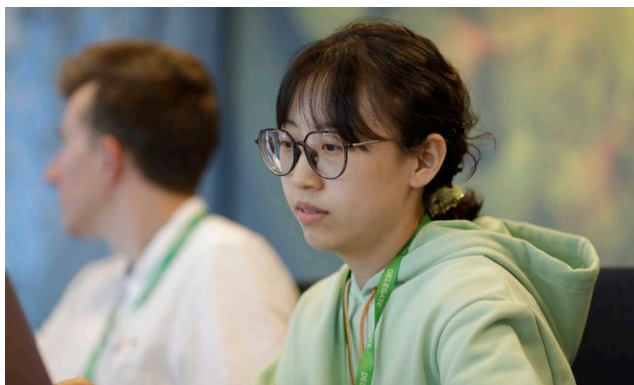
we joined the Turing's Data Skills Taskforce. We sit on the NHS England's Developing Data and Analysis as a Profession Board and have commenced resource-sharing, for example, European Molecular Biology Laboratory (EMBL)-EBI now host our bitesize videos.

Masters

Our seven master's programmes, delivered in partnership with six leading universities across the four nations, is now in its second year having successfully imparted students with quantitative skills. A third of students came from non-biomedical backgrounds and from the first cohort 25% went on to a PhD, 20% into graduate research positions, 30% into industry, 10% into NHS-based careers, and 5% are medical students.

HDR UK/Turing Wellcome PhD programme

Our flagship programme is now in its second cohort with nine students and seven universities. The first cohort demonstrated their skill through student-led collaborations, the development of unique PhD proposals and the inclusion of research papers at international conferences in 2021. With diversity and inclusion considerations at its core, over half of accepted students identify as either female, BAME or having a disability.





Championing diversity in data science

Black Internship Programme

Run by HDR UK and the Alliance in partnership with [10,000 Black Interns initiative](#), our Health Data Science [Black Internship Programme](#) continues an exciting endeavour for early career Black data scientists who are currently heavily underrepresented within the health data science sector. Our Health Data Science Black Internship Programme is just one of several steps that we are taking to start to address this underrepresentation.

In 2021, we recruited 54 interns (including 30 women) to work in placements at 25 of HDR UK's partner organisations including members of the Alliance.



This pioneering programme opened the door to significant opportunities for the interns, with many going on to full-time employment or further education in health data science. The programme positively impacted diversity in the health data science landscape not only by providing opportunity for the 50 but creating new future role models that many more can aspire to. HDR UK is repeating the programme in 2022, which is bigger and more impactful than the previous year.

Biomedical Vacation Scholarships

Launched in 2022, this programme gives six undergraduates in the middle of quantitative degrees the chance to undertake health data research projects. Promoting diversity, priority is given to candidates from socio-economic backgrounds currently underrepresented within health data science and those attending non-Russell Group universities. Two-thirds of the successful candidates met widening access criteria.





Plans for the future

In 2022/23 HDR UK will transition to its future five-year strategy. HDR UK's long-term mission, *to unite the UK's data to make discoveries that improve people's lives*, remains unchanged as we move forward.

From 2023-28, the Institute will continue to work with its partners and funders towards a future vision in which large-scale data and advanced analytics benefits every patient interaction, clinical trial, biomedical discovery and enhances public health.

The breadth of support for the Institute's vision and strategy is clearly demonstrated by the 59 letters of support received during our 2021 stakeholder engagement process, covering a wide range of NHS, industry, academic, charity and international partners. In its next phase, the Institute will continue to assemble and deliver user-led health data research infrastructure and services across

different geographies. By bringing together diverse health data science, engineering and delivery expertise, HDR UK will advance research discoveries and accelerate insights that benefit patients and the public across the UK and globally.

Through partnerships, the Institute will continue to find ways to narrow the gap between new research discoveries and direct benefits for patients and the public. Across common and rare infectious and non-communicable diseases, including cardiovascular disease, cancer and dementia, patients and the public will have safer, better and faster healthcare solutions to deliver equality in health outcomes.



In the second five years, HDR UK will focus on **three integrated areas of activity** to deliver this ambition:

Research Data Infrastructure and Services

Providing the UK-wide and global co-ordination and leadership of health data infrastructure and services required to make health-relevant data findable, accessible, interoperable and re-usable (FAIR). This will be built on the convening, collaborative and co-ordinating role of the UK Health Data Research Alliance and will comprise four Pillars of activity:

- **Assembling the technology services ecosystem**
- **Trust and transparency**
- **Developing the tools required to make data useable**
- **Building skills and capacity.**

Research Driver Programmes

Advancing research discoveries through high impact UK-wide programmes that:

- **Address major health and societal challenges**
- **Guide the development of the infrastructure and services for the benefit of other researchers**
- **Are outward-looking with global reach.**

One Institute Partnerships

Through national leadership with a clear vision and ambition to assemble an ambitious health data research ecosystem with enduring benefits for all researchers.

- **As an innovative distributed UK-wide and increasingly global institute, HDR UK will act as a flagship for team science, drawing on skills, resources, and expertise from academic, NHS, industry and government partners.**



OUR APPROACH FOR THIS AMBITIOUS FIVE-YEAR STRATEGY IS SET OUT IN THE [OVERVIEW OF THE QUINQUENNIAL REVIEW](#) AND OUR [STRATEGIC DELIVERY PLAN FOR 2022/23](#), BOTH PUBLISHED ON OUR WEBSITE.



Health Data Research UK is the national institute for health data science and registered charity (no. 1194431), funded by UK Research and Innovation, the Department of Health and Social Care in England and equivalents in Northern Ireland, Wales and Scotland, and leading medical research charities.

www.hdruk.ac.uk