



ONE INSTITUTE STRATEGY 2019/20

9 April 2019

Contents

Executive Summary	3
One Institute: Our 20-year vision	5
Science – activating world-class health data science	7
People – empowering the health data scientists	10
Infrastructure – uniting the UK’s health data	13
Public Engagement and Involvement	17
Horizon Scanning	18
Financial Plan	19

Purpose of document

This document describes Health Data Research UK’s strategy for 2019/20 and how this aligns to our 20-year vision and 5-year priorities. This One Institute Strategy 2019/20 is written with a professional audience in mind and is aimed at colleagues working with Health Data Research UK. It has received input from Health Data Research UK’s Board, Public Advisory Board, Science Strategy Board, research leads, our funders and other key advisors. It is supported by an internal delivery plan, which can be requested by emailing enquiries@hdruk.ac.uk

We are funded by the Medical Research Council (MRC); the health research departments of England, Scotland, Wales and Northern Ireland (National Institute for Health Research (NIHR), Chief Scientist Office (CSO), Health and Care Research Wales, HSC Research and Development respectively), the Economic and Social Research Council (ESRC), the Engineering and Physical Sciences Research Council (EPSRC), Wellcome, The British Heart Foundation and UK Research and Innovation.

Executive Summary

Health Data Research UK is uniting the UK's health data to make discoveries that improve people's lives.

By bringing together the sharpest scientific minds and providing safe and secure access to rich health data, we aim to better understand diseases and discover new ways to prevent, treat and cure them.

Over the next 20 years, our vision is for large scale data and advanced analytics to benefit every patient interaction, clinical trial, biomedical discovery and enhance public health.

We are building a data innovation engine for the NHS, academia and industry to make this happen. This includes a set of ambitious priorities that we have set out to deliver **over the next five years**:

- Improving discovery of new diseases and treatments
- Improving evaluation of safe, effective, affordable treatments for patients
- Improving healthy life expectancy for people living with a common disease
- Improving clinical innovations that enhance NHS and social care services
- Developing and applying advanced health data science to address major health challenges
- Creating more than 10,000 health data scientists
- Creating a world-leading data infrastructure and UK-wide approach to secure data services
- Earning public trust by engaging and involving people in all aspects of our research and innovation

Why is this important? Antimicrobial resistance, mental health issues, cardiovascular disease, cancers, and dementia are just some of the extensive healthcare problems facing today's society.

Health Data Research UK can help find solutions to these problems by forging partnerships between researchers, innovators, data scientists, patient groups and the public and empowering them with access to quality, large-scale health data. Insight-driven science is crucial for advances in understanding health and disease and how we can tackle them with life-changing personalised medicines, treatments and technologies.

The UK is home to the NHS, a strong life sciences industry, world-leading academic institutes, and a vibrant tech community. The population's health data will provide the UK's pioneers with an opportunity to drive innovation, grow the industry, and take a holistic view to improving the health of the nation.

We are reinventing healthcare research; changing the outdated silo mentality to an open, collaborative culture working towards a common goal – to improve people's lives.



We are delivering our mission by bringing together the following themes underpinned by a culture of transparency, engagement and involvement to earn public trust in the use of health data for research:

Science

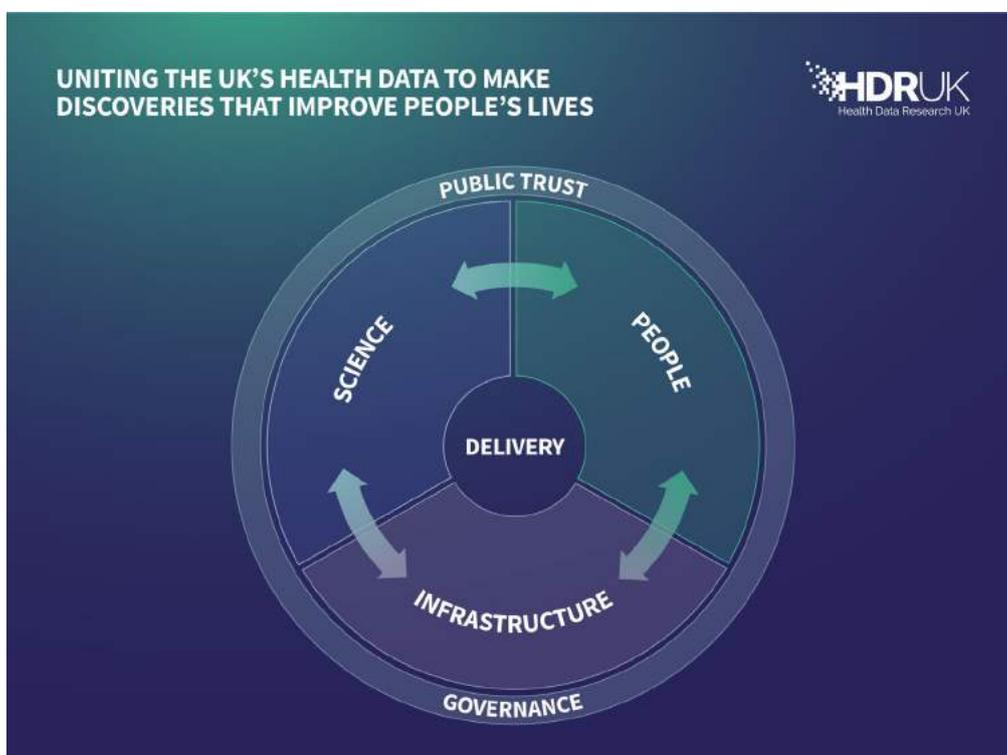
The UK has a rich scientific heritage and holds some of the most robust and diverse health data of anywhere in the world. The stimulation of world-leading scientific research, environmental data, insight from wearables and NHS routine data will be the catalyst in the discovery of new diseases, increase in clinical trial participation, insights into the best treatments, and the elimination of some of the most devastating diseases. We will do this by developing innovative ways of extracting new knowledge from complex and diverse health data and empowering the community of NHS and other healthcare innovators to harness its potential.

