COVID-19 Health Data Research

29 September 2020 - Weekly update for SAGE & UKRI/DHSC

Authors:

Andrew Morris, Health Data Research UK
Ben Gordon, Health Data Research UK
Carole Morris, Public Health Scotland
Caroline Cake, Health Data Research UK (lead)
Cathie Sudlow, BHF Data Science Centre
Charlie Davie, DATA-CAN
Clara Fennessy, Health Data Research UK
David Seymour, UK Health Data Research Alliance
John Aston, Home Office (SAGE sponsor)
John Deanfield, NICOR
Mark Parsons, Scotland National Safe Haven
Members of the HDR UK Public Advisory Board
Melissa Lewis-Brown, Health Data Research UK
Ming Tang, NHS England and Improvement
Nilesh Samani, British Heart Foundation
Rhoswyn Walker, Health Data Research UK
Ronan Lyons, SAIL Databank (UKRI/DHSC sponsor)
Sara Hiom, Cancer Research UK
Tom Denwood, NHS Digital
Alice Turnbull, Health Data Research UK
Ian Young, Health & Social Care Northern Ireland

COVID-19 Health Data Research recommendations – 29 September 2020

Health data research insights on COVID are continuing, with 111 research questions, 158 (+8) projects active within the national data Trusted Research Environments (TRE), 158 (-21) in development, and 781 pre-print publications and 46 published papers. Incremental progress across all recommendations.

HDRUK
Health Data Research UK

#	5 Recommendations endorsed by SAGE on 11 June	Progress on SAGE actions identified on 11 June
1	All **swab & antibody testing programmes** data to be securely linked and used for research. Requires unparalleled cooperation across all four nations between NHS organisations, PHE, data custodians, academic endeavours, and technology partners, whilst building public trust.	**SAGE ACTION: HDR UK** to work with partners to plan and create a serology and testing data research asset that is linkable to other data sources. **PROGRESS:** Award made by UKRI/NIHR - proposal led by Philip Quinlan, Emily Jefferson and partners will commence 19 Oct.
2	Further research, undertaken collaboratively with international partners where appropriate, should address why **BAME groups** have a higher rate of severe COVID-19 outcomes. This will help to target the best interventions and inform the response to future public health crises.	**PROGRESS**: Further insights being generated (see next page). Special Interest Group, to include UK Health Data Alliance Data Officers Group, being set up to review current landscape and issues around ethnicity coding, towards enacting the Alliance Board commitment to improve consistency and quality of ethnicity coding, enabling data use to increase the representativeness of research.
3	Enhance data capture on patients and staff in **care homes**, in particular interconnections between settings, to enable research on health, transmission and outcomes. Clarify appropriate use of national Trusted Research Environments for consolidation of relevant care home COVID-19 data.	**PROGRESS**: Initial reviews of the NHS Digital adult social care management coronavirus status data collection show rapidly increasing data quality and completeness across CASPA members. CASPA are exploring using an opt-out model to increase provider coverage. Ongoing work to understand how care providers can access and use the data to inform service provision and to identify driver research questions to further explore data utility. Establishing a broader research response still requires more accessible individual-level data for care home staff, residents and domiciliary care (care within own home) sector.
4	Accelerate access to **restricted national datasets**, since lack of availability is holding back crucial research.	**PROGRESS**: England Testing data and CHESS available to priority studies to request but not yet via routine access requests and continued issues with missing data. Ongoing engagement with NHS Digital to understand and resolve issues.
5	Commission large scale collaborative analyses of the **long-term impacts** of health and social care changes during the COVID-19 lockdown on major diseases. This will require access to linked data from a range of sources (including from COVID-19 laboratory tests, primary and secondary healthcare, death registries, disease-specific audit/registry data). In addition, linkages to cross sectoral data beyond health will be essential to understand the wider impacts of COVID-19.	**SAGE ACTION: HDR UK** to work with ONS and others to accelerate linkage of cross-sectoral datasets. **PROGRESS:** HDR UK continues to identify and prioritise datasets and linkages to support priority research questions. BHF Data Science Centre: 19 cardiovascular analysts from 6 institutions actively working via access to linked datasets in the NHS Digital TRE, including primary care GPES data which includes over 4 billion journal data entries across population of over 56 million representing 97.5% of GP practices in England.

Priority research questions with new insights generated this week – 29 September 2020

Health data research on COVID-19 continues to grow, now reaching 781 (non peer-reviewed) pre-prints & 46 published papers

