














SECTION 1: QUINQUENNIAL REVIEW **OVERVIEW OF HEALTH DATA RESEARCH UK**

Navigation page

 All to read	 Infrastructure committee to read	 National Priority/ Driver committee to read
Section 1 Overview of HDR UK	Section 2: First five years (2018 – 2023): thematic report and individual programme reports	Section 3: Second five years (2023 – 2028): thematic proposal and individual programme proposals
1.1. Overview  1.2. The origins of HDR UK and its structure 1.3. Achievements, impact and added value 1.4. Future plans for the second five years 1.5. Future aims and objectives 1.6. Research governance and maintaining quality science	2.1. First five years thematic report  2.2. Individual programme reports: Research Data Infrastructure and Services  2.3. Individual programme reports: National Priorities  2.4. Individual programme reports: Leveraged programmes enabled by additional funding 	3.1. Second five years thematic proposal  3.2. Individual programme proposals: Research Data Infrastructure and Services  3.3. Individual programme proposals: Driver Programmes  3.4. Individual programme proposals: One Institute Partnerships 
Appendices 		

The HDR UK Quinquennial Review has been led, coordinated and delivered by Rhoswyn Walker, Alice Turnbull, Hollydawn Murray and Andrew Morris with significant contributions and support from team members across the institute and partner organisations.

Design, images and content support have been led by Amanda White and Joel Beckman. Financial reporting and planning was led by Alison Hopkinson and Lucy Gould.

HDR UK is hugely grateful to everyone for their leadership and support.

4 February 2022

Table of Contents

1	
Navigation page	2
Section 1: Overview of Health Data Research UK	5
1.1. Overview	5
1.2. The origins of HDR UK and its structure.....	7
1.2.1. Origins of HDR UK	7
1.2.2. HDR UK's first five years.....	8
1.2.3. Leadership, governance and delivery model	10
1.3. Achievements, impact and added value	13
1.3.1. Impact 1: Assembly of a UK-wide and international data infrastructure for health research	14
1.3.2. Impact 2: Accelerated the pace and scale of health and biomedical data science	15
1.3.3. Impact 3: Working in partnership as 'One Institute'	16
1.3.4. Additional investment secured	17
1.4. Plans for the second five years.....	18
1.4.1. Lessons learned	18
1.4.2. Strategy for the second five years (2023 - 2028)	21
1.4.3. Research Data Infrastructure and Services	23
1.4.4. Research Driver Programmes.....	25
1.4.5. One Institute Partnerships	27
1.5. Future aims and objectives	28
1.5.1. Metrics for success	28
1.6. Governance and maintenance of high-quality services and research	30
1.6.1. Alignment milestones	30
1.6.2. Research best practice	30
1.7. Literature cited	33

List of Figures

Figure 1: HDR UK first five years Institute programmes	12
Figure 2: HDR UK Black Internship Programme cohort 2021	17
Figure 3: Harnessing the UK's diverse health data ecosystem.....	21
Figure 4: HDR UK's strategy for the second five years.....	22
Figure 5: Second five years Research Driver Programme, Research Data Infrastructure and Services and One Institute Partnerships interconnected strategy	23
Figure 6: Uniting health data through delivery of FAIR Research Data Infrastructure and Services	24
Figure 7: Research Driver Programmes: Advancing health research discoveries and leaving a legacy	25
Figure 8: UK Regional Networks	27
Figure 9: HDR UK's journey from the first five years (2018-2023) to the second five years (2023-2028)...	29

See **Appendix 1: Glossary of acronyms and key terms** and **Appendix 2: Summary of figures**.

Section 1: Overview of Health Data Research UK

1.1. Overview

Health Data Research UK (HDR UK) is the national institute for health data science, established by a consortium of nine UK core funders¹ in 2018. **HDR UK's charitable mission is to unite the UK's health data to enable discoveries that improve people's lives.** At launch, the Director's vision was:

"To exploit the extraordinary capability of informatics to create a national medical research and innovation Institute that leads the international agenda in health data science.

The Institute will create a thriving, high-energy, UK-wide network of interdisciplinary research expertise.

This will disrupt traditional science and transcend disciplines, enable knowledge and predictions from large multi-dimensional datasets and harness cutting-edge technologies to catalyse scientific discovery at scale, accelerate solutions to UK and global health challenges, drive the creation of business opportunities and support informed policy making.

The ambition is to transform science, practice and health globally."

Director's vision – Andrew Morris 2017

In less than four years, HDR UK has developed advanced infrastructure and applied cutting-edge data science approaches to clinical, biological, genomic, and other multi-dimensional health data to address the most pressing health research challenges facing society, including:

- 1. Assembly of a UK-wide data infrastructure and services for health research** – including 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](#), the [Health Data Research Innovation Gateway](#) and has convened a network of Trusted Research Environments (TREs) across the UK. This has enabled discovery and safe research access to over 720 datasets held by 60 data custodians (out of an estimated 5,000 UK health datasets currently in research use²). For the first time ever, research is now possible using linked, UK-wide data on a population of more than 65 million people.
- 2. Accelerating the pace and scale of health and biomedical data science** – 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. From April 2018 to March 2021, HDR UK researchers published

¹ The Consortium, convened by the Medical Research Council, includes the British Heart Foundation.. Chief Scientists Office (Scotland), Department of Health and Social Care (National Institute of Health Research), Economic and Social Research Council, Engineering and Physical Sciences Research Council, Health and Care Research Wales, Public Health Agency (Northern Ireland), and Wellcome

² UK health data assets include a broad range of data types and custodians across health and care system, academia, industry and medical research charities. It is also possible to create derived datasets, so the permutations are considerable. As a benchmark, there are >8,000 studies on UK data archive <https://www.data-archive.ac.uk/>; among these, the search 'health' returns nearly 3,000. The HRA website also identifies around 350 research databases that have received favourable opinions. In addition to these research products, data are collected as part of routine care within electronic health record systems (e.g., within ~300 NHS Hospital Trusts and 10,000 general practices across UK). It seems plausible that there could be >5,000 datasets. However, only a small proportion of these data are either collated as national datasets or accessed for research purposes.

>1,300 papers³ with >26,000 citations and have delivered research that has directly informed policy decisions and clinical practice during the COVID-19 pandemic.

3. **Working in partnership as One Institute** – HDR UK has created a multi-disciplinary, geographically distributed, UK-wide Institute of over 1500 researchers⁴ across >39 organisations⁵. 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. Patients, public and practitioners have been embedded at the centre of all work to ensure transparency and earn trust in Institute programmes with >43,000⁶ interactions through public events across HDR UK. The impact and return on investment from the £52.7 million Core Award has been significantly increased with £108.5 million of additional grant income obtained by HDR UK, and an additional £89.6 million in-kind contributions from partner organisations⁷.



³ This number is based on the ResearchFish report received 25 March 2021. This includes outputs by HDR UK National Priority members and HDR UK fellows published April 2018 (HDR UK establishment) -11 March 2021 and attributed by submitters as HDR UK Institute outputs. The snapshot of number of citations taken from Dimension AI on 28 April 2021. ResearchFish is an external online reporting system used by funders to collect information from researchers on the outcomes and the impact of their work. It is currently used by UKRI and many other public and charitable research funding agencies. ResearchFish data submission and collection take place only once per year – as determined by UKRI. HDR UK research outputs data to March 2022 will be available at the time of the Quinquennial Review in-person meetings. Dimensions is the world's largest linked research information dataset.

⁴ Institute members are defined in the HDR UK Attribution Policy. As a distributed institute that is not all 'under one roof' the HDR UK community is defined as people who are in receipt of funds from HDR UK (e.g. salary, research costs), or involved in an HDR UK project (including where salary is provided as an in-kind contribution). All outputs require HDR UK to be acknowledged as an author affiliation and/or funding support.

⁵ Includes research organisations that are members of HDR UK Substantive Sites, who are signatories to the HDR UK Institute Agreement. This does not represent all delivery partners within Substantive Sites and does not include partners involved in HDR UK activities beyond the Substantive Sites e.g. delivery partners in non-core funded infrastructure programme, host organisations to HDR UK MSc students, fellows.

⁶ Includes number of public attendees and participants engaged and involved in work across the National Priorities, Hubs, and Institute Office. Data is collected through monthly KPIs and represents the total number of people interactions through workshops, webinars and deliberative engagement activities since April 2020.

⁷ Further details on HDR UK's achievements and impacts are provided in Section 1.3 and Section 2.

Despite this progress, the transformative potential of health data research is far from being realised in full. Only a fraction of NHS, biomedical and health-relevant data is accessible to inform research. Data is of variable quality. Many datasets are still held, unconnected, in individual Institutions and/or on data platforms that lack the computing infrastructure required for advanced analysis. There are major research and technological skills shortages. Public trust and confidence in the use of health data for research remains vulnerable.

To fully understand the determinants of health, it is critical that the full range of NHS-generated, ‘real-world’ data can be accessed population-wide and linked with other diverse data. This includes research-generated multi-omics, imaging and pathology data, as well as environmental, social, administrative and citizen-generated remote monitoring data drawn from multiple sources across institutions and sectors and accessible as part of a trustworthy health data research ecosystem. This data ecosystem, if successfully designed, convened and coordinated, will act as the foundation for a new era of health data science to advance the discovery and practice of health, medicine and care.

HDR UK is on track to find a better way to create value and impact from distributed data assets by developing reliable, safe, and secure approaches to the assembly of a trustworthy data ecosystem. But there is much more to do.

In its proposal for the second five years (2023-2028), HDR UK sets out an ambitious and innovative programme that will massively increase the scale, quality, speed and impact of insights derived from health data research in the UK and internationally. By 2028, the Institute will be well positioned to move forward to the next stage of its >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’***, and to be at the centre of a global, collaborative network of health data science.

1.2. The origins of HDR UK and its structure

1.2.1. Origins of HDR UK

The UK has some of the richest healthcare and research data assets worldwide. While some major research data assets – such as [UK Biobank](#) and the [UK Data Service](#) – are internationally recognised, well organised, readily accessible and widely used, only a small proportion of health research data and data from the NHS is accessible for innovation and research to benefit the public’s health.

Over the last ten years, a wide range of stakeholders have become increasingly impatient to see value successfully extracted from the lifelong NHS health records of the UK’s population. This impatience has been accompanied by a growing concern that the UK could lose its global competitive advantage, its ability to perform ground-breaking research and capacity to attract significant inward investment all of which underpin better treatments and more efficient health and social care.

The UK is also well positioned to make a distinctive contribution to global health data science because of its strong research base. The UK’s science and research impacts are amongst the best in the world, and its universities feature prominently in global rankings for teaching and research. Institutes such as [The Alan Turing Institute](#), [Wellcome Sanger Institute](#), [European Bioinformatics Institute \(EMBL-EBI\)](#) and [UK Dementia Research Institute](#), and initiatives such as [Genomics England](#), [Our Future Health](#), [OpenSAFELY](#), and the capabilities of the [Office for National Statistics \(ONS\)](#), [National Institute for Health Research \(NIHR\)](#) Biomedical Research Centres, [NHSX](#) and [NHS Digital](#) collectively offer strong foundations for delivering research insight and impact from large scale data.

However, these strong foundations also bring significant complexity. There are hundreds of stakeholders and health-relevant data assets across a complex geographical, institutional, and sectoral landscape, including 164 UK universities, more than 1,200 NHS organisations and numerous public and private sector organisations. There is different legislation and structures across the four health services in England, Northern Ireland, Scotland and Wales. Data is of variable quality and is often difficult and slow to access for research.



In 2017, the core funders came together to make an initial £52.7 million, five-year investment, to establish a national institute for biomedical and health data science that would bring coherence to this complex landscape. The vision was for the Institute to provide national leadership, influencing the research community as well as UK scientific and health policy. The funders' objectives were to:

- Create a multi-disciplinary, geographically distributed and UK-wide research institute
- Build an integrated network to provide core capability
- Develop aligned partnership sites and other strategic investments
- Develop cutting-edge analytical tools and methodologies, required to integrate complex and diverse health-relevant data at an unprecedented depth and scale
- Develop a workforce of skilled data scientists through capacity building and career development
- Advise on, and collaborate with, complementary initiatives in data science led by the funders.

1.2.2. HDR UK's first five years

HDR UK's structure and strategy were designed to address the funders' objectives. As a new type of distributed organisation, the Institute recognised the need to be flexible in a fast-moving and complex environment and to adapt its governance and operating model to meet emerging needs.

At its inception, HDR UK was modelled on a multiple 'MRC University Unit' model. Following open competition, six Substantive Sites were established in Cambridge, London, Midlands, Oxford, Scotland, and Wales & Northern Ireland. The sites brought together 23 universities, research institutes and NHS partners and were governed by a single HDR UK Institute agreement. Each was led by a Research Director (John Danesh, Harry Hemingway, Simon Ball, Martin Landray, Cathie Sudlow, Ronan Lyons) supported by a team of Associate Directors and with overall leadership and coordination provided by Andrew Morris and Rhoswyn Walker (**Appendix 3: Leadership summary; Appendix 19: Leadership CVs**).

HDR UK's initial approach built a strong, UK-wide inter-disciplinary community. But important changes were necessary for the Institute to effectively deliver on its mission. Essential stages in the evolution included:

1. Pivoting from six geographical sites with multiple independent projects into [six national research priorities each led by a Research Director\(s\)](#)⁸, creating a **united 'One Institute' with an integrated national scientific and community offering**. These priorities were identified as opportunities to use health data in all its forms, including NHS patient data, genomics, biomedical and wearable data to ask important research questions that only a national institute could address (2019)
2. **Establishing Infrastructure and Services as a core HDR UK activity**, including: (i) the UK Health Data Research Alliance, convening data custodians across the UK; (ii) the Health Data Research Innovation Gateway (the Gateway), enabling discovery and access to data for research and innovation; and (iii) Health Data Research Hubs for expert data curation and data access services (2019)⁹
3. Holding an open competition to bring two more regions – North of England and South-West of England – into the Institute, to strengthen the Better Care National Research Priority and increase geographic coverage to help build an **inclusive and representative UK-wide scientific community** (2020)
4. Developing a **four nations network of national health data TREs** (secure platforms with the analytical and statistical tools to support researchers in doing their work) in partnership with [Administrative Data Research UK](#) (ADR UK) / ONS and the British Heart Foundation (BHF) Data Science Centre (DSC) / NHS Digital (2021)¹⁰
5. Bringing the Gateway team in-house to develop the **Gateway as a core infrastructure service, creating a dedicated computational and data engineering team** and increase operational efficiency and value for money (2021)
6. Forming **HDR Global**, building upon a successful initiative to establish the [International COVID-19 Data Alliance \(ICODA\)](#) to enable and support health data research, teams and networks in the Global South (2021, co-funded by the [COVID-19 Therapeutics Accelerator](#), Bill & Melinda [Gates Foundation](#) and [Minderoo Foundation](#)).

In 2021, HDR UK was awarded charitable status by the Charities Commission (charity number: 1194431), recognising its contribution to the public good.

Collectively these developments add long-term value for the wider health data research ecosystem and

⁸ National Research Priorities are (i) Applied Analytics (Dave Robertson and Chris Holmes); (ii) Better Care (Jonathan Sterne, Munir Pirmohamed, Simon Ball); (iii) Clinical Trials (Martin Landray); (iv) Improving Public Health (Ronan Lyons); (v) Phenomics (Cathie Sudlow and Harry Hemingway); and (vi) Understanding the Causes of Disease (John Danesh)

⁹ Funded by core funds and the Industrial Strategy Challenge Fund (ISCF) - Digital Innovation Hub (DIH)-programme; and the British Heart Foundation Data Science Centre to enable UK-wide health data science across a broad spectrum of cardiovascular and related domains.

¹⁰ Co-funded by the Core Award, UK Research and Innovation (UKRI) National Core Studies and British Heart Foundation Data Science Centre.

provide strong foundations to enable the Institute's plans for the second five years. The Institute's rapid evolution is in the context of a continually changing ecosystem, including the recent merger of NHS England & Improvement with NHSX and NHS Digital and the impact of the COVID-19 pandemic. HDR UK's agility, robust governance and strong leadership has allowed the Institute to thrive during this rapid period of change. However, the sequential integration of new grant funding has inevitably led to operational complexity. A key part of what the Institute sets out to achieve in the next five years is an integrated, coherent, and scalable programme of delivery.

1.2.3. Leadership, governance and delivery model

HDR UK aims to be highly innovative and a global exemplar of team science by creating an 'Institute without walls'. It is committed to being accessible, highly networked, and open to a range of collaborative models with NHS, public and a range of small and large private sector partners. The **Institute's values** (box below) create an inclusive and open community that weaves together talent across diverse professions, sectors, career stages and geographies. This approach has proved highly successful, building capacity and capabilities to showcase the power of collaboration, open science and data sharing.

HDR UK's values (**TORCH**) act as the foundation of the Institute's strategy and approach:

Transparency

Optimism

Respect

Courage

Humility



HDR UK is independent of the NHS, industry, government, and any single academic institution and is the only UK Institute that includes England, Northern Ireland, Scotland, and Wales. Maintaining an ‘institutionally-agnostic’ approach is invaluable, allowing HDR UK to convene, encourage standardisation, and serve a range of partners while capitalising on the diverse expertise required to generate new research insights from complex and distributed data assets.

HDR UK’s first review by its funders, the [Establishment Review](#), took place in September 2020. It assessed the progress made on setting up the Institute, including its operations, governance and strategic direction. The resultant strong positive endorsement by the independent review panel, chaired by Professor Dame Wendy Hall, is a testament to what has been achieved since the Institute was formed in 2018:

“HDR UK has made impressive progress since inception. The Panel congratulated HDR UK for the achievement of bringing together a highly committed and collaborative UK community with the ambition of improving health of the UK public by realising the public benefits from research using biomedical and health data at scale.”

Establishment Review, 2020

HDR UK operates a distributed model – with a coordinating team of directly employed and seconded staff based at the Institute Office at Wellcome in London, plus UK-wide partner organisations delivering research, training, and knowledge transfer activities. Coordinating functions provided by the Institute Office include scientific leadership and executive, finance, operations, communications, legal and major partnerships, and investment support. The majority of HDR UK’s core budget is distributed to partner research organisations, in particular to deliver the Institute’s National Research Priorities and Health Data Research Hubs (**Figure 1**).



The Institute has also successfully established several partnership programmes across the UK and globally. HDR UK's success in winning additional grant funding has resulted in growth, and diversification of skills and expertise within the Institute's directly employed teams to lead, manage and deliver new programmes, including the Gateway and Alliance (Figure 1).

Figure 1: HDR UK first five years Institute programmes



HDR UK's leadership and governance structure are set out in detail in **Section B.4.1** of the [Establishment Review](#). A summary of the Institute's leadership is provided in **Appendix 3: Leadership summary**. HDR UK's governance structure is provided in **Appendix 4: HDR UK's Current and Future Governance Model**.

The Institute is governed by an independent [Board of Trustees](#) with supporting [Advisory Groups](#). The Institute's [Executive Committee](#) (ExCo) is responsible for setting and monitoring HDR UK's strategic direction. The Institute's [Delivery Group](#), comprising a UK-wide leadership team (Research Directors, Hub and Data Science Centre Directors, and Programme Directors), enables HDR UK to function as a unified Institute. It provides clear accountability and single 'line of sight' of all programmes to the Board and core funders. The structure also provides a clear route for partners to work with HDR UK, with the Institute identifying opportunities for further engagement through regular updates on all activities at the UK Health Data Research Alliance Board and Delivery Group meetings.

HDR UK has a very active and engaged [Public Advisory Board \(PAB\)](#) chaired by Angela Coulter. The PAB plays a key role in earning trust and transparency through involvement in all Institute activities and ensuring that it is ultimately accountable to patients and the public.

The Institute has a distinguished [International Advisory Board \(IAB\)](#), chaired by Dr Amy Abernethy, formerly Chief Medical Officer at [Flatiron Health](#), Deputy Commissioner at the [US Food and Drug Administration](#) (FDA) and now President of [Verily Life Science's](#) Clinical Research Platforms. The IAB has met three times since HDR UK's inception, and the members provide invaluable advice on scientific and infrastructure strategy and HDR UK's role in training and mentorship. Members also highlight the distinctive strengths of the Institute which inform HDR UK's future strategic priorities:

“Serve as a template for the world – HDR UK has an opportunity to serve as an exemplar in infrastructure, standards, strategy and more that other sectors and geographies could learn from.”