People

We will inspire the next generation of health data scientists; attracting innovative thinkers who will use large-scale data in novel ways to solve health problems facing our society. We are in a bright new world of original ideas, ground-breaking discoveries and high-profile partnerships. Our training programmes will support and empower people from across the world – including mathematicians, scientists, data analysts and engineers – to lead the data science revolution.

Infrastructure

We are developing cutting-edge infrastructure and a strong governance framework co-created with the public to enable secure, trusted research data services and support improvements in patient care. Research and innovation will flourish when boosted with comprehensive national data from electronic health records, medical images, wearables, genomics, environmental sources and biomedical research.



Delivery highlights

Health Data Research UK was established in April 2018. In 2018/19 we delivered:

- **Science:** we have established six HDR UK substantive research sites, comprising 22 UK universities, NHS and research institute partners, with a common Institute Agreement. We are transitioning to national programmes and have already: established national implementation projects; developed joint grants, for example a £10 million grant proposal with the British Heart Foundation; played international leadership roles on Alzheimer’s Disease; and developed joint national initiatives with the Alan Turing Institute and Medicines Discovery Catapult.
- **People:** we have created our initial cohort of 180 Principal Investigators (PIs) and co-Investigators, and 46 non-clinical UKRI and Rutherford Fellows. We have built a strong leadership team of Research Directors and Associate Directors across the UK. We are setting up Master’s degree programmes and studentships in health data science and jointly submitted grants for a PhD programme. We have built national partnerships with the NHS, Software Sustainability Institute and Open Data Institute.
- **Infrastructure:** we were commissioned to lead delivery of the UK-wide Digital Innovation Hubs Programme. During the year we launched 10 sprint exemplar projects, which will deliver by November 2019; and created the UK Health Data Alliance with national data custodians from all four nations.
- **Engagement:** we established a Public Advisory Board following an application process that attracted 100 members of the public. We engaged over 2,500 people in the design and development of the Digital Innovation Hub Programme from the public, academia, industry, NHS, government and charities.

One Institute: Our 20-year vision

We will be a world-leading, international health data research institute, distinguishing the UK as the core of the health data science revolution. Over the next 20 years, access to rich, far-reaching data combined with new scientific insights will transform research and innovation, and lead to significant advancements in our understanding, preventing, diagnosing and treating of disease.

To achieve this aspiration, Health Data Research UK is bringing together the NHS, the public, industry and academia as one combined force. A collaboration of this scale has never been seen before and it will significantly transform the future of healthcare by developing and championing the best in *science*, *people* and *infrastructure* underpinned by earning public trust. Each of these elements has a crucial part to play in advancing healthcare improvements and will fuse together to create an unparalleled ecosystem for health data science and innovation to flourish.

Our Strategic Framework outlines how we will develop these three areas and our plan for 2019/20, our five-year big ambitions, and our 20-year vision.

STRATEGIC FRAMEWORK FOR 1 YEAR, 5 YEARS AND 20 YEARS

OVER THE NEXT 20 YEARS our vision is for large scale data and advanced analytics to benefit every patient interaction, clinical trial, biomedical discovery and enhance public health. We are building an innovation engine for the NHS, academia and industry to make this happen.

5 YEAR AMBITIONS

- IMPROVING DISCOVERY** of new diseases and treatments
- IMPROVING EVALUATION** of safe, effective, affordable treatments for patients
- IMPROVING HEALTHY LIFE EXPECTANCY** for people living with a common disease
- IMPROVING CLINICAL INNOVATIONS** that enhance NHS and social care services for people

WHAT WE ARE DOING IN 2019/20	SCIENCE			
	Understanding the causes of disease National Project: Understanding disease using multi-omics + local	Better, faster, efficient clinical trials National Project: Building a national trials strategy + local	Improving public health National Project: Measuring and understanding multi-morbidity + local	Digital health insights National Project: Engaging with NHS, research and innovation leaders + local
	Human Phenome Project National Project: Human phenome online library + local			
	Applied Analytics National Project: Reproducibility of Artificial Intelligence and Machine Learning in medicine + local			
	Health data science to tackle disease , including dementia, cardiovascular disease, cancer and rare genetic disorders			
	PEOPLE			
	Data science literacy, NHS analytics community, HDR UK UKRI/Rutherford Fellows, Doctoral research			
	School leavers, postgraduate, post-doctoral and pathways to seniority programmes			
	INFRASTRUCTURE			
	Deliver Hubs Deliver DIH Sprints: Myeyesite; Cloud-based integration of phenotype and genotype data for rare disease research; Developing an integrated data repository and analysis platform to maximise the impact of large-scale cancer cohort studies	Deliver Hubs Deliver DIH Sprints: Using Routine NHS Data to Accelerate Clinical Trial Recruitment	Deliver Hubs Deliver DIH Sprints: Using data to improve care for patients with heart failure	Deliver Hubs Deliver DIH Sprints: PED4PED; Enhancing Clinical Responses through Digital Information Transfer and Alerting by a Hospital-wide Smartphone Application
Infrastructure layer				
UK Health Data Research Alliance				
ONE INSTITUTE				
Public engagement	Communications	Professional leadership		
PUBLIC TRUST				