HDRUK
Health Data Research UK

Topic	Insights from ongoing studies (links provide further details):
Surveillance	• Evidence of substantial ethnic inequalities was found in the risk of testing positive for SARS-CoV-2, ICU admission, and mortality, which persisted after accounting for explanatory factors, including household size. It is likely that some of this excess risk is related to factors not captured in clinical records such as occupation, experiences of structural discrimination, or inequitable access to health and social services. Prioritizing linkage between health, social care, and employment data and engaging with ethnic minority communities to better understand their lived experiences is essential for generating evidence to prevent further widening of inequalities.
Immunity	• The GenOMICC consortia study has identified genetic traits in those experiencing critical illness with COVID-19, relating to key host antiviral defence mechanisms, and mediators of inflammatory organ damage in Covid-19. Both mechanisms may be amenable to targeted treatment with existing drugs. Large-scale randomised clinical trials will be essential before any change to clinical practice. • Evidence is emerging that suggests a moderately increased risk of COVID-19 mortality amongst people living with HIV. However in some cases, HIV antiretroviral therapy regimens were associated with a lower risk of adverse COVID-19 outcomes – these analyses are susceptible to confounding by comorbidities, so require further study.
Longitudinal health	• A study looking at access to health services and the influence of sex, ethnicity, socio-economic position (SEP) and burden of co-morbidities, found that the UK's lockdown approach during the COVID-19 pandemic appears to have deepened existing health inequalities, impacting predominantly females, ethnic-minorities and those with chronic illnesses. Authors recommend that public health authorities need to implement urgent policies to ensure equitable access to health and care for all in preparation for a second wave. • Analysis of primary care data from a deprived urban population, found that diagnoses of common conditions decreased substantially between March and May 2020, suggesting a large number of patients have undiagnosed conditions. A rebound in future workload could be imminent as COVID-19 restrictions ease and patients with undiagnosed conditions or delayed diagnosis present to primary/secondary health-care services. Such services should prioritise the diagnosis & treatment of these patients to mitigate potential indirect harms to protect public health.
Treatments	• A significant proportion of COVID-19 patients have hypertension and are treated with medicines (e.g. angiotensin-converting enzyme I inhibitors, aka ACE inhibitors), which have been postulated to influence susceptibility to SARS-CoV-2. Analysis of a large primary care database showed no significant associations between prescription of either ACE inhibitors or ARBs (angiotensin II type-1 receptor blockers) and all-cause mortality during the peak of the COVID-19 pandemic.

Patient and Public Voice Feedback

Priority area of focus should be to explore chronic effects of COVID-19, particularly given the increasing number of people who are now exhibiting signs of Long Covid.

20 COVID-19 taskforce calls with **90** clinical and health data research leaders engaged

1430 academic, industry and NHS participants in COVID-19 Slack channel with 10 sub-channels

111 health data research questions identified – 42 prioritised

781 COVID-19 pre-print publications

Click here for a link to the full prioritised list of questions, status, and prioritisation process

COVID-19 dataset availability and status of projects using the data – 29 September 2020

Progress on data linkage to key UK wide project datasets including COVID-19 Clinical Information Network (CO-CIN) and COVID-19 Genomics (COG-UK) Consortium (COG-UK) as the study cohorts continue to increase in size. Overall reduction in TRE project pipeline.

HDRUK
Health Data Research UK

KEY

- Data flows specified but not yet agreed
- Data flows agreed but not yet available for linkage
- Fully available

1. Pillar 2 testing data for England continues to be available only by exception.

2. Daily flow of Welsh COVID-19 Test, Trace and Protect (CTTP) data now available in SAIL Databank.

Core COVID-19 Datasets available for linkage	England (NHS Digital Data Processing Service)	Scotland (National Data Safe Haven)	Wales (SAIL Databank)	Northern Ireland (Honest Broker Service)
Primary Care				
Pillar 1 COVID-19 Testing Data				
Pillar 2 Testing data (UK Gov)	N/C Delayed: Was expected Aug			Missing results prior to 26 Apr
Pillar 3 & 4 Testing data	N/C No confirmed date		Pillar 3 available	Data flow in place, no tests yet
Community Prescribing				
Critical Care (CHESS, ICNARC, SICSAG)	N/C CHESS expected in DARS Sep			N/C - Under review
Personal Demographic Service				
Secondary Care	SUS only - not HES			
Death registry				
Census 2011				

3. Further reduction of 'in development' projects as more requests progress to active research and small number are no longer being pursued.

# of COVID-19 Projects by TRE stage (change from previous report)	England (NHS Digital Data Processing Service)	Scotland (National Data Safe Haven)	Wales (SAIL Databank)	Northern Ireland (Honest Broker Service)	Total
In development	37 (-6)	29 (-13)	86 (-2)	6 (-)	158 (-21)
- a/w researcher	*28 (-4)*	*Not available*	*63 (+1)*	*6 (-)*	*N/A*
- a/w data custodian	*9 (-2)*	*Not available*	*23 (-3)*	*0 (-)*	*N/A*
Submitted for IG approval	10 (+6)	9 (+1)	1 (-)	0 (-)	20 (+7)
Approved but not yet active	1 (-1)	0 (-)	4 (+4)	2 (-)	7 (+3)
Active research taking place	**50 (+6)**	**40 (+3)**	**68 (-1)**	**0 (-)**	**158 (+8)**

KEY UK WIDE PROJECTS:

- RECOVERY
- CO-CIN (ISARIC 4C)
- COG-UK
- CARDIOVASCULAR CONSORTIUM
- COVID-19 symptom study
- GENOMICC

Datasets available for COVID-19 research via national TREs for Wales Scotland and England

4. COVID-19 Clinical Information Network CO-CIN (ISARIC 4C) now has data from over 82,000 hospitalised patients available for research. Application for linkage to routine English health data in advanced stage and already in place for Scotland, with request to Wales to follow. This will enable high impact research such as outcomes for patients on the shielding list, association between diabetes and in-hospital mortality and long term consequences for hospitalised survivors of covid-19.

5. COVID-19 Genomics (COG-UK) Consortium (COG-UK) has sequenced 58,175 (almost 70%) of 84,932 global viral genomes. This is now linked to key data fields from routine electronic health records to add detail on host (patient). Viral sequencing being provided to key studies including ONS Infection Survey and GENOMICC.

6. **134 active users on COVID-19 related projects within SAIL** with mean time to project approvals of 3.4 days for September

NOTES

N/C – No change
TRE - Trusted Research Environment
IG - Information Governance
DPN – Data Provision Notice
CHESS - COVID-19 Hospitalisations in England Surveillance System
SICSAG - Scottish Intensive Care Audit Steering Group
HES – Hospital Episode Statistics
SUS – Secondary Uses Service

Paper C:

A National Health Data Research Capability to Support COVID-19 Research Questions

SAGE Reporting: 21 April 2020

Executive Summary:

There is considerable need for COVID-19 research questions to be rapidly answered to guide national (and international) decision making. HDR UK has teamed up with NHS Digital, the UK Health Data Research Alliance, NHS national data custodians in Scotland, Wales and Northern Ireland, and national providers of specialist data to provide: a) a process to streamline and prioritise the most important health data research questions; b) an approach to link data; and c) provide access to secure analytical environments for researchers to answer these questions to improve understanding and treatment of coronavirus. This approach was presented to SAGE on 14 April and described in the following paper.

This week is the first report on three areas of progress:

A. The Research Funnel (described in Appendix 1), prioritised questions as at 14 April 2020 (listed in Appendix 2) and their status in the process: This contains 44 research questions, (of which 19 were prioritised) against four SAGE priority areas (direct impact, indirect due to health care pressures, indirect due to socio-economic factors and other)
B. Linked NHS data in English, Scottish, Welsh & Northern Irish Trusted Research Environments: CHESS, SGSS, NICOR (6 sub-sets) datasets have been added into NHSD since last week
C. Information Governance and access: Agreement of streamlined approach with a single front door for each nation has been achieved

The primary insights for SAGE at this stage are:

- Research questions are currently dominated by ‘direct impact’ questions, however, others are starting to feed through for indirect impacts associated with cardiovascular and cancer
- The current rate limiting step for prioritised questions is getting to well-defined data request submissions to submit to the custodians. We are building the health data research community around this to support and accelerate this process
- Our focus for the next week is to: increase indirect impact questions (particularly in cardiovascular and cancer) and to establish the approved researcher route so that more researchers can safely access the data to answer the questions

Authors:

John Aston, Home Office (SAGE sponsor)

Caroline Cake, Health Data Research UK (lead)

John Deanfield, NICOR

Tom Denwood, NHS Digital

Clara Fennessy, Health Data Research UK

Ben Gordon, Health Data Research UK

Ronan Lyons, SAIL Databank

Andrew Morris, Health Data Research UK

Nilesh Samani, British Heart Foundation

David Seymour, UK Health Data Research Alliance

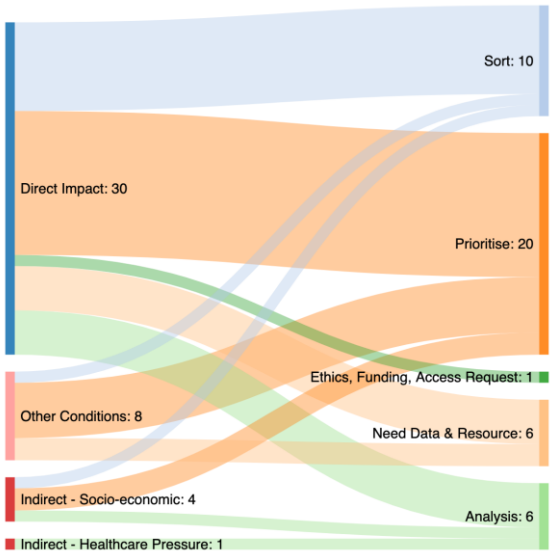
Cathie Sudlow, BHF Data Science Centre

A. Research Funnel and prioritised questions

The key messages from the first iteration of the prioritisation process are:

- We have received 44 health data research questions
- The questions address each of the four SAGE priorities, with the largest group of questions addressing the “Direct Impact of COVID-19". Currently the 10 highest prioritised questions are associated with “Direct Impact of COVID-19", however, we anticipate that this will change over time as the impact and knowledge of the virus progresses
- Questions are coming from a wide variety of sources, including from front-line clinicians. 11 are at stage 4 or beyond in the process, meaning that work is required to shape the scope of the project and identify the potential data resources required
- There has been a specific focus on identifying cardiovascular research questions, with the support of the National Institute for Cardiovascular Research (NICOR), British Heart Foundation, the BHF Data Science Centre and NHSD. In the week ahead we will be increasing focus on indirect impact associated with other diseases, in particular cancer, supported by DATA-CAN, the Health Data Research hub for cancer.
- We have also convened a discussion between national research experts and NHS Digital/Public Health England team, together with the ISARIC Study team to accelerate provision of a robust, validated answer to the high priority question – ‘can we investigate and quantify variation in the incidence (test-positive rate) and outcome of COVID-19 on individuals from Black, Asian and Ethnic Minority (BAME) populations in the UK?’ (See Appendix 2).