Amy Abernethy, IAB Chair, Verily (July 2021, IAB)

1.3. Achievements, impact and added value

The Institute has demonstrated the added value of integrating UK-wide infrastructure and health data science capabilities within a single organisation which can draw on and convene diverse skills and expertise. HDR UK's mission is to enable data-driven discoveries across a broad range of health challenges. The Institute's ability to rapidly bring together UK-wide and global infrastructure, research and collaborative partnerships has been recognised.

After less than two years of operation, the COVID-19 pandemic created a clear focus of activity that has acted as an accelerator of the Institute's impact. The case study summarising HDR UK's contribution to the COVID-19 pandemic (**Section 2.1.2 Delivering impact through the COVID-19 pandemic**) demonstrates powerfully what can be achieved through the linkage and analysis of diverse and large-scale UK-wide health-related data to influence health policy and clinical practice in near real time.

The opportunity is to use this model, exemplified through the COVID-19 pandemic, to tackle other diseases and societal challenges that cause significant morbidity and mortality.

"Thank you, without HDR UK's work a lot of what we have achieved this year during COVID would not have been possible"

Sir Chris Whitty, Chief Medical Advisor UK Government (March 2021 to the COVID-19 Taskforce)

"What you are doing on data linkage, governance and coverage across the four nations is crucial and impressive. Your work is absolutely making a difference and it is being noticed. It's not going to get quieter though, it's only going to get busier."

Sir Patrick Vallance, Chief Scientific Advisor UK Government (February 2021 to the Data & Connectivity Delivery Group)

Highlights of the Institute's achievements and impacts include (Section 2 for full details):

1.3.1. Impact 1: Assembly of a UK-wide and international data infrastructure for health research

- Over 60 leading healthcare and research organisations have joined the **UK Health Data Research Alliance** (established by HDR UK 2019) (**Appendix 16: UK Health Data Research Alliance member organisations**), uniting to establish best practice for the ethical, trustworthy, and streamlined use of UK health data for research at scale. The Alliance has worked in partnership across five priority areas, to identify common challenges and barriers to progress, develop and promote shared solutions to overcome these barriers and convene the community to share progress (**Section 2.2.3**). Outputs have included establishing a joint consensus statement on TREs in partnership with NHSX, towards harmonising the landscape of infrastructure resources
- The Gateway has delivered a single, efficient entry point for researchers and innovators to discover and request access to large-scale UK health-related data (established 2020) **across multiple data platforms** and data custodians. At present **720 datasets have been made discoverable** (January 2022) (whereas prior to the establishment of HDR UK, the UK had just 18 datasets discoverable on Health Data Finder, which provided the starting point for the Gateway). The UK's first harmonised **data access request** process across multiple data custodians has been integrated into the Gateway ([Bailey et al 2022](#))
- HDR UK has led the development of nine Health Data Research Hubs as UK centres of excellence providing expertise, tools, knowledge and services to maximise the insights and innovations developed from health data. The Hubs act as a focus for interface with industry and have curated over 80 high value datasets, securing over **450 contracts in their first 24 months**, 26% of which are with industry
- HDR UK leads the [BHF Data Science Centre](#) (Director Cathie Sudlow) that has created for the first time a centre of interdisciplinary expertise that provides national research capability in cardiovascular data science
- HDR UK has engaged a range of industry partners in the design and delivery of the research data infrastructure and services, including enabling new data resources to be made available for clinical trials (in partnership with AstraZeneca and NHS Test and Trace) and supporting a range of small to medium-sized enterprises (SMEs) in collaboration with KQ labs and the Medicines Discovery Catapult (**Section 2.1.7 Knowledge transfer and exchange**)
- HDR UK Institute members have supported the UK and international response to the COVID-19 pandemic through the [UKRI Government Office of Science National Core Study on Data and](#)

[Connectivity](#) - enabling use of over 90 COVID-19 relevant datasets in five UK TREs¹¹, supporting >1,500 studies which have directly impacted policy and healthcare practice across the world (**Section 2.1.2 Delivering impact through the COVID-19 pandemic**)

- Via the HDR UK BHF Data Science Centre's [CVD-COVID-UK/COVID-IMPACT Consortium](#), linked health data from more than 54 million people in England is available for the first time for researchers through NHS Digital's TRE ([Wood et al 2021](#)), and extended this to link and enable access to 62 datasets across three national TREs¹². The Consortium has >250 members across >40 institutions with >70 analysts working on >25 projects, including supporting research to better [understand vascular risks of COVID vaccines in 46 million adults in England](#), which directly informed UK government policy
- HDR UK led the establishment of the International COVID-19 Data Alliance, to create a trustworthy international health data infrastructure that supports 12 COVID-19 research programmes across 19 countries.

1.3.2. Impact 2: Accelerated the pace and scale of health and biomedical data science

- From April 2018 to March 2021, HDR UK researchers published > 1,300 papers¹³ with >26,000 citations
- HDR UK has delivered science that has broken down barriers between disciplines. Institute outputs have directly informed the UK government policy decisions and clinical decision making during the COVID-19 pandemic; continue to enable research breakthroughs and provide exemplars of the new ways of working with large-scale health data which are required to deliver insights directly benefiting patients and the public:
- Over 20 publications demonstrating the effectiveness and safety of COVID-19 vaccines through the EAVEII studies led by HDR UK Hub Director Aziz Sheikh ([Simpson et al 2021](#); [Sheikh et al 2021a](#); [Sheikh et al 2021b](#)¹⁴)
- The RECOVERY trial, co-led by HDR UK Research Director Martin Landray, provided the world's first COVID-19 treatments proven to save lives and was pivotal in the selection of effective treatments for COVID-19 ([The Recovery Collaborative Group, 2021](#))
- Providing the Welsh Government with up-to-date insights and accurate spatial data to inform regional lockdown restrictions, using the research-ready Welsh national e-cohort linked health data asset, representing 3.1 million people in Wales and securely stored in Secure Anonymised Information Linkage Databank ([SAIL Databank](#)), led by HDR UK Research Director Ronan Lyons ([Lyons et al 2020](#))
- Mobilising 15 electronic health record (EHR)-enabled population cohorts (comprising >800,000 participants from across the UK and internationally) with extensive genomic and other molecular data to support multiple high-impact discoveries, including the identification and prioritisation of novel drug targets and clinically relevant determinants of human disease ([Gaziano et al 2021](#); [MacTel Consortium 2021](#); [Zheng et al 2020](#))

¹¹ TREs in England ([NHS Digital TRE service for England](#)), Wales ([Secure Anonymised Information Linkage \(SAIL\) databank](#)) and Scotland ([Scottish National Data Safe Haven](#)) Northern Ireland ([Northern Ireland Honest Broker Service](#)), [Office of National Statistics](#)

¹² TREs in England ([NHS Digital TRE service for England](#)), Wales ([Secure Anonymised Information Linkage \(SAIL\) databank](#)) and Scotland ([Scottish National Data Safe Haven](#))

¹³ This value is based on the ResearchFish report received 25 March 2021. This includes outputs by HDR UK National Priority members and HDR UK fellows funded by UKRI published April 2018 (HDR UK establishment) -11 March 2021 and attributed by submitters as HDR UK Institute outputs. The snapshot of number of citations taken 28 April 2021.

¹⁴ Supported by HDR UK core funds, and UKRI Data and Connectivity National Core Studies

- Providing tools to make data usable. Spiros Denaxas (HDR UK London) and colleagues have developed the [Phenotype Library](#), the largest national standards-driven library of citable phenotyping algorithms, metadata and tools for defining human disease in EHRs research ([Denaxas et al 2020](#))
- Improving scientific reproducibility and transparency associated with the rapidly growing use of polygenic scores to enhance the prediction and understanding of diseases and traits. Mike Inouye (HDR UK Cambridge) and colleagues, worked with EMBL-EBI and partner institutions nationally and internationally to launch the open access [Polygenic Score Catalog](#) in 2019. This is now a widely used, international database ([Lambert et al 2021](#)).

1.3.3. Impact 3: Working in partnership as ‘One Institute’

- HDR UK has created a UK-wide, interdisciplinary research community. The Institute has shared its research insights at its own conferences and events that have attracted >13,500 people from 68 countries. The wider HDR UK community has grown to >500 partner organisations internationally
- The Institute has created and enabled:
 - A cohort of 50 cross-UK postdoctoral fellows, with shared learning opportunities with the NIHR Academy
 - A UK-wide Wellcome Trust funded PhD programme in partnership with the Turing Institute
 - Seven MSc programmes with partner universities across the four nations
 - HDR UK Futures, established in 2021 is an online learning platform which has already provided learning and training opportunities for >6,000 people.
- The Institute’s national leadership of the [Black Internship Programme](#) for health data science supported 48 interns embedded across 23 partner organisations in its first year (2021), with 60 partners agreeing to host interns in 2022. It has championed diversity and inclusion in health data science (**Figure 2**)
- HDR UK has shown leadership across the sector in demonstrating transparency and trustworthiness to patients and by embedding public and patient involvement 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
- On the global stage, HDR UK has contributed to developing the health data science policies of the World Health Organization (WHO), the US National Institutes of Health (NIH), the G7 and G20 countries, Singapore, France and countries in Africa, South America and Asia. HDR Global has collaborated with The Global Health Network (TGHN) to provide leadership in developing federated partnerships with low and middle-income country (LMIC) partners and enabled researchers in the global south to harness data science to address locally identified health challenges.

Figure 2: HDR UK Black Internship Programme cohort 2021



1.3.4. Additional investment secured

HDR UK's core-funded activities have been enhanced with significant, additional directly funded programmes of £108.5 million to support delivery of the Institute's mission. These programmes have been essential to the creation of the infrastructure and services that underpin HDR UK's second five years, including the establishment of the Gateway and Alliance.

In addition, HDR UK has been supported by a further £89.6 million of in-kind contributions to Institute programmes from partner organisations. This in-kind contribution takes the form of staff time, facilities and resources which directly enable HDR UK led work, but are not funded by the Core Award or the additional funded programmes.

1.4. Plans for the second five years

1.4.1. Lessons learned

Alongside the challenges of delivering health and care services during the worst pandemic for a century, other pressing societal challenges include increasing numbers of people living with multiple, long-term conditions, widening health inequalities and the vulnerability of populations to climate change.

During the COVID-19 pandemic, the world has engaged with data and information more intensively than ever before. Large-scale integration and analysis of health-relevant data across organisational and national boundaries is vital to deliver insights which can help address not only pandemic recovery but also the social and economic security of the global population.

Health data science is at a significant inflection point, with the opportunity to improve health and care through research across the world. HDR UK has demonstrated that it can serve as a facilitator to connect, share excellence, learn from what has worked elsewhere and enable a convergence towards best practice. The Institute's coordinated approach has delivered impact on a scale not witnessed before and is aligned with the aims of the UK's [National Data Strategy](#), [National AI Strategy](#) and [Life Sciences Vision](#). But there is more to be done. Fragmentation, lack of inter-operability, as well as legal, regulatory, cultural and philosophical barriers make data access and data sharing slow.

Over 50 partner organisations are interviewed each year by the HDR UK leadership to help define the Institute's distinctive contribution to health data research and inform its strategy and priorities. The insight from these interviews has helped to identify seven unmet needs that HDR UK is well positioned to tackle:

1. Diversity of data

Insight and learning: Ground-breaking research insights require diverse, linked data from multiple sectors. This includes molecular, behavioural, environmental and social data in many forms, from codified EHRs to pathology, imaging, free text and wearables. The Institute's focus to date has been on enabling access to and curating structured national NHS data. This is of huge value and remains a work in progress but is only one component of a rich data ecosystem. We are just scratching the surface. HDR UK is well placed to expand the diversity of data available for research in collaboration with ecosystem partners from NHS, academia, industry, and government.

In the second five years, the Institute will continue to make strides towards enabling data for research on a scale not witnessed before. It will not only enable linkage of priority NHS data to diverse research datasets but also, working with ONS, ADR UK, universities, industry and others, extend beyond NHS data assets to broader health-related data. This will enable society to address major challenges including resilience for future health shocks, inequalities across the life course and synergistic approaches to disease prevention (**Section 3.1.1**).

2. Bringing alignment to fragmented data systems

*Insight and learning: The health data ecosystem is very complex. HDR UK's convening power and ability to generate coherence in a complex landscape is highly valued as demonstrated by the 59 letters of support received from a wide range of partners (**Appendix 18**). Representing and engaging partners across the whole of UK gives the Institute an important and distinctive opportunity to enable ecosystem alignment.*

Good progress has been made establishing national TREs in England, Northern Ireland, Scotland and Wales but most data remain scattered, fragmented, and documented in a wide variety of formats. HDR UK will work with the NHS, industry as innovators and technology providers, the Turing Institute, ONS, and other partners to assemble and deploy enabling technologies and other infrastructure across the data ecosystem. Data custodians will be encouraged to deposit datasets to an appropriate secure data repository to increase reuse and to ensure interoperability between TREs and between the TRE network and the Gateway. Having already made progress towards increasing the findability and accessibility of health data, in HDR UK's second five years the Institute will build on this and work with the above partners and wider stakeholders to develop a blueprint for an interoperable health data research infrastructure which is open and accessible to all (initially delivered by Research Infrastructure and Service programmes, **Section 3**).

3. Improving data quality and utility

Insight and learning: Health data often requires expertise to engineer and 'clean' the data before use.

HDR UK will **create a team of data engineering expertise**, distributed across the UK, committed to open tools and technologies development that produce research-ready data for advanced data analytics, including AI. HDR UK will support the development of a critical mass of data curation and data wrangling capabilities in TREs (initially delivered by Research Driver Programmes and Infrastructure and Services and supported by the Capacity Building Infrastructure and Service programme, **Section 3**).

4. Demonstrating trustworthiness

Insight and learning: Public trust is integral to successful, high impact delivery, but needs to be continually earned. Patients can be highly effective advocates for health data research and involving the public in every stage of data use results in better and more trustworthy research.

The UK needs transparent governance approaches that address the complex and dynamic landscape of confidentiality and data privacy underpinning health data research within and between institutions, and countries. HDR UK will provide **national leadership for Findable, Accessible, Interoperable, Reuseable (FAIR) data in health** ([Wilkinson et al 2016](#)).

Trust and Transparency will be a major focus. An Infrastructure and Service research and delivery programme will address the public involvement, governance, ethical and policy requirements that large-scale data sharing requires, whilst addressing legitimate concerns about data privacy. As an extension of this approach, HDR Global will collaborate with TGHN and regional partners to enable knowledge sharing and development of governance approaches that demonstrate trustworthiness on a global scale.

HDR UK will evaluate, in partnership with the Turing Institute and wider engineering and computational sciences community in academia and industry, how advances in privacy-enhancing technologies can be applied to the sharing and use of health data for research.

5. Integrative approaches to improve the research experience

Insight and learning: Assembling data infrastructures needs to follow the principle of user-centred design. "Build it and they will come" isn't enough. Data integration needs to be for a purpose. Any new data infrastructure should be developed with specific health research questions and user needs in mind.

Establishing focused research programmes to answer distinctive research questions is critical for efficient and effective delivery.

In the next five years the Institute will ensure tight integration of research, infrastructure and services by sponsoring six focused, Research Driver Programmes that will create and test new approaches using health data for research. The programmes were selected based on having UK-wide representation, global relevance, scientific novelty with potential for direct impact on public health and clinical practice, and capacity to contribute resources and expertise to test and develop HDR UK's Infrastructure and Services. The aim is to leave a legacy of improved health data research experience.

6. Meeting the global opportunity

Insight and learning: challenges are global. The UK needs to be outward looking and willing both to partner with and learn from similar health data science programmes globally.

HDR UK's second five years will strengthen the UK's international contribution to health data science through the quality and scale of its health data research insights. The Institute will achieve this initially by developing the concept of HDR Global, working with international partners to learn from others and to support collaboration and knowledge sharing on a global scale.

7. Health data research culture

Insight and learning: HDR UK as a new type of Institute is well positioned to champion team science, develop new approaches to research culture and provide representative and inclusive leadership across the health data science community.

The Institute will champion a research culture that celebrates creativity, inclusivity and the values of Transparency, Optimism, Respect, Courage and Humility. The Institute will continue to grow a diverse talent pool and embed best practice in research, building on the policies, values and approaches established in the first five years.

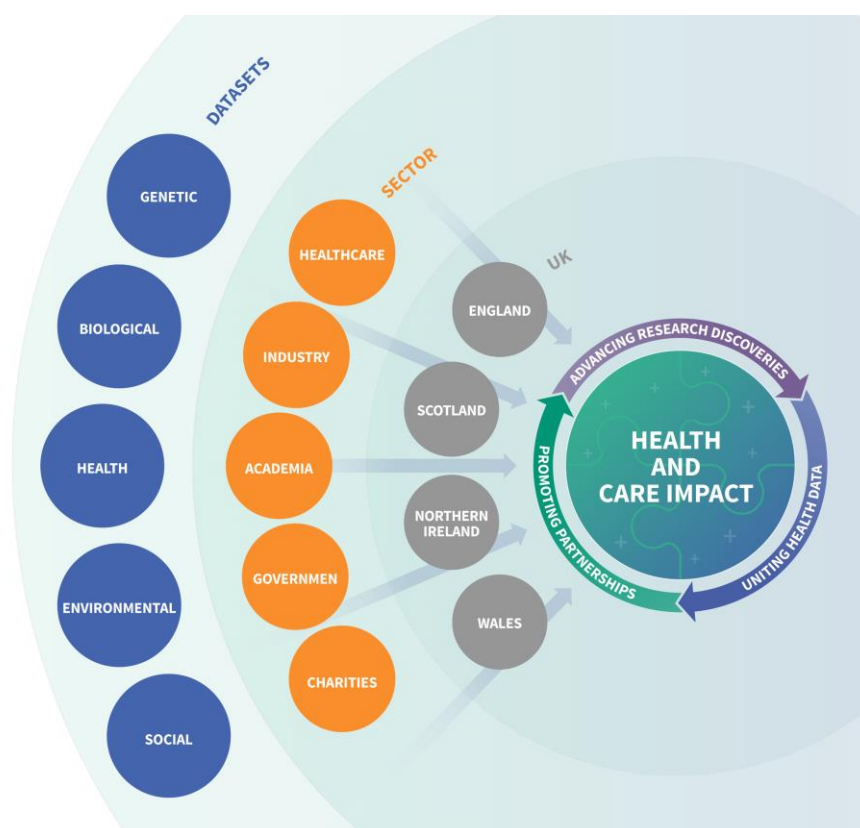
1.4.2. Strategy for the second five years (2023 - 2028)

HDR UK's long-term mission, to **unite the UK's data to make discoveries that improve people's lives**, remains unchanged as it moves forward into its second five years. 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'.

HDR UK will convene partners across different geographies and sectors to unite diverse health-related data (**Figure 3**). The breadth of support for the Institute's vision and strategy is clearly demonstrated by the 59 letters of support received from a wide range of NHS, industry, academic, charity and international partners as part of preparing this proposal (**Appendix 18**). The Institute will continue to assemble and deliver user-led health data research infrastructure and services. 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.

The Institute will narrow the gap between new research discoveries and direct impacts for patients and the public. Across pandemics and infectious diseases, non-communicable diseases - including cardiovascular disease, cancer and dementia, common and rare conditions - patients and the public will have safer, better and faster answers to causes, treatment and prevention of disease to deliver equality in health outcomes.

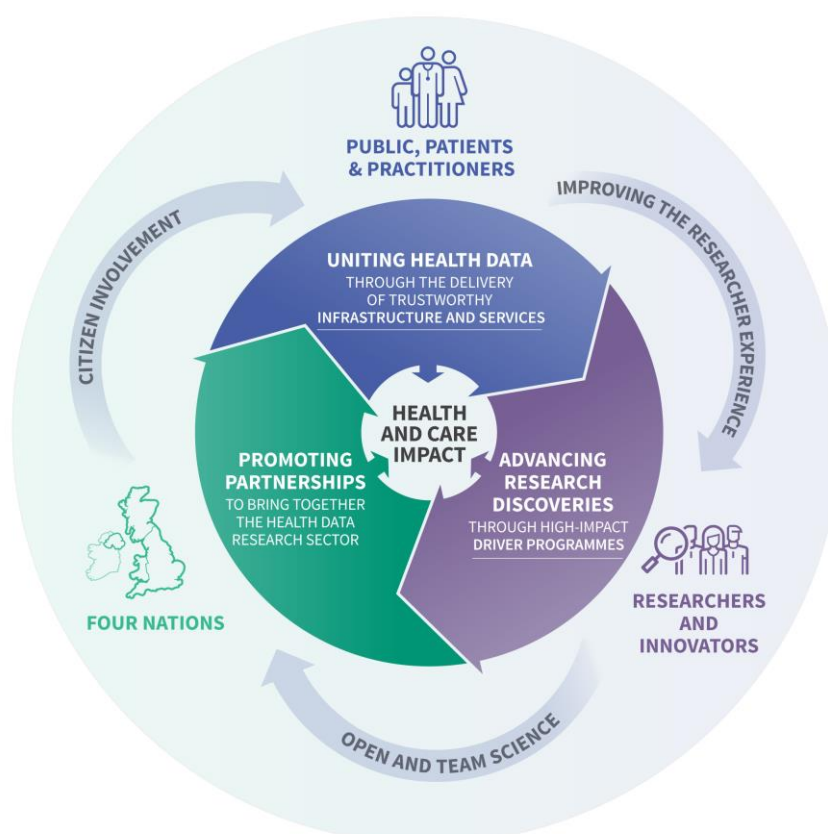
Figure 3: Harnessing the UK's diverse health data ecosystem



In the second five years, HDR UK will focus on **three integrated areas of activity** to deliver this ambition (Figure 4):

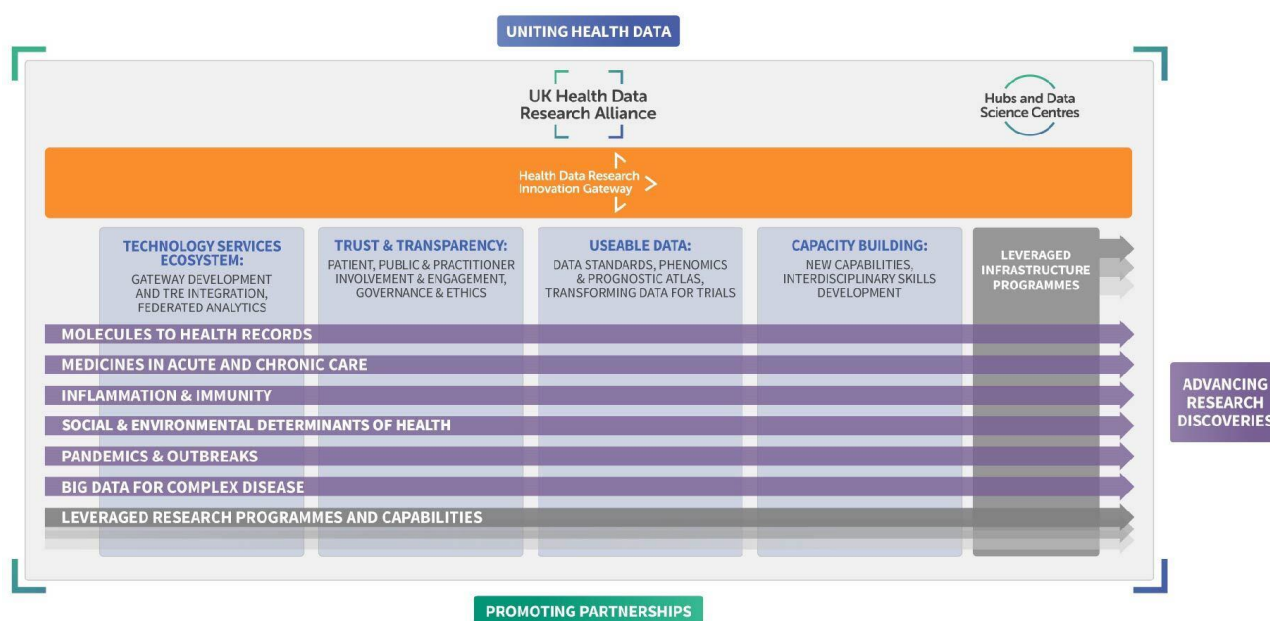
1. **Research Data Infrastructure and Services:** Providing the UK-wide and global co-ordination and leadership of health data infrastructure and services required to make health-relevant data FAIR. This will be built on the convening, collaborative and co-ordinating role of the **UK Health Data Research 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.
2. **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.
3. **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.

Figure 4: HDR UK's strategy for the second five years



The core-funded Infrastructure and Services will enable the core Research Driver Programmes, and in turn the Research Driver Programmes will inform and drive the core Infrastructure and Services, enabled by UK-wide and global partnerships (Figure 5).

Figure 5: Second five years Research Driver Programme, Research Data Infrastructure and Services and One Institute Partnerships interconnected strategy



These interconnected areas of research activity represent a refinement and evolution [of the first five-year strategy](#). They will increase coherence and the efficiency of UK health data research, moving from one-to-one relationships between data custodians and research teams to a scalable, many-to-many approach, where linked datasets are easily accessible to a wider number of researchers. The impact will be to considerably boost the number and originality of emerging health research insights.

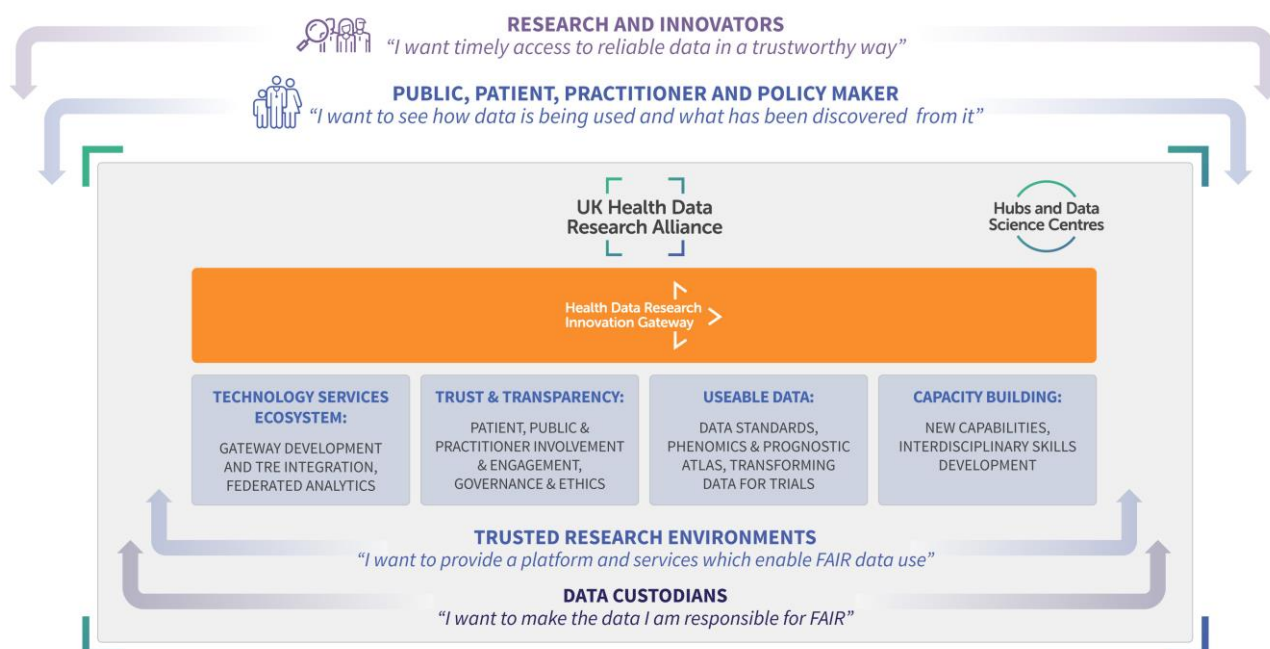
The programmes and services will be delivered by a leadership team from across the UK (**Appendix 3: Leadership summary**), which has been assembled based on excellence, equality and diversity and supported by the 'One Institute' approach to interdisciplinary team science. The ability to integrate new programmes and capabilities funded by non-core investments (leveraged activities) will also create new leadership opportunities, ensuring that the Institute remains able to adapt and respond quickly to changes in the external environment.

1.4.3. Research Data Infrastructure and Services

HDR UK will focus on the research, design and assembly of high-performance Research Data Infrastructure and Services to unite health-relevant data so that they are FAIR.

The Infrastructure and Services will be assembled on behalf of the members of the UK Health Data Research Alliance, and the wider research community. The Alliance is central to the delivery of HDR UK's aims. HDR UK proposes to strengthen the Alliance by increasing its membership, making it open to all custodians of health-relevant data, and by evolving its governance and oversight to reinforce its inclusive, cooperative structure (Figure 6).

Figure 6: Uniting health data through delivery of FAIR Research Data Infrastructure and Services



The Alliance (led by Ben Gordon, David Seymour, Kay Snowley and Paola Quattroni), will provide the convening, collaborative and coordinating role required to support four Infrastructure and Service pillars:

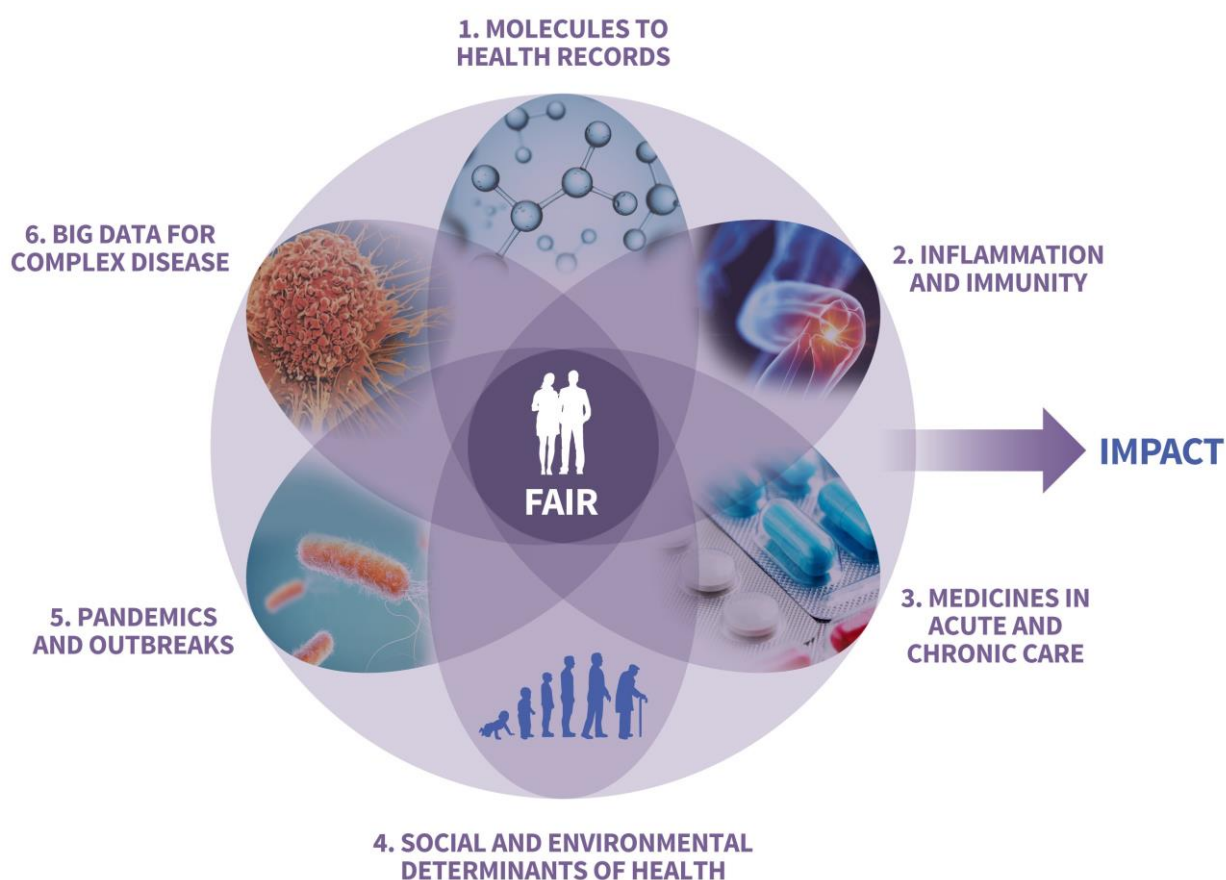
- **Pillar 1: Technology Services Ecosystem** (led by: Carole Goble, Charles Gibbons, Emily Jefferson, Helen Parkinson, Luke Readman, Mark Parsons, Michael Chapman, Pete Stokes, Phil Quinlan, Simon Thompson, Susheel Varma)
 - Assembly of an ecosystem of integrated data resources and services. This ecosystem will be the centrepiece of HDR UK and wider UK health landscape, universally compatible across TREs, trusted, interoperable (tools) and flexible (data).
- **Pillar 2: Trust and Transparency** (led by: Amanda White, Andy Boyd, Cassie Smith, Chris Monk, Sinduja Manohar)
 - Leadership and innovation in patient and public involvement and engagement that demonstrates trustworthiness
 - Leadership in defining principles and best practice of a UK-wide ethical and governance framework, supporting the harmonisation, adoption and automation of health-related data sharing.
- **Pillar 3: Useable Data** (led by: Ben Gordon, Emily Jefferson, Harry Hemingway, Marion Mafham, Matt Sydes, Monica Jones)
 - Tools and data engineering capability that enable seamless access to FAIR data, including Phenomics and Prognostic Atlas capabilities (dataset search, classification, and efficient metadata browsing tools described via open dataset catalogues, common data models and data dictionaries), and transforming data, and providing resources for clinical trials.
- **Pillar 4: Capacity Building** (led by: Christopher Yau, Sarah Cadman, Tim Frayling)
 - Develop a workforce of skilled data scientists, technologists and health data professionals through capacity building and career development with global reach.

The Infrastructure and Services will be delivered by HDR UK teams, bringing together a wide range of skills, including data scientists, ethicists and computational researchers working in partnership with academia, industry and the NHS across the UK.

1.4.4. Research Driver Programmes

The six high-impact health data research programmes will not only harness large-scale data to advance understanding of human health but will also strategically drive and advance the development of the FAIR Infrastructure and Services, collectively spanning discovery, prevention, early diagnosis, and treatment (Figure 7).

Figure 7: Research Driver Programmes: Advancing health research discoveries and leaving a legacy



The programmes have been selected based on being ambitious and disruptive, in anticipation of delivering ground-breaking scientific and infrastructure innovation. Each programme creates an alignment with UK-wide partnerships of academic organisations, NHS partners, industry, charities and policy makers.

Importantly the programmes have been designed to deliver impact and excite and inspire the public, academia, and industry. They will solve issues that are important to society and will provide a balanced and complementary range of data uses, beneficiaries, research outputs and infrastructure challenges to overcome. To ensure the programmes rapidly deliver impacts, they build on the most successful and impactful elements of the first five years' health research portfolio and focus on areas where UK scientific expertise is poised to enable future breakthroughs.

The initial Research Driver Programmes are:

- **Molecules to Health Records** (led by: John Danesh and Sarah Lewington)
Create new informatics infrastructures and data science methods that help achieve a deep integration of biology, biomedicine, and population health sciences. The ‘Molecules to Health Records’ approach will unlock the potential of major global investments, in the UK and worldwide, in large-scale molecular bioresources of populations and patients.
- **Medicines in Acute and Chronic Care** (led by: Elizabeth Sapey and Munir Pirmohamed)
Understand and transform the use of medicines for patients of all ages, especially those multiple long-term conditions and those facing health inequalities, by creating a real-time data foundation of all medicines for the UK.
- **Inflammation and Immunity** (led by: Aziz Sheikh and Jenni Quint)
Transform the UK’s capabilities to improve understanding of inflammatory mechanisms and health outcomes using respiratory and allergic disorders as exemplar clinical domains.
- **Social and Environmental Determinants of Health** (led by: Paul Elliott and Ruth Gilbert)
Provide national infrastructure and methods to link place-based information on social and environmental determinants of health to existing, longitudinal administrative data from health and related services, as well as cohort and survey data, at extremely high spatial resolution.
- **Pandemics and Outbreaks** (led by: Kenny Baillie and Sharon Peacock)
Develop sustainable Outbreak Data Analysis Systems within the UK, through an Anti-Microbial Resistance (AMR) exemplar, to demonstrate their effectiveness and to facilitate adoption worldwide.
- **Big Data for Complex Disease** (led by: Cathie Sudlow and Mark Lawler)
Deploy the increasingly extensive, diverse, health systems data now being brought together at national/country level across the UK to better predict diseases such as cancers and cardiovascular diseases (CVD), to stratify risk for better screening, early detection and diagnosis, and to enhance understanding of the interactions between different diseases.

Coherence and integration across the programmes will be ensured through UK-wide coordination (led by Andrew Morris and Rhoswyn Walker). Each programme takes advantage of the range of datasets available across the UK that are relevant to human health, across a range of modalities and dimensions exemplifying the power of data integration for a health research purpose.

Each programme must leave a legacy by enhancing existing data resources and enabling access to high value priority datasets to create novel, linked **data foundations which are FAIR**¹⁵. Priority UK-wide data assets including, near real-time hospital datasets, GP data, imaging, pathology and device datasets will be curated, some for the first time. Linkage of cross-sectoral datasets enabled through partnerships with ONS, ADR UK and the Geospatial Commission will be progressed (**Appendix 18: Letters of Support**). The programmes will tackle a range of methodological challenges and develop new ways of curating and analysing diverse large-scale data. They will ensure that Infrastructure and Services are co-designed to meet user needs and will test a replicable approach that will allow the Institute to address additional challenge areas, such as social care and mental health, by on-boarding further, externally funded, Driver Programmes.

¹⁵ Data foundations are data that is fit for purpose, recorded in standardised formats on modern, future-proof systems and held in a condition that means it is findable, accessible, interoperable and reuseable as defined in the UK National Data Strategy, 2020.

The programmes will create global collaboration opportunities, share insights and expertise for wider benefit and co-develop approaches which enable global impact.

1.4.5. One Institute Partnerships

HDR UK will build on its existing structure of a coordinating team, based at Wellcome in London, while distributing the majority of its resource to Driver Programme, Infrastructure and Service leads and eight Regional Networks across the UK. This allows the Institute to benefit from the health data science expertise and infrastructure across the UK and globally, working with regional NHS, academic and industrial ecosystems to deliver the Institute's mission in partnership.

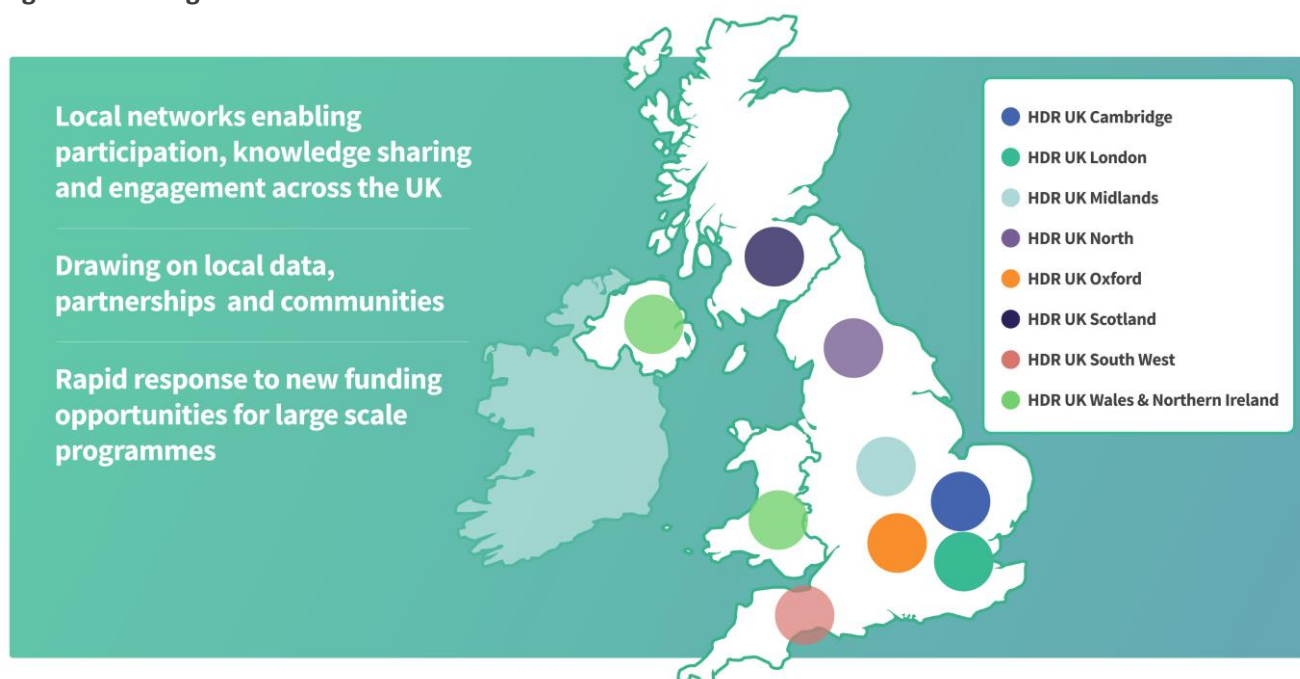
UK Regional Networks (led by: Alastair Denniston, Andy Clegg, Angela Wood, Cecilia Lindgren, Dave Robertson, David Ford, Dermot O'Reilly, Emily Jefferson, Eva Morris, Fiona Pearce, Harry Hemingway, John Danesh, Jonathan Sterne, Munir Pirmohamed, Rachel Denholm, Sinead Brophy, Sinead Langan)

HDR UK regional funding has been transformational in promoting joint working between partners, including engagement of hard-pressed NHS colleagues to create the partnerships necessary for health data research. These networks will be maintained and further developed in HDR UK's second five years.

The eight UK Regional Networks (**Figure 8**) will:

- Enhance regional partnership and collaboration with HDR UK that will enable participation in existing and future Driver Programmes and active participation in Infrastructure and Services to address regional and national priorities
- Work with regional partners including universities, Academic Health Science Networks (AHSNs) and NHS bodies to speed up adoption of health data science innovation into practice
- Act as a focus of investment leverage for future HDR UK scientific, training and infrastructure.

Figure 8: UK Regional Networks



The regional leads will join the Institute's Delivery Group, with a specific remit to review milestones, opportunities for collaboration and extension of Driver Programmes and Infrastructure and Services. The Regional Networks will inform best use of UK-wide local, regional and national data assets and engage researchers across the four nations to ensure that the benefits reach the whole UK population. The Networks will represent the health research data community in their region. They will build and maintain strong links with colleagues in key regional data custodian organisations and facilities, as well as NIHR Biomedical Research Centres, companies, charities and academic groups across the UK.

HDR Global (led by: Anne Wozencraft, Neil Postlethwaite, Trudie Lang)

HDR Global, is a global federated health data partnership supporting data governance, capacity building and system strengthening by connecting experts and excellence across and between projects, programmes, regions and organisations. It provides open access to tools, resources, technical and data analytical support to enable new analyses from existing data sets. Initially, HDR Global is engaging with three regional centres of excellence in Africa, South America and Asia to address local health changes and embed approaches that ensure findings from new data analysis are taken into practice.

It has adopted the Driver Programme 'showing-by-doing' approach that generates evidence through asking new questions of existing datasets in the world's most resource-limited settings. HDR Global is closely partnering with TGHN to deliver this ambitious data science programme.

The core funds will support a small leadership team to ensure integration with the Institute's wider expertise and activities. The strategy is that HDR Global will maintain its track record of obtaining major third-party investment to deliver programmatic activities.

Institute Office (led by: Amanda Borton/Alison Hopkinson (interim), Andrew Morris, Amanda White, David Seymour, Victoria Platt)

The Institute Office comprises a core team of permanent staff (employees and secondees) covering a range of skills from strategic development and oversight, communications, operations, finance, scientific strategy, technology strategy and programme delivery. Broadly, they provide four key functions across the Institute:

- Institute leadership including providing secretariat to the Board of Trustees and seeking medium-to long-term development opportunities
- Cross-Institute operations, finance and communications including leading on HDR UK's commitment to achieve net zero and diversity and inclusion programmes
- Joining up strategy delivery mechanisms across the Institute to ensure alignment and operational efficiency
- Leading major UK and international strategic partnerships.

1.5. Future aims and objectives

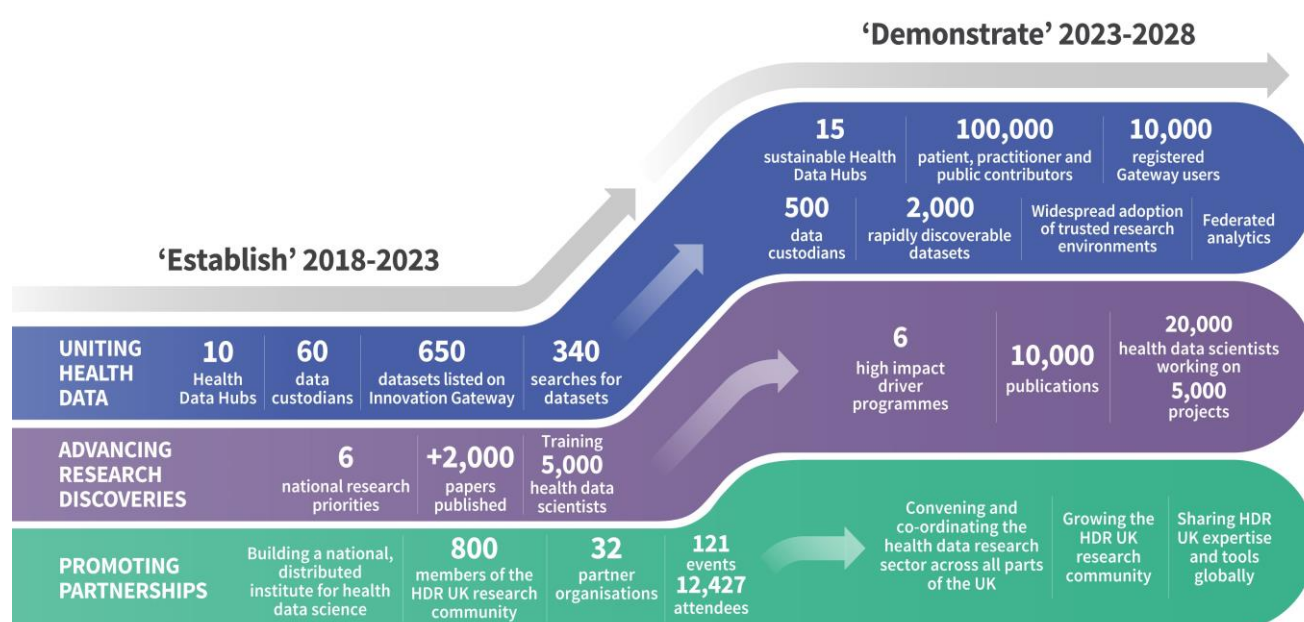
1.5.1. Metrics for success

Through this strategy, HDR UK will continue to deliver step changes in the speed and scale of benefits realised by the public from health data (**Figure 9**). This will be achieved by providing better, more robust and increasingly insightful research through a distributed, FAIR data infrastructure. The interconnected programmes will: support linked data assets relevant to the entire UK population and enable more than 10,000 publications with public and patient impact; embed HDR UK expertise in over 5,000 projects involving 20,000 health data scientists; make more than 2,000 well curated datasets discoverable and

accessible via the Gateway and attract and train 20,000 people to advance their health data science careers¹⁶.

HDR UK will continue to enable research that directly benefits patients and the public, informing clinical decision making and influencing health policy. This will be demonstrated through a range of qualitative evidence, including high-impact case studies, stakeholder interviews and causality assessments and will be informed by the needs and priorities of HDR UK's partners, including members of the public.

Figure 9: HDR UK's journey from the first five years (2018-2023) to the second five years (2023-2028)



HDR UK's bold ambition to transform health discoveries through data translates into tangible goals, outcomes and measures of success, which the Institute will use to gauge progress and impact.

The Progress and Impact Framework (PIF), implemented in HDR UK's first five years, was welcomed by the [Establishment Review](#) panel (Establishment Review Report Section B.7). It provides quantitative and qualitative data which has successfully delivered a balanced view of the Institute's progress against the vision, goals and outcomes agreed with the core funders and HDR UK Board.

Through the second five years, the PIF and associated key performance indicators will be further refined and applied monthly, quarterly and annually to continually improve and increase HDR UK's likelihood of achieving its goals, outcomes and vision (**Appendix 5: Progress and Impact Framework (PIF) – overview;** **Appendix 6: Second five years Progress and Impact Framework (PIF): Goals, Outcomes and Metrics**).

¹⁶ Numerical metrics will be balanced against a commitment to delivering high value impacts with relevance to clinical care and health policy. They offer a range of more readily measurable dimensions which offer a proxy for total impact.

The high-level outcomes HDR UK expects to achieve in its second five years include:

Infrastructure and Services:

1. **Scale** – a trustworthy, effective and efficient approach enabling many researchers to safely use large-scale, linked data
2. **Service** – frictionless services that are valued by users across the ecosystem (data custodians, funders, researchers, innovators, public, patients, practitioners, policymakers).

Research Driver Programmes:

3. **Impact** – Multidimensional evidence of better health research through use of a wide range of health data: number and quality of publications; case studies demonstrating influence on health policy, clinical practice and improving health outcomes; and longer-term demonstrable positive effects on the economy, society, culture, public policy or services, health, the environment or quality of life.

One Institute Partnerships:

4. **Sustainability** – Leverage of substantial resources beyond core funding, contributing to the UK being recognised globally as the most impactful nation for high quality health data science and establishing equitable, federated partnerships with collaborations worldwide
5. **Inclusivity** – Inclusive, UK-wide and increasingly global, team-oriented culture built on the values of transparency, optimism, respect, courage and humility.

HDR UK recognises the challenges of assessing and measuring a wide range of more and less tangible impacts. The Institute will work with its community and key partners to understand, evaluate, and communicate the full extent of impact and value for money arising from the Institute's second five years.

1.6. Governance and maintenance of high-quality services and research

1.6.1. Alignment milestones

The whole institute will work towards three high-level milestones which will be evaluated by the HDR UK Board, supported by the IAB and the UK Regional Network Leads:

- **Milestone 1: Established** (30 June 2023 - three months after the start of the second five years)
Demonstrate that the initial activities are underway and aligned to the programme and service aims.
- **Milestone 2: Delivery** (31 March 2025 – two years after the start of the second five years)
Demonstrate progress in the service delivery and development and in the research activities towards answering the research goals.
- **Milestone 3: Impact and scalability** (30 June 2026 - 3.25 years after the start of the second five years)
Evidence that the service has enabled an improved user experience, and that the research programmes have enabled better science across all dimensions.

More detail on the objectives and assessment proposed for each Milestone is provided in **Section 3.1.1**.

1.6.2. Research best practice

HDR UK is committed to ensuring that research ethics, integrity and governance have a high profile and are firmly embedded in the Institute's ethos and culture. The Institute is a signatory to six research-related concordats¹⁷.

¹⁷ Concordat to Support Research Integrity; Concordat for Engaging the Public with Research; Concordat on Openness on Animal Research in the UK; The Declaration on Research Assessment (DORA); Concordat to Support the Career Development of Researchers and the Technician Commitment.

As a distributed UK-wide Institute, HDR UK's research is conducted through partnerships with academic, NHS and commercial organisations across the UK. This is delivered through contractual agreements, which include a requirement to meet the principles and standards established in the concordats as well as any regulatory or legal requirements associated with data management.

HDR UK monitors these agreements annually through its assurance reporting – each of the Institute's research partners must be a signatory to the concordats and report on any scientific misconduct or research and data governance issues. Where required, research projects must seek approval through appropriate ethical approval processes. This approach supports a clear line of accountability between researchers and their employing institution, and between the employing institution and HDR UK. It also ensures that HDR UK minimises duplication and bureaucracy by making best use of existing processes.



HDR UK's [open access arrangements](#) for research outputs and data management are grounded in its [Attribution Policy](#) supported by the contractual relationships that the Institute has with its research partners. HDR UK is committed to being an exemplar of the Declaration of Research Assessment ([DORA](#)) principles and UKRI's commitment to [open access](#), open data and responsible research assessment.

Financial support is provided to research partners to enable open access arrangements. For its second five years, HDR UK has required open access charges to be included in budgets for Driver Programmes and Infrastructure and Services.

In addition to meeting any regulatory or legislative requirements for data management and sharing, all data custodians contributing to Driver Programmes must also make their metadata available on the Gateway to enable further use and reuse of their data, and will be required to subscribe to the UK Health Data Research Alliance [Principles for Participation](#), including the FAIR data principles.

Further details on HDR UK's approach to data management is set out in the Institute's Data Management Plan (**Appendix 7: Data Management Plan – Health Data Research UK**).

The [HDR UK International Advisory Board](#) provides international benchmarking to maintain quality across all research and infrastructure activities. Any new capabilities added to the Institute are selected based on robust independent peer-review, typically delivered in partnership with the contributing funder(s).

1.7. Literature cited

Bailey, A., Coleman, G., Harbinson, A., Khodiyar, V., Mill, N., Morris, C., Orton, C., Quattroni, P., Seymour, D., Stokes, P., & Varma, S. (2022). Five Safe Data Access Request application form.

<https://doi.org/10.5281/ZENODO.5946891>

Denaxas, S., Gonzalez-Izquierdo, A., Direk, K., Fitzpatrick, N. K., Fatemifar, G., Banerjee, A., Dobson, R. J. B., Howe, L. J., Kuan, V., Lumbers, R. T., Pasea, L., Patel, R. S., Shah, A. D., Hingorani, A. D., Sudlow, C., & Hemingway, H. (2019). UK phenomics platform for developing and validating electronic health record phenotypes: CALIBER. *Journal of the American Medical Informatics Association*, 26(12), 1545–1559.

<https://doi.org/10.1093/jamia/ocz105>

Gaziano, L., VA Million Veteran Program COVID-19 Science Initiative, Giambartolomei, C., Pereira, A. C., Gaulton, A., Posner, D. C., Swanson, S. A., Ho, Y.-L., Iyengar, S. K., Kosik, N. M., Vujkovic, M., Gagnon, D. R., Bento, A. P., Barrio-Hernandez, I., Rönnblom, L., Hagberg, N., Lundtoft, C., Langenberg, C., Pietzner, M., ... Casas, J. P. (2021). Actionable druggable genome-wide Mendelian randomization identifies repurposing opportunities for COVID-19. *Nature Medicine*, 27(4), 668–676. <https://doi.org/10.1038/s41591-021-01310-z>

Harron, K., Gilbert, R., Fagg, J., Guttmann, A., & van der Meulen, J. (2021). Associations between pre-pregnancy psychosocial risk factors and infant outcomes: A population-based cohort study in England. *The Lancet Public Health*, 6(2), e97–e105. [https://doi.org/10.1016/S2468-2667\(20\)30210-3](https://doi.org/10.1016/S2468-2667(20)30210-3)

Lambert, S. A., Gil, L., Jupp, S., Ritchie, S. C., Xu, Y., Buniello, A., McMahon, A., Abraham, G., Chapman, M., Parkinson, H., Danesh, J., MacArthur, J. A. L., & Inouye, M. (2021). The Polygenic Score Catalog as an open database for reproducibility and systematic evaluation. *Nature Genetics*. <https://doi.org/10.1038/s41588-021-00783-5>

Lyons, J., Akbari, A., Torabi, F., Davies, G. I., North, L., Griffiths, R., Bailey, R., Hollinghurst, J., Fry, R., Turner, S. L., Thompson, D., Rafferty, J., Mizen, A., Orton, C., Thompson, S., Au-Yeung, L., Cross, L., Gravenor, M. B., Brophy, S., ... Lyons, R. (2020). Understanding and responding to COVID-19 in Wales: Protocol for a privacy-protecting data platform for enhanced epidemiology and evaluation of interventions. *BMJ Open*, 10(10), e043010. <https://doi.org/10.1136/bmjopen-2020-043010>

MacTel Consortium, Lotta, L. A., Pietzner, M., Stewart, I. D., Wittemans, L. B. L., Li, C., Bonelli, R., Raffler, J., Biggs, E. K., Oliver-Williams, C., Auyeung, V. P. W., Luan, J., Wheeler, E., Paige, E., Surendran, P., Michelotti, G. A., Scott, R. A., Burgess, S., Zuber, V., ... Langenberg, C. (2021). A cross-platform approach identifies genetic regulators of human metabolism and health. *Nature Genetics*, 53(1), 54–64.

<https://doi.org/10.1038/s41588-020-00751-5>

Sheikh, A., McMenamin, J., Taylor, B., & Robertson, C. (2021a). SARS-CoV-2 Delta VOC in Scotland: Demographics, risk of hospital admission, and vaccine effectiveness. *The Lancet*, 397(10293), 2461–2462.

[https://doi.org/10.1016/S0140-6736\(21\)01358-1](https://doi.org/10.1016/S0140-6736(21)01358-1)

Sheikh, A., Robertson, C., & Taylor, B. (2021b). BNT162b2 and ChAdOx1 nCoV-19 Vaccine Effectiveness against Death from the Delta Variant. *New England Journal of Medicine*, 385(23), 2195–2197.

<https://doi.org/10.1056/NEJMc2113864>

Shi, T., Pan, J., Katikireddi, S. V., McCowan, C., Kerr, S., Agrawal, U., Shah, S. A., Simpson, C. R., Ritchie, L. D., Robertson, C., & Sheikh, A. (2021). Risk of COVID-19 hospital admission among children aged 5–17 years

with asthma in Scotland: A national incident cohort study. *The Lancet Respiratory Medicine*, S2213260021004914. [https://doi.org/10.1016/S2213-2600\(21\)00491-4](https://doi.org/10.1016/S2213-2600(21)00491-4)

Simpson, C. R., Shi, T., Vasileiou, E., Katikireddi, S. V., Kerr, S., Moore, E., McCowan, C., Agrawal, U., Shah, S. A., Ritchie, L. D., Murray, J., Pan, J., Bradley, D. T., Stock, S. J., Wood, R., Chuter, A., Beggs, J., Stagg, H. R., Joy, M., ... Sheikh, A. (2021). First-dose ChAdOx1 and BNT162b2 COVID-19 vaccines and thrombocytopenic, thromboembolic and hemorrhagic events in Scotland. *Nature Medicine*. <https://doi.org/10.1038/s41591-021-01408-4>

The RECOVERY Collaborative Group. (2020). Effect of Hydroxychloroquine in Hospitalized Patients with Covid-19. *New England Journal of Medicine*, 383(21), 2030–2040. <https://doi.org/10.1056/NEJMoa2022926>

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., ... Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3(1), 160018. <https://doi.org/10.1038/sdata.2016.18>

Wood, A., Denholm, R., Hollings, S., Cooper, J., Ip, S., Walker, V., Denaxas, S., Akbari, A., Banerjee, A., Whiteley, W., Lai, A., Sterne, J., & Sudlow, C. (2021). Linked electronic health records for research on a nationwide cohort of more than 54 million people in England: Data resource. *BMJ*, n826. <https://doi.org/10.1136/bmj.n826>

Zheng, J., Haberland, V., Baird, D., Walker, V., Haycock, P. C., Hurle, M. R., Gutteridge, A., Erola, P., Liu, Y., Luo, S., Robinson, J., Richardson, T. G., Staley, J. R., Elsworth, B., Burgess, S., Sun, B. B., Danesh, J., Runz, H., Maranville, J. C., ... Gaunt, T. R. (2020). Phenome-wide Mendelian randomization mapping the influence of the plasma proteome on complex diseases. *Nature Genetics*, 52(10), 1122–1131. <https://doi.org/10.1038/s41588-020-0682-6>

HDRUK

Health Data Research UK

Gibbs Building, 215 Euston Road, London NW1 2BE
+44 (0) 20 3371 1393 | @HDR_UK | hdruk.ac.uk

4 February 2022