We will develop and apply advanced health data science to address major health challenges

We will create 10,000+ health data scientists

We will create a world-leading data infrastructure and UK-wide approach to secure data services

We will earn public trust by engaging and involving people in all aspects of our research and innovation

HORIZON SCANNING

Looking ahead at next 5-20 years: International Advisory Board, scientific discoveries, funder priorities

Science – activating world-class health data science

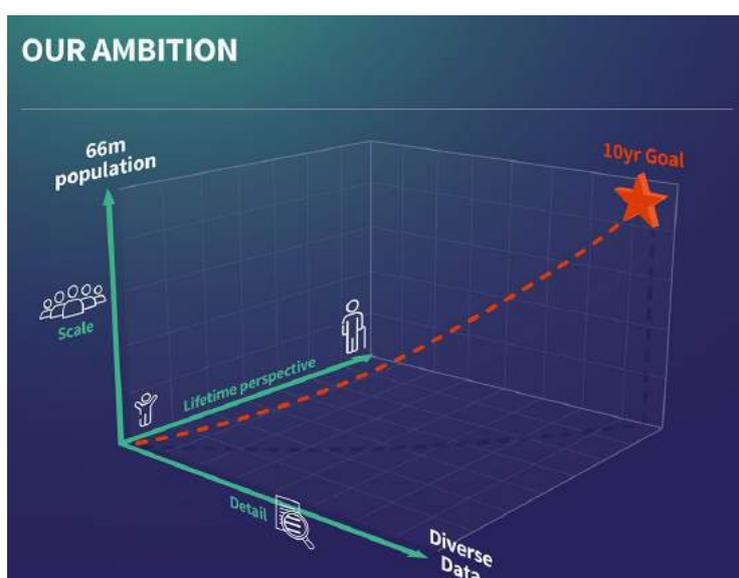
Ambition: We will develop advanced health data science tools and technologies to analyse complex and diverse health data and seize its power to solve health challenges

We are leading a new era of scientific convergence, where computational, population, social and biomedical research communities come together to tackle complex health questions using large scale health data. HDR UK is combining health expertise with new tools and technologies to enable a data-led approach to reduce the time from scientific discovery, to intervention, and through to innovative patient treatments. We aim to revolutionise the ability to understand, prevent, diagnose and treat the most prevalent and devastating diseases, such as dementia, cardiovascular disease, cancer and rare genetic disorders.

Underpinned by our Infrastructure Programme, we will attract the leading maths, statistics, computer science and interdisciplinary teams and provide development opportunities, through our talent and training programmes to maximise their strengths. We will challenge them to develop analytical methods and support them to produce and scale tools to shape the future of healthcare.

The power of health data science at scale will be demonstrated by:

- **Scale of population:** we will use the largest health research and patient data available, at a national and international population scale.
- **Scale of data:** we will find new ways to use diverse data from electronic health records, to genomics, imaging, mobile and wearable data, and environmental and other social data.
- **Scale over time:** we will deliver a life-course perspective, from early to adult life and across generations, to increase the duration and frequency of data available for research use.
- **Scale of services:** the tools and technologies we develop will be engineered to be reproducible, reliable and scalable across different datasets, and made available to researchers globally.
- **Scale and speed of impact:** the outcomes of our research will be rapidly applied to frontline decision-making for patients, health providers, populations, policymakers and regulators.



Our immediate focus is to shift from science delivered by individual sites to a new national approach. We will connect new interdisciplinary teams, initiate national implementation research projects and build new partnerships that capitalise on the world-leading research that exists in our partner universities and research institutes.

This is the start of a journey, we want to demonstrate early impact, and we will iterate our research questions as we learn along the way.

Priorities for 2019/20 and beyond

Health Data Research UK will create the new health data science tools and technologies needed to address major health challenges.

Health Data Science Tools & Technologies

We will develop safe, secure and scalable technologies to keep pace with the exponential increase in the generation and use of data, from imaging to wearables, medical devices and social media. We will utilise mathematical, statistical and computational skills to support advanced curation, data aggregation, integration and analysis. These tools and technologies will transform health research and innovation. Akin to the advent of the microscope it will allow researchers to ask previously unanswerable questions, generating new knowledge and insights to better understand, prevent and treat disease.

We are focusing on two areas of data science that have the greatest potential to unleash insight and inform decision making:

Priority 1: HDR UK will drive forward the Human Phenome Project – applying the science of measurement to robustly and reproducibly characterise all aspect of human health and disease.

Unlike other research disciplines that have taken advantage of the power of large scale data - be that the Human Genome Project or particle physics enabled by CERN - the study of human health characteristics (the phenome) is significantly hindered by the lack of a well-curated resource for researchers and clinicians to deposit algorithms, tools, methods and training material required to reproducibly interpret complex health data.

To address this, HDR UK will create a critical mass of data scientists, engineers and health data researchers to develop an online open-access, standards-driven library of complex health and disease characteristics (phenotypes) from diverse health data. This will significantly reduce duplication of effort and will be essential to the delivery of health research at scale.

- In 2019/20, we will generate the first early prototype of a national human phenome online resource.
- Our long-term goal is to create the global-leading human phenome web-based resource, used by millions of clinicians and researchers world-wide.

Priority 2: Applied Analytics - HDR UK will bring the power of Machine Learning (ML) and Artificial Intelligence (AI) to large and complex health data.

Working in partnership with the Alan Turing Institute and other data science leaders, we will develop and demonstrate how AI and ML can support the detection, diagnosis, and treatment of disease. Development of these methods require safe and secure access to large scale health data. We will drive best practice through the creation of transparent, reproducible, ethical, and effective algorithms, and ensure the scalability of these new tools across different data and populations.

- In 2019/20 we will commission a national project, in partnership with the Alan Turing Institute, that aims to improve the reliability and reproducibility of Artificial Intelligence and Machine Learning in medicine.
- Our long-term goal is to dramatically reduce the time taken to develop, validate and apply advanced analytical and statistical methods to deliver robust and reliable health insights.

Demonstrating the power of large-scale health data to address major health challenges

Large-scale health data has huge potential to generate new knowledge, address ambitious research questions and speed up ground-breaking discoveries. We will bring distinctive data science expertise to contribute to advancements in the understanding of the causes of disease, clinical trial delivery, public health and digital health insights.

We will work in partnership with different research communities to ensure that data science-led insights transform the speed and approach to discovery, prevention, treatment and evaluation of health and care interventions.

Priority 3: Understanding the Causes of Disease – Uniting data to identify new causes of disease, leading to the development of new or repurposed treatments, diagnostics and therapeutic interventions

We will advance understanding of disease prediction, cause, progression and response to therapies by linking diverse molecular data with routine patient data. This deep analysis will shift disease classification away from symptoms and pathology to the molecular causes of disease. This approach will accelerate insights into the causes of common and rare diseases and lead to new or repurposed interventions.

- In 2019/20 we will develop a national approach that brings together the UK's diverse and rich molecular research cohorts with routine health data, to create a single UK multi-omics & electronic health record (EHR) consortium.
- Our long-term goal is to speed-up the discovery of new causes of complex disease, leading to the rapid development of new or repurposed therapies.

Priority 4: Better, Faster and More Efficient Clinical Trials – In partnership with data custodians, we will improve the recruitment, assessment and monitoring of clinical trials across the UK

Health data science will transform randomised clinical trials and improve patient engagement and assessment of trial participants. Data-driven processes, supported by EHRs, will pick-up issues more quickly and contribute to improving the safety and efficacy of health interventions. Working initially with NHS Digital and the NIHR Clinical Research Network we aim to scale this approach UK-wide.

- By Autumn 2019, we will deliver a digital trial feasibility service based on the Hospital Episode Statistics dataset managed by NHS Digital.
- Our long-term goal is to make possible the delivery of larger, higher quality trials, at lower cost and in less time using routine electronic health data.

Priority 5: Improving Public Health – Informing the development of policies and interventions to increase health life expectancy and preventing the onset of disease

We will develop the tools required to deliver societal and personalised approaches to public health research, where questions can be asked in the knowledge that the right data will be available safely and securely. We will build linked, longitudinal data across biological, clinical, social and the environmental determinants of disease up to the whole 66M UK population. We will demonstrate how new technologies

can be exploited to improve public health, including person-centred data capturing social behaviours, relationships and perceptions.

- In 2019/20, we will build the first of our population cohorts, covering 20m+ lives across all four UK nations, to investigate the growing problems of multi-morbidity and falling life expectancy, particularly in the most socially-deprived regions of the UK.
- Our long-term goal is to increase healthy life-expectancy, using diverse and representative data across the whole of the UK to inform evidence based public health prevention.

Priority 6: Digital Health Insights – We will translate routine health data into insights required for personal, clinical and health system decision making.

Access to large-scale ‘research ready’ patient data will allow a range of academic, industry and clinical innovators to develop the robust and regulated digital health tools and algorithms need to (i) enhance clinical decision support at the point of care (ii) reduce clinical variation and increase efficiency of care (iii) provide patients with the high quality information they need to better manage their condition and engage with the health and care system. We will evaluate the impact that these new data, tools and innovative models of care have on longitudinal patient outcomes in healthcare settings.

- In 2019/20 we will deliver our first portfolio of digital health insights applications and engage with NHS, research and innovation leaders to build the 21st Century NHS analysts’ workforce.
- Our long-term goal is to accelerate the development of Learning Health Systems across the UK.

People – empowering the health data scientists

Ambition: We will create 10,000+ health data scientists spanning all career stages, from school-leaver to senior research manager and international opinion leaders.

The UK has a rich and diverse scientific talent base, thanks to the strength of the NHS, our academic institutions and innovative scientific and digital industries. We aim to harness this, and bring on board our international peers, to create a community of health data scientists with new skills that will dramatically change medical research and open up new, faster, smarter pathways to patient care.

Attracting those with transformative mindsets in biology, research and clinical medicine, computer science, data science and analytics, statistics, engineering, physics, astronomy and related disciplines, we will build a cohort to lead the health data science revolution.

Our training strategy will focus on integrating the three fields needed to create the health data science teams of the future: statistics, informatics, and clinical and health science. The strength of training will fill the skills gaps in the NHS, academia and industry and translate data into societal benefit.

The Cohort will be built in partnership with our delivery partners and champions including NIHR, UKRI, NHS Digital, the NHS Digital Academy, the Open Data Institute, the Alan Turing Institute and the Software Sustainability Institute. We will develop great people by identifying those with curious minds, technological appetites, and a keenness to be at the forefront of revolutionising patient care. We will create a new type of syllabus to address health data science needs of the 21st Century.

Priorities for 2019/20 and beyond

Priority 1: School leavers

Apprenticeships

Particularly attractive to large organisations, such as the NHS, our apprenticeships will provide young people with training and experience in health data science skills. Partnerships between local employers and higher education institutes will deliver new, practical programmes, with qualifications ranging from A level to Masters equivalent.

Undergraduate medical curricula

Data science skills are increasingly central to clinical practice and we want MBChB courses to reflect this. We will work with the Royal Colleges, General Medical Council, Faculty of Clinical Informatics and COPMed to promote a modern approach to data science training as a core component of an MBChB curriculum.

- In 2019/20 we will meet with employers to understand their requirements for apprenticeships in health data science; put out a call for development grants to allow employers and providers to design schemes that meet these requirements; and seek to establish or join a multi-agency working group on designing a core data science curriculum for the MBChB degree.

Priority 2: Postgraduate

Masters degree schemes

We will fund new one-year masters-level programmes in health data science that enable life sciences or quantitative sciences graduates to be effective members of health data research teams. We will stimulate programmes that genuinely integrate statistics, informatics and health science, aimed at medical students or life sciences graduates keen to develop their quantitative skills, or at core maths, physics, statistics or computing graduates who want to move into the health science arena. This will create a ‘shop window’ of modular MSc offerings across our initial 22 University partners and beyond.



The HDR UK/Turing doctoral training programme

We will work with the six HDR UK sites and the Alan Turing Institute to develop a proposal for an HDR UK/Turing Doctoral Training Programme in Health Data Science under the Wellcome Trust’s current call for PhD studentships. The programme will recruit 12 of the UK’s strongest mathematical, statistical and computational science graduates each year.

- In 2019/20 we will run an open competition for HDR UK-funded Masters programmes and bid to the Wellcome Trust for an HDR UK/Turing Doctoral Training Programme in Health Data Science with UK-wide reach.

Priority 3: Postdoctoral

The NIHR/HDR UK Incubator in Health Data Science

Our NIHR/HDR UK Incubator in Health Data Science will create 40 annual NIHR clinical fellowships over the next three years to add to our existing cohort of 46 non-clinical UKRI/Rutherford Fellows. We will invite other research fellows working in the health data science field to become affiliated to HDR UK and to

participate in a series of training events for this combined cohort. This will encourage networking and cross-fertilisation of ideas across multiple disciplines, giving the fellows ownership of shaping the agendas.

Career-break returner fellowships

We will issue an annual call for two-year fellowships to support health data scientists who are returning to work after a career-break, either to re-establish their health data science research programme or to re-train as a health data scientist.

Co-funding with Research Councils

Several of the UK Research Councils run annual fellowship competitions whose remit includes, but is not restricted to, health data science. We will offer to co-fund individual fellowships whose research agenda fits within our declared priority areas.

Summer schools

We will deliver a week-long summer school, hosted each year by a different HDR UK site. The programme for each summer school will include scientific presentations, short courses and networking opportunities. The summer schools will also foster good social and working relationships amongst colleagues from different cities, institutes and sectors.

- In 2019/20, we will formally launch the NIHR/HDR UK Incubator in Health Data Science, issue the first call for career-break returner fellowships, write to UK Research Councils to set out our co-funding offer and run the first annual HDR UK Summer School, hosted by St Andrews University.

Priority 4: Supporting Career Pathways

Career Pathways for Data Science Specialists

Data scientists are essential to the success of data-driven research in multi-disciplinary teams, yet the lack of opportunity to get a significant mention on journal papers or large grants, or to supervise PhD students, can be a barrier to promotion. We will work with the Medical Research Council, academia and industry to promote the importance of attracting and retaining senior data science specialists in the health and biomedical science domain.

Transition to and from industry

We will create training opportunities for individuals at all career levels to transition to and from industry, with a focus on pharma and tech companies. Possible activities will include: co-funded PhD studentships between HDR UK and industrial placements for HDR UK Fellows to work on short-term projects, and hackathons for HDR UK and industry staff to combine their technical expertise to address specific challenges.

Opportunities for continuing professional development

We will conduct a training gap analysis of existing provision and support the development of new training solutions, such as face-to-face courses, hackathons and online learning, as needed. To demonstrate impact, we will develop, implement, monitor, analyse and report on a comprehensive set of key performance indicators for HDR UK-badged training events.

- In 2019/20, we will produce a position paper on Career Pathways for Data Science Specialists, hold a series of HDR UK and industry hackathons, and implement a single system for monitoring and evaluating all HDR UK-badged training activities.

Infrastructure – uniting the UK’s health data

Ambition: We will create a world-leading data infrastructure and UK-wide approach to secure data services to accelerate scientific research and digital innovation

We will improve health data access by creating rich, robust and reliable infrastructure and embedding a strong governance framework with leadership and engagement of the public. Researchers from across disciplines, institutes and geographies will be able to access data and valuable insights through:

- **Digital Innovation Hubs:** specialised research services and trusted research environments based on areas of expertise, such as disease area, therapeutic area, research activity, and region.
- **UK Health Data Research Alliance:** facilitating partnerships across NHS organisations and a consistent approach to tools, technologies, governance, data provision and public engagement
- **Infrastructure layer:** providing access, including access to the HDR UK Phenomics online library resource, and de-identification, policy, standards, tools and best practices for using data

Priorities for 2019/20 and beyond

Priority 1: Establish an ecosystem of Digital Innovation Hubs

Funded by the UK Research and Innovation Industrial Strategy Challenge Fund (ISCF), the Hubs will be centres of excellence that provide expert research data services to enable NHS, academia and industry to create real-world insights, innovations and improvements from a range of health data. Each Hub will be a safe haven for research and innovation that harnesses large datasets, fast computation, and new approaches to data curation and analysis.

The Hubs will provide expert services needed by academic, industry, healthcare professionals and NHS users, and will build on areas of specialty. This includes, but is not limited to, UK-wide curated disease datasets, clinical trial services, regional real-world evidence insights, high-value domain expertise and analytical research platforms. The Hubs will make the UK an attractive destination for international investments and actively stimulate the start-up market.

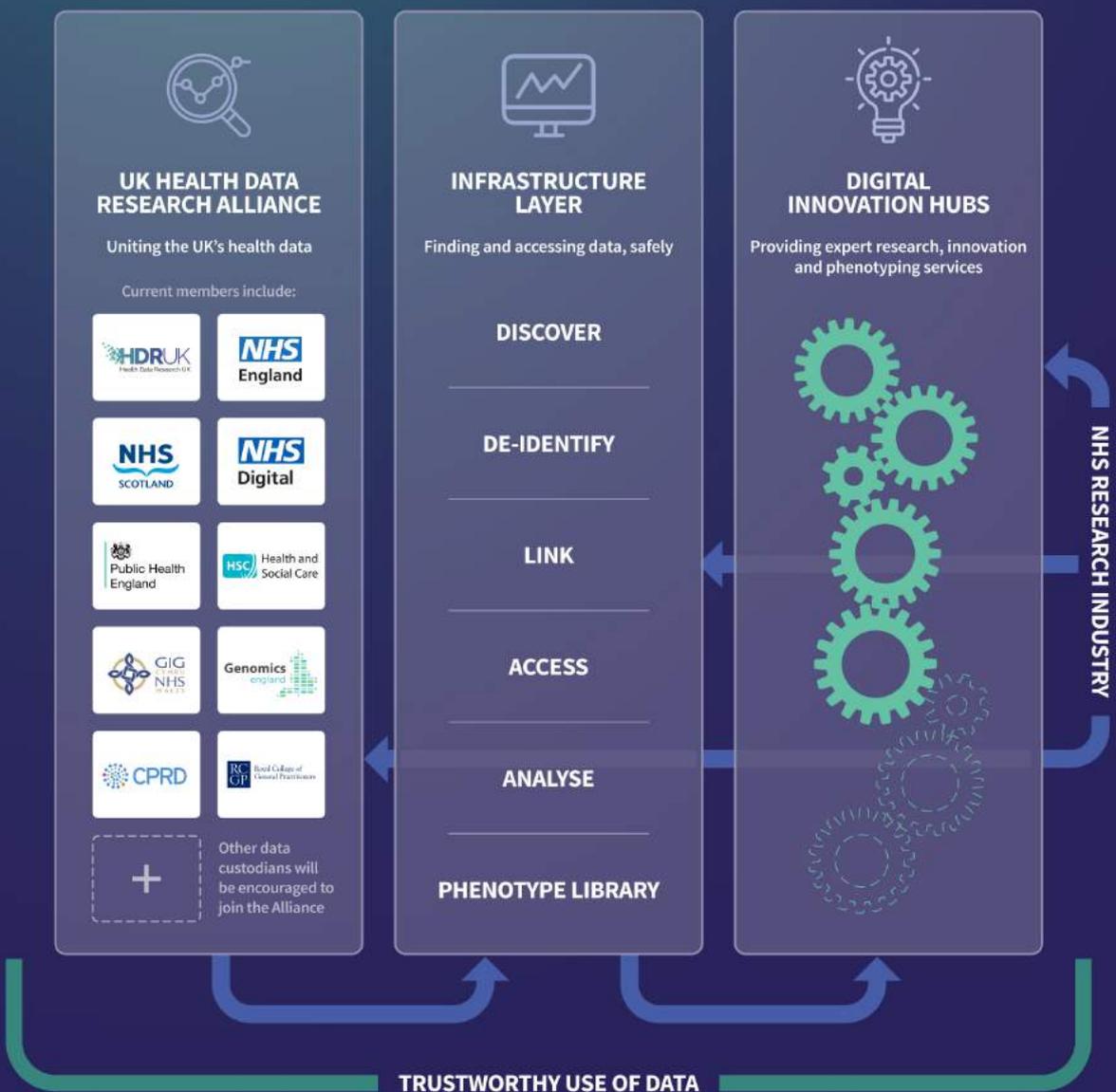
The UK Digital Innovation Hub infrastructure will provide services for specific medical technologies, disease areas or data innovation such as the proposed UK Cardiovascular Data Science Knowledge Hub being developed in partnership by the British Heart Foundation (BHF) and HDR UK.

For the Hubs to be effective and have the capability to scale to the UK level, they are reliant on the UK Health Data Research Alliance and Infrastructure levels.

- In May 2019 we will launch a competition call, outlining the criteria for selection, a participation agreement, information governance and infrastructure. Following selection, we expect successful Hubs to be live in 2019, with the opportunity to access data at scale on common diseases, including cardiovascular disease and cancer.



UNITING THE UK'S HEALTH DATA TO MAKE DISCOVERIES THAT IMPROVE PEOPLE'S LIVES



Priority 2: The UK Health Data Research Alliance

The UK Health Data Research Alliance was set up to accelerate progress in data research and innovation. Its founding members include NHS England, NHS Scotland, NHS Wales, Health and Social Care Northern Ireland, NHS Digital, Genomics England, the Royal College of General Practitioners and Public Health England. They will unite in expertise and establishing best practice in the stewardship of the UK's health data – including patient data from the NHS, genomic data and molecular data – to enable faster, more efficient access for research at scale. As the Alliance grows and evolves, further health data custodians will join, bringing more datasets to contribute to regional and local depth.

- In 2019/20, the Alliance, working closely with the Information Commissioner's Office, the National Data Guardian and with public engagement, will develop the first library of datasets discoverable through the infrastructure layer and demonstrate improved user access to datasets.

Priority 3: Develop an infrastructure layer which provides safe and secure access to the UK's health data

The infrastructure layer will provide a single, secure, accessible data infrastructure, underpinned by a governance framework and advanced data discoverability tools to provide information on the available data and its insights. Whilst the layer doesn't hold data, it provides access to a standards-driven library of complex phenotypes being developed by HDR UK, delivered to robust industrial standards, cloud-enabled and available to access either directly or through a hub.

The design will be developed by the national data providers, such as NHS England, NHS Digital and equivalent in the devolved nations; existing large infrastructures in the NHS, such as Genomics England; major research resources, such as UK Biobank; and industry. Wherever possible, we will work in partnership to evolve existing infrastructure, assets and commercial agreements.

- From May 2019, we will develop the initial delivery phase, focusing on proof of concept projects across four areas to define the critical functions. This will focus on a single point of access, privacy and de-identification, and interoperability to support mapping between datasets.

Priority 4: Maximise the impact and learning from the Sprint Exemplar projects

In 2018/19, we funded ten sprint exemplar projects to demonstrate proof of concepts for technology, methodology and research services that can inform the design of the DIH programme. The successful projects will receive support and learning opportunities. Case studies of the sprint exemplars will be used to generate public interest and increase the profile of HDR UK.

- By the end of 2019 we will have held three exemplar days and produced case studies for the ten successful projects to share their learning and support delivery. These insights will be used to develop the DIH programme and generate public interest.

Priority 5: Enable proportionate, rigorous, efficient and transparent information governance across the UK Health Data Research Alliance

Our early Information Governance (IG) work with the Alliance data custodians has found barriers to accessing data including different legal jurisdictions, insufficient resources to manage the volume of requests, complex processes that are poorly understood by users, and users with insufficiently robust governance processes to securely access the data. We will prioritise improving access across data

custodians and establishing trustworthy and proportionate IG in new organisations funded through the DIH programme.

IG is already harmonising between custodians where they are working together to link datasets, for example through NHS Digital's Research Advisory Group. In the short term, we will accelerate this by using the DIH programme to provide practical, high-priority data access and linkage problems for custodians to resolve. The high-profile nature of this work will help overcome organisational constraints and the practical work will encourage dialogue and sharing approaches between IG teams. Custodians will be expected to resolve IG in a replicable way, so that once data is accessible for one purpose and user, it is accessible for others who meet the necessary standards.

The long-term expected increase in health data usage cannot be achieved through the existing IG set-up by the UK's health data custodians. Existing processes are often complex, manual and reliant on small numbers of individuals. We will work with Alliance members to create an aligned IG strategy and delivery plan, which will include use of automated tools that are consistent across organisations, consistent interpretations of legal requirements, an understanding of user requirements while maintaining public trust, and IG skills, resource and career management.

- In 2019/20 we will establish an IG framework that encourages harmonisation across data custodians within the Alliance. This framework will include IG principles that all custodians and users of health data, including DIHs, sign up to and comparative performance measures, agreed by the Alliance, that demonstrate accessibility. These guiding principles will be developed with the HDR UK Public Advisory Board, and with the HDR UK International Advisory Board, complementing our work with the ICO and National Data Guardian.

Priority 6: Facilitate fair capture of value by the NHS and government for the benefit of patients and UK tax payer

While the initial £37.5m funds over four years gets the programme started, it will only be sustainable through a Participation Agreement that encourages data use and feeds value back to the NHS and wider UK society. Our model will build on and help to operationalise the principles set out in the Life Sciences Sector Deal 2^[1] and will remain relevant as the health, technology and life science sectors change. The Hubs are a new service, reaching new customers, and we need a model that allows us to learn and respond to the market. The starting Participation Agreement has three parts; each will be observed during the first year and each aspect of the model will be revisited at the end of the first year.

- **Hub operating models:** The Hubs will be encouraged to develop their own, sustainable models that provide a valued service for small-medium sized enterprises and individual researchers to access data in a secure, affordable way.
- **Use of the infrastructure layer:** Users will subscribe to an annual agreement and processing arrangement, both starting at zero for the first 12 months, to use the infrastructure layer.
- **Accreditation and Assurance:** A kite mark will be developed to accredit researchers, data users, safe havens, and quality of datasets and will in the future be used for assurance. There will be a fee for assurance and use of this kite mark.

^[1] commercial arrangements will: Have an explicit aim to improve the health and care of patients in the UK; Be mutually-beneficial and fair terms; Not inhibit the ability to maximise the value or use of NHS data; Be transparent, clearly communicated; Fully adhere to all national level legal, privacy and security obligations

There will be a minimum set of consistent standard terms and conditions for accessing the data through the infrastructure layer, that will be aligned with the National Centre of Expertise, and include: no exclusivity arrangements over the data made available through the Alliance and the Infrastructure Layer; adherence to the non-negotiable HDR UK Intellectual Property principles; and no replication of the full dataset.

Any user of the data who generates a new, derived dataset can seek to make this a part of the Alliance, that could also be discoverable through the Infrastructure Layer or the Hubs, with terms of access set by the data custodian, provided they meet Alliance data stewardship criteria.

HDR UK will have full visibility of the data use through the Infrastructure Layer and will collect data on the operation of the market. HDR UK will provide reports on the market and will use these to guide the annual terms of use.

- In 2019/20 we will create and implement the Participation Agreement through workshops with policy bodies (HM Treasury, the Office for Life Sciences and the National Data Guardian), industry, NHS and legal advisors

Public Engagement and Involvement

Ambition: We will encourage more effective health data science by involving people at all stages of the innovation process

Interactions with the public will provide opportunities for mutual learning and instil the benefits of using health data for research. The public will be with us at every step to inform our planning and delivery and ensure we are acting in their best interests and with their backing.

Working in partnership with national organisations and charities, including the NHS, Understanding Patient Data, Genomics England, and the Association of Medical Research Charities, we will maximise opportunities for public engagement. Our narrative and delivery will be aligned to strengthen our combined reach and provide clarity on how and why UK organisations are using their health data. We will learn from and share best practice with UK and international organisations and seek opportunities for strategic partnerships with the commercial sector.



Priorities for 2019/20 and beyond

Priority 1: Patients and the public will inform our direction, policy, decisions, planning and delivery

Our Public Advisory Board, established in January 2019, will guide our strategic delivery, provide advice on delivering our programmes, and ensure that our work is focused on improving patient and population health outcomes. We will develop a network of public engagement experts across the UK Institute, mapping existing expertise and supporting the creation of regional groups, where needed, in our six research sites, the Digital Innovation Hubs, and in the UK-wide infrastructure. Patients and the public will be invited to participate in all key activities and events.

- In 2019/20, we will embed public engagement into our governance via our Public Advisory Board; clarify the expectations for public engagement in the Digital Innovation Hubs; establish a public engagement network across all sites; and ensure members of the public are present at all major external events.

Priority 2: We will be transparent, open and accountable on how and why data are gathered, stored, secured, safeguarded, accessed, used and shared

We will take the values of open research to the next level by sharing our learnings, ideas, opinions, research projects, strategy, and expertise with the public via our different communications channels. We have a commitment to lead a culture of transparency across Health Data Research UK and will work with our Public Advisory Board, and other representatives, to ensure our information is delivered in accessible and innovative ways.

- In 2019/20, we will work with our partners to publish case studies with patient testimonials of how data is used for health research, clearly showcasing the benefits to individuals. We will run briefings for the press and media on the benefits of trustworthy health data research. We will build a dedicated area on the website to explain how HDR UK researchers use data and share examples and learnings of this.

Priority 3: We will be innovative in our approach and use mechanisms and stories that help earn public trust

Technology is creating opportunities for people to engage in new ways, with people trusting influencers, reviewers and sector leaders, over established institutions. We will capitalise on this by working with clinicians, researchers, academics and thought leaders to lead our engagement activity, advocate health data research, and instil the benefits of using data for research. Our scientific and research outputs will be shared through newsletters, case studies, podcasts, video and animation as examples of why our health data research will make a difference to current and future generations' lives.

- In 2019/20, we will review international best practice for public engagement. We will create a series of materials, including an animation, that clearly explain HDR UK's strategic approach and the Digital Innovation Hub Model to a public audience. We will identify clinical, academic and public spokespeople to share our messages and engage people in debate.

Horizon Scanning

Ambition: We will look ahead to the next 5-20 years to make HDR UK a world-leading, international health data research institute

While our current focus is on establishing HDR UK within the UK, we are already engaging on the International Stage, for example as part of the Alzheimer's Disease Data Initiative Interoperability Working Group with Gates Ventures.

- In 2019/20 we will establish the International Advisory Board and agree our international strategic priorities.

Financial Plan

Ambition: We aim to develop One Institute with excellent stewardship of resources

The Institute has a small but strong core supporting the large strategic programmes, and our organisation branches out into and works with numerous other organisations and cross-sector partners. We maximise our impact by building on existing activities, developing current programmes for better efficiency and leveraging external resources, including matched contributions and complementary funding.

Our success is dependent on demonstrating the efficient use and maximum impact of resources across all partner organisations.

Priorities for 2019/20 and beyond

Priority 1: Excellent stewardship of resources

Resources will be controlled and managed effectively and responsibly by an efficient, expert and professional team. A steady-state cost base will be established and maintained.

Priority 2: Demonstrating impact of funds

We will monitor our funds to demonstrate the use of contributions in kind, complementary, leveraged and matched funds to maximise the impact of our activities.

Priority 3: Financial sustainability

We will develop a financial sustainability plan to support the future of the Institute, incorporating the on-going funding from our Core Funders and commercial income development. Where appropriate opportunities arise, we will seek to diversify our income and attract significant additional direct and leveraged funding to support our programmes.

2019/20 Expenditure