The following figure shows the 43 questions by SAGE area, and by stage in the funnel:



B. Linked NHS data in English, Scottish, Welsh & Northern Irish Trusted Research Environments

NHS Digital data update summary

- The major new datasets detailed in the last paper, have now landed within NHS Digital: CHESS, SGSS, NICOR (6 sub-sets)
- SUS+, the 'raw and more timely form' of Hospital Episode Statistics is being prepared for analysis
- Focus this week to conclude the cardiovascular analysis, working in partnership with NICOR and NHS England
- These datasets are being put to use as per other examples below (convalescent plasma trial, predictive modelling)
- Progress is being made landing new data sets for the purposes of COVID-19 response (GP data, social care)
- Focus is on linkage and use of data, while the Data Processing Services (DPS) is rapidly matured
- Detail on each point below

Convalescent Plasma Program (Data Access)

- Linked dataset sent to NHS Blood and Transplant (NHSBT) on 17/04 – identified ~8,000 recovered COVID-19 patients whose serum could be provided to critically ill COVID-19 patients as part of the therapeutic trial. Demographic data was linked to test and intensive care data with specific exclusions agreed by NHSBT and NHSD (i.e. national data opt out, age, S-flag, Shielded Patients List, age parameters). Recruitment started on 19/4 by NHSBT.

Intensive care and testing data (Data Collection and Analysis)

- COVID-19 Hospitalisation in England Surveillance System (CHESS – intensive care) and Second Generation Surveillance System (SGSS - testing) are now routinely being collected from PHE by NHSD, enabling *linked* analysis and dissemination. PHE lead on dissemination of these un-linked data.

National Institute for Cardiovascular Research (NICOR) / NHS England / BHF (Data Collection and Analysis)

- Exercise has been undertaken by NICOR to encourage all sites to continue to submit timely data to NICOR. All contributing Trusts have been contacted by email supported by individual phone calls.
- HQIP / NHSE, and Information Governance agreement secured for transfer and analysis of 6 NICOR datasets (covers ~3m patients) to NHSD. Data cannot be onward shared at individual level (aggregate outputs can). Transfer completed 18/04 with environments being loaded, with relevant analytical tools, and analysis commenced w/c 20/04. NHS England analytical teams, in conjunction with Prof Colin Baigent (Oxford), Prof Christopher Gale (Leeds) and Prof Mamas Mamas (Keele), leading on analysis of first questions.
- Supporting key service line questions, jointly prioritised by NICOR, NHS England and agreed with BHF

Strengthening Analytics Capability (Data Analysis)

- Advanced analytics cell established – redeployed specialist resources within NHS Digital and seconded resources from PHE. Initial focus on service and clinical questions to support SAGE e.g. indirect effects of COVID-19 on cancer outcomes and screening programmes.

National Diabetes Audit (NDA) (Data Analysis)

- Linking NDA data with intensive care data – supports urgent research request to investigate potential associations / specific risk factors for diabetic patients.

Predictive Modelling (Data Analysis and Access)

- Successfully piloted predictive demand modelling for ventilators and bed capacity (developed in partnership with Cambridge University) – national, regional & trust-level deployment by 19/04
- Informal indicators for COVID-10 death data – working closely with PHE to refine approach to enable more accurate modelling of the outbreak
- Population health analytics – continuing to investigate potential association between ethnicity and poorer outcomes

GP data (Data Collection and Access)

- NHS Digital to centrally distribute GP data for planning and research during COVID-19 - proposal agreed 15/04 at BMA/RCGP Joint IT committee and reviewed by National Data Guardian on 17/04 (outcome expected 21/4). Centralisation reduces burden on GPs to ensure legitimate, controlled and proportionate data release. Covers all GP practices in England and will require their positive action to participate. Data Protection notice (DPN) being drafted and endorsement will be sought for DPN from BMA/RGCP w/c 20/04. Plan to deliver first weekly tactical extract w/c 11/05, subject to ongoing support from the profession.

Residential and Domiciliary Care (Data Collection and Access)

- Mechanism established for daily collection of aggregate count of impact of COVID-19 on residents, service users and staff – covers ~6,000 care settings (~15% of overall market). DPN and communications plan being agreed to support start of data collection by 24/04. Solution being developed for remaining sites.

C. Information Governance and access

'Single front door' access request routes are now operational in each devolved authority with expedited data access processes:

- England via NHSX covid-19datasharing@nhsx.nhs.uk (established for COVID-19)
- Scotland via Public Benefit and Privacy Panel (PBPP) for Health and Social Care (existing process)
- Wales via SAIL databank Information Governance Review Panel (IGRP) (existing process)
- Northern Ireland vi Honest Broker Service (existing process)

Work initiated to ensure consistent definition of ‘safe user’ (accredited researcher) across the 4 nations and other COVID-19 health data science initiatives.

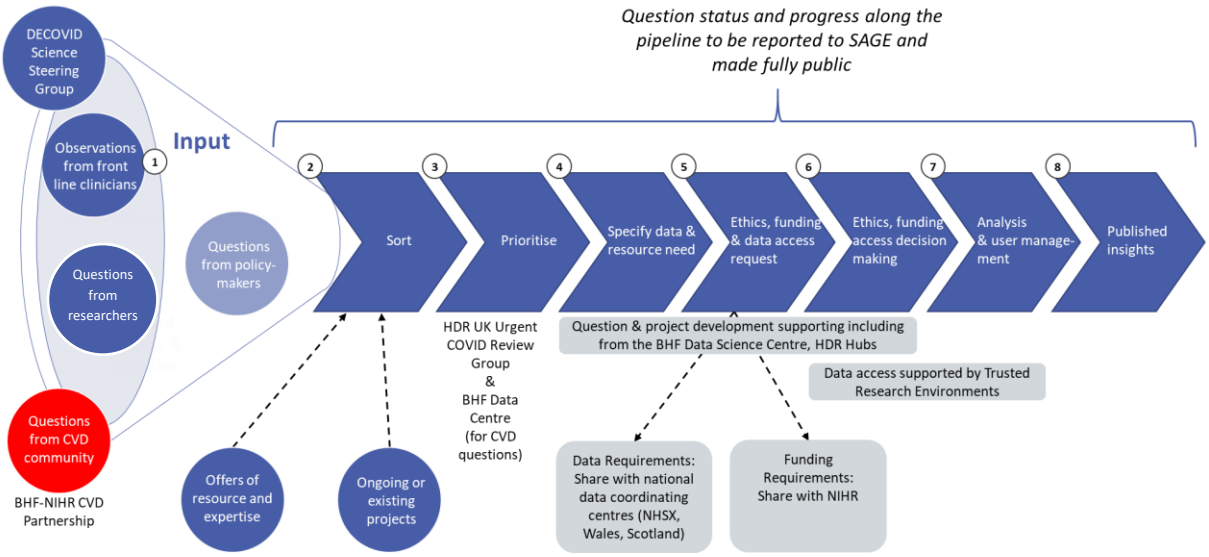
Initial set of priority research and policy questions to guide application of Information Governance policy and legal frameworks and identify any differences in interpretation or application across four nations.

Appendix 1 – Health Data Research question funnel

This process has been developed in response to requests from SAGE, NIHR and NHSX to provide a means to prioritising the volume of questions that are received based on new and ongoing research activity, observations from the frontline and areas of emerging national importance.

The aim through this process is to provide a list of prioritised questions for SAGE and others to track and to provide resources to support this work. Data custodians, such as NHS Digital/NICOR, and Trusted Research Environment providers, such as Secure Anonymised Information Linkage Databank (SAIL) in Wales will provide appropriate data access. The input stage will be aligned with other questions submitted through various communities such as the BHF-NIHR COVID Review Group and the DECOVID Scientific Steering Committee, chaired by Bryan Williams, Medical Director of UCLH.

The aim of this process is to align the various routes for questions coming in and support the direction of resources to the most urgent areas. The process is set out below:



1. Input

All questions will be directed through a single, simple webform. This will also receive input from other groups, as referenced above, in order to ensure there is a single 'front door' to the national health data research prioritisation process. There will be separate routes to collect information about ongoing projects or offers of resources.

2. **Sort**

Questions will be quickly reviewed to ensure completeness of information, lack of duplication and consistency of format. All questions, along with projects and offers of resources, will be displayed on the HDR UK Matchmaking tool, to foster national collaboration.

3. **Prioritise**

New questions will be presented to the HDR UK Urgent COVID Review Group (membership in appendix 2) for an initial prioritisation decision

4. **Specify data and resource need**

Prioritised questions will be supported to develop an understanding of the data and resources that will be needed to undertake the research. This support will be provided by elements of the health data research community, including the BHF Data Science Centre and the Health Data Research Hubs.

5. **Ethics, funding and data access request**

Once the project protocols have been developed, support will be required to ensure appropriate information governance is in place. Funders, such as NIHR, and organisations supporting data access, such as NHSX, will support at this stage.

6. **Ethics, funding and access decision making**

Once approvals have been received, the relevant data custodian Trusted Research Environment will support secure data access.

7. **Analysis and user management**

Throughout the duration of the project, progress will be tracked and reported to SAGE and other groups.

8. **Published Insights**

The Principal Investigator and project team will rapidly publish the insights from the project for peer scrutiny and to support decision-making at a national and international level.

Appendix 2 - List of Priority Questions

The questions are being prioritised by the HDR Urgent COVID Review Group (membership provided in appendix 2). The top prioritised questions (prioritised 8 out of 10 and above) and their status is provided below. Questions that had already prioritised via NIHR are shaded in green, questions related to Cardiovascular Disease are shaded in orange. In future reports we will provide a short overview of the emerging insight for each of the priority questions. The full list of 50 active questions is available in our Matchmaker Tool.

SAGE Category	Keyword	Date of input	Question	(Expected) data requested	Prioritisation Score (Median)	Days since input	Moved stage? ↑/–/↓	Current funnel stage	Question posed by (lead)	Question lead role	HDR UK lead
Direct Impact	Treatments RECOVERY	02/04/20	RECOVERY Can Lopinavir-Ritonavir vs Interferon β vs lowdose corticosteroids be effective in treating COVID 19 test +ve hospitalised patients?	Complete	Auto prioritised	22	–	7	Peter Horby	Professor of Emerging Infectious Diseases and Global Health, University of Oxford	NHS Digitrials (Martin Landray)
Direct Impact; Other conditions	Clinical characteristics Patient outcomes	02/04/20	ISARIC-CCP What are the clinical characteristics of COVID-19 positive patients; what are the determinants (genetic, other omic, prior medical history, other) of good and poor outcome; and how can knowledge of this	England, Scotland population / demographic datasets	Auto prioritised	22	–	7	Cathie Sudlow	HDR UK Scotland, Edinburgh; and BHF Data Science Centre	Cathie Sudlow

SAGE Category	Keyword	Date of input	Question	(Expected) data requested	Prioritisation Score (Median)	Days since input	Moved stage? ↑/–/↓	Current funnel stage	Question posed by (lead)	Question lead role	HDR UK lead
			help to target clinical and public health strategies?								
Direct Impact	Virus genome	02/04/20	COG-UK Can study of the whole virus genome enable scientists to monitor changes at a national scale, reveal how the virus is spreading and whether different strains are emerging?	Population clinical datasets (e.g. HES)	Auto prioritised	22	–	7	Ewan Harrison	HDR UK Cambridge; HDR UK fellow	Ewan Harrison
Direct Impact; Indirect - socio-economic	BME Patient outcomes	06/04/20	Why do BME groups appear to have increased risk of severe COVID outcomes (e.g. ventilation and mortality)? Is this caused by social, environmental and/or genetic factors? Are BME outcomes the same or different across the UK? And internationally? Does this tell us anything about the different outcomes? Initial aim: To investigate and quantify variation in the incidence (test-positive rate) and outcome of COVID-19 on individuals from Black, Asian	Administrative health datasets available in Public Health England and NHS Digital (CHESS, HES & SUS+, Mortality)	10	11	↑	4	Rhoswyn Walker	Chief Science Strategy Officer, HDR UK	Eva Morris, University of Oxford

SAGE Category	Keyword	Date of input	Question	(Expected) data requested	Prioritisation Score (Median)	Days since input	Moved stage? ↑/–/↓	Current funnel stage	Question posed by (lead)	Question lead role	HDR UK lead
			and Ethnic Minority (BAME) populations in the UK.								
Direct Impact	COVID-19 testing	07/04/20	How do we support the scale-up of COVID-19 testing, by making sure that the data that is provided on the confirmed state of COVID-19 diagnosis and antibody levels is robust and reliable?		9.5	10	↑	3	Philip Quinlain	Head of Digital Research Service at University of Nottingham	Philip Quinlain?
Direct Impact	Vaccines	02/04/20	Is the rubella vaccination (or prior exposure to German measles) protective against COVID-19 due to shared capsid sequence homology between SARS-CoV2 and Rubella?		9	22	↑	5	Adam Young, Yorgo Modis, Bjoern Neumann and Robin Franklin	University of Cambridge	Cathie Sudlow
Direct Impact	Antibody diagnostic	07/04/20	How can we accurately measure the ongoing prevalence of COVID-19 in the population following identification of a "good enough" antibody diagnostic?		9	10	↑	3	Rhoswyn Walker	Chief Science Strategy Officer	

SAGE Category	Keyword	Date of input	Question	(Expected) data requested	Prioritisation Score (Median)	Days since input	Moved stage? ↑/–/↓	Current funnel stage	Question posed by (lead)	Question lead role	HDR UK lead
			(This requires representative and random sampling from the whole or at-risk sub-populations)								
Direct Impact; Other conditions	Vulnerable patients Shielding	06/04/20	Understanding vulnerable patients: How are underlying conditions defined, and what is the impact of infection on a range of outcomes, and what are the benefits of 'shielding' and other preventive interventions?		9	11	↑	4	Harry Hemingway	Professor of Clinical Epidemiology at UCL	
Direct Impact	Patient outcomes	07/04/20	Are there any treatments which show evidence of improving outcomes for patients infected with coronavirus? Clinicians are having to make real-time decisions today, on the best possible treatment options for critically ill patients without robust evidence of harm or potential benefits of the therapeutic interventions. Better use of routine medication data could provide additional evidence to inform these decisions		9	10	↑	3	Liz Sapey, Alastair Denniston, Tanya Pank Hurst	Director of PIONEER and Reader in Acute and Respiratory Medicine at University of Birmingham; Director of INSIGHT and Consultant Ophthalmologist at University of Birmingham	

SAGE Category	Keyword	Date of input	Question	(Expected) data requested	Prioritisation Score (Median)	Days since input	Moved stage? ↑/–/↓	Current funnel stage	Question posed by (lead)	Question lead role	HDR UK lead
			prior to the definitive outcomes of clinical trials.								
Direct Impact; Indirect - healthcare pressures; Other conditions	EHRs, Patient Outcomes, Co-morbidities	02/04/2020	Where hospitals have EHRs is it possible to provide real time data on outcomes per COVID-19 admission by age and by co-morbidities by hospital? To understand whether there are hospitals that appear to have better outcomes for co-morbidity sub-groups (indicating potentially more effective interventions to learn from)?		8.5	22	–	3	Jose Sousa	CTU Manager, School of Medicine, Dentistry and Biomedical Sciences, Queens University Belfast	j.sousa@qub.ac.uk
Indirect - healthcare pressures; Other conditions	Cardiovascular disease, MI, Stroke, Disease Management	02/04/2020	What is the influence of COVID 19 epidemic in the UK and the NHS response to this on presentation, management and prognosis of non-COVID disease, in particular cardiovascular diseases such as MI and stroke?		8.5	22	↑	6-7	Cathie Sudlow	HDR UK Scotland, Edinburgh; and BHF Data Science Centre	Cathie.Sudlow@hdruk.ac.uk
Direct Impact; Indirect -	ICU, Ventilation	02/04/2020	Can we use data science to support front line decision making in Intensive Care Units? E.g. at the point of		8	22	–	4	Simon Ball: Chris Holmes; John Bradley;	Medical Director, University Hospitals	chris.holmes@stats.ox.ac.uk, john.bradley@addenbrookes.nhs.uk,

SAGE Category	Keyword	Date of input	Question	(Expected) data requested	Prioritisation Score (Median)	Days since input	Moved stage? ↑/–/↓	Current funnel stage	Question posed by (lead)	Question lead role	HDR UK lead
healthcare pressures			peak need, if patient requirements outstrip ventilation capacity, how should hospitals stratify and prioritise patients for ventilation?						Liz Sapey; Axel Heitmueller	Birmingham; Health Data Science/Alan Turing Institute; Gut Reaction – HDR Hub for Inflammatory Bowel Disease; Pioneer – HDR Hub for acute care; Discover-NOW – HDR Hub for Real World Evidence	E.Sapey@bham.ac.uk Axel.Heitmueller@imperialcollegehealthpartners.com
Other conditions	Patient outcomes	02/04/2020	What is the influence of pre-existing cardiovascular disease on outcomes of COVID-19 infection?		8	22	–	4	Cathie Sudlow	HDR UK and BHF Data Science Centre	Cathie.Sudlow@hdruk.ac.uk

SAGE Category	Keyword	Date of input	Question	(Expected) data requested	Prioritisation Score (Median)	Days since input	Moved stage? ↑/–/↓	Current funnel stage	Question posed by (lead)	Question lead role	HDR UK lead
Indirect - socio-economic	Socioeconomic, psychological	02/04/2020	What are the psychological, social and economic consequences of policies to limit the spread and flatten the peak of COVID 19?		8	22	↑	7	David Porteous and Cathie Sudlow	HDR UK Scotland (on behalf of Generation Scotland and other UK cohorts)	Cathie.Sudlow@hdruk.ac.uk
Direct Impact; Indirect - socio-economic	Socioeconomic, communication	02/04/2020	Socioeconomic inequalities: Analysis by postcode IMD. What's the best way to provide targeted and tailored messages to diverse communities?		8	22	–	4	Linsey Hovard		Linsey Hovard DAIS

SAGE Category	Keyword	Date of input	Question	(Expected) data requested	Prioritisation Score (Median)	Days since input	Moved stage? ↑/–/↓	Current funnel stage	Question posed by (lead)	Question lead role	HDR UK lead
Direct Impact; Indirect - socio-economic	Socioeconomic, response, population	02/04/2020	How can we ensure that we fully understand variations in response to COVID-19 infection at the molecular, environmental, social and economic levels, by effectively coordinating the UK's longitudinal population studies to gain a much richer understanding of disease progression and outcomes?		8	22	–	3	Mary De Silva, Debbie Lawlor, Martin Tobin. John Danesh, Nic Timpson, & David Porteous	Wellcome Trust COVID-19 Longtidudinal Population Study Steering Group	jd292@medschl.cam.ac.uk;
Direct Impact	Genomic studies	02/04/2020	How can we maximise the speed and power of host genomic studies internationally to inform drug development?		8	22	–	3	Martin Tobin	With the (International) COVID-19 Host Genomics Initiative	

SAGE Category	Keyword	Date of input	Question	(Expected) data requested	Prioritisation Score (Median)	Days since input	Moved stage? ↑/–/↓	Current funnel stage	Question posed by (lead)	Question lead role	HDR UK lead
Other conditions	Immunosuppressants	02/04/2020	Current guidelines recommend shielding is carried out for patients receiving immunosuppressants, however there is mixed evidence as to whether these patients will have poorer outcomes following coronavirus infection. Could we compare the outcomes data for patients who are receiving / not immunosuppressants and validate whether this population group are more vulnerable?		8	22	–	3	Liz Sapey, Alastair Denniston, Tanya Pank Hurst	Director of PIONEER and Reader in Acute and Respiratory Medicine at University of Birmingham; Director of INSIGHT and Consultant Ophthalmologist at University of Birmingham	E.Sapey@bham.ac.uk Alastair.denniston@insight.hdrhub.org
Other conditions	Patient outcomes	02/04/2020	Are there any concomitant treatments/ongoing prescribed medication which are making the outcomes of coronavirus infection worse for patients? This information would help clinicians to understand if there are any ongoing treatments which should be stopped as a priority when patients present with suspected		8	22	–	3	Liz Sapey, Alastair Denniston, Tanya Pank Hurst	Director of PIONEER and Reader in Acute and Respiratory Medicine at University of Birmingham; Director of INSIGHT and Consultant Ophthalmologis	E.Sapey@bham.ac.uk Alastair.denniston@insight.hdrhub.org

SAGE Category	Keyword	Date of input	Question	(Expected) data requested	Prioritisation Score (Median)	Days since input	Moved stage? ↑/–/↓	Current funnel stage	Question posed by (lead)	Question lead role	HDR UK lead
			COVID-19. For example, help to better understand existing theoretical associations between anti-hypertensives and NSAIDs and COVID-19 outcomes.							t at University of Birmingham	

Appendix 3: Urgent COVID Review Group Membership

Comprising the HDR UK Uniting and Improving the Data Delivery Group (IDG) and Using the Data Delivery Group (UDG), in addition to 4 rotating members of the HDR UK Public Advisory Board

Name	Role	Sector	Nation/Region	Data Expertise/ Disease area
Simon Ball	Executive Medical Director at University of Birmingham	Clinical practice & research	England-Midlands	Better Care Nephrology
John Bradley	Consultant Physician at Cambridge University Hospitals	Clinical practice & research	England - Cambridge	Gut Reaction Hub Renal
Caroline Cake	Chief Executive Officer of Health Data Research UK	Central Management	UK	One Institute
John Danesh	Professor of Epidemiology and Medicine and Head of the Department of Public Health and Primary Care at the University of Cambridge	Clinical research	England - Cambridge	Understanding Causes of Disease Cardiovascular
Charlie Davie	Managing Director of UCLPartners, practising Consultant Neurologist at the Royal Free London NHS Foundation Trust.	Clinical practice & research	England - London	DataCan Hub Neurology
Alastair Denniston	Director of INSIGHT, Consultant Ophthalmologist at University of Birmingham	Clinical practice & research	England - Midlands	Insight Hub Ophthalmology
Ben Gordon	Digital Innovation Hub Programme Director	Central Management	UK	Improving Health Data
Axel Heitmueller	Director of Discover-NOW and Managing Director at Imperial College Health Partners at Imperial College London	Clinical delivery & research	England - London	DiscoverNow Hub Strategy & Business Development
Harry Hemingway	Professor of Clinical Epidemiology at UCL	Clinical research	England - London	Human Phenome Cardiovascular

Chris Holmes	Health Data Science and AI Lead	Research	UK	Applied Analytics Statistical Genetics
Martin Landray	Professor of Medicine and Epidemiology at the University of Oxford, HDR UK's Science Priority Lead for Clinical Trials	Clinical practice & research	England - Oxford	NHS DigiTrials & Clinical Trials Cardiovascular
Ronan Lyons	Clinical Professor of Public Health at the University of Swansea	Clinical Research	Wales	Public Health Accident & Emergency
Andrew Morris	Director of Health Data Research UK	Clinical Research	Scotland/UK	One Institute Diabetes
Gerry Reilly	Chief Technology Officer	Central Management	UK	Uniting Health Data - Technology
David Robertson	Chair of Applied Logic, Vice Principal and Head of College of Science and Engineering at The University of Edinburgh	Research	Scotland	Applied Analytics Computing – applied logic
Elizabeth Sapey	Director of PIONEER and Reader in Acute and Respiratory Medicine at University of Birmingham	Clinical practice & research	England - Midlands	Pioneer Hub Acute Care
Neil Sebire	Professor of Pathology at UCL Great Ormond Street Hospital Institute of Child Health, Chief Research Information Officer and Director of the Digital Research, Informatics and Virtual Environment (DRIVE) Unit at GOSH.	Clinical practice & research	England - London	Standards Paediatric Pathology
David Seymour	Partnership Director at Health Data Research UK	Central Management	UK	Uniting Health Data
Aziz Sheik	Professor of Primary Care Research & Development	Clinical research (?)	Scotland	Breathe Hub

	and Director of the Usher Institute at The University of Edinburgh			Paediatric Allergy & Asthma
Cathie Sudlow	Director of the Scottish site of Health Data Research UK and Director of the British Heart Foundation UK Centre for Cardiovascular Health Data Science	Clinical practice & research	Scotland/UK	BHF Centre Neurology (stroke)
Rhoswyn Walker	Chief Science Strategy Officer	Central Management	UK	Using Health Data

X 4 Public Advisory Group Members (rotating) 75% female, 25% male

Diversity (not including public members): Gender balance: ~75% male, 25% female
Ethnicity: 90% non-BAME

Health Data Research UK:

Health Data Research UK is the national institute for health data that includes England, Wales, Scotland and Northern Ireland. Its mission is to unite the UK's health data to enable discoveries that improve people's lives. It is a not-for-profit public benefit company funded by UK funded by UK Research and Innovation, the Department of Health and Social Care in England and equivalents in Northern Ireland, Wales and Scotland, and leading medical research charities www.hdruk.ac.uk.

Paper D: Data Access Management

Aim

Health Data Research UK is working in partnership with NHS organisations and other data controllers across the UK to lead a £37.5 million government investment on behalf of UK Research and Innovation to improve the safe and responsible use of health-related data at scale for research and innovation. To achieve this, we must address the difficulties in accessing data quickly (mentioned as a major barrier by 70% of respondents across researcher and direct industry engagement [1]), and identifying the location of data and understanding data quality (mentioned as a major barrier by 55% of respondents).

Through the DIH programme, seven Health Data Research Hubs and over 25 members of the Health Data Research Alliance have made their datasets discoverable through the Innovation Gateway MVP.

Our ambition is for the Gateway to support a streamlined, proportionate approach to access requests based on the five safes model for research and innovation uses with a clear public benefit, in line with the Principles for Participation[2]. We aim to make life easier for both requestors and decision makers through a combination of automation, built-in validation, transparency of progress and the capability to host virtual data access request panels. The intention is to build on existing cross-sector best practice both nationally across the UK and internationally.

The project

In the next development phase of the Gateway, HDR UK will work with the Technology Partner to provide an Access Management module that enables users to request access to datasets, submit required information and track the progress of applications directly through the Gateway. For data custodians with existing 'in house' solutions, the Access Management module would provide validated inputs to their approvals processes, whilst for data custodians that do not currently have an automated, web-based workflow it would provide a 'best of breed' web-based access management request solution.
In order to inform next steps for the development of this phase of the Gateway, we intend to work with the health data research community including data users, data custodians, health data research hubs, regulators and public and patients.

---

[1] Industry Engagement conducted as part of the DIH Design and Dialogue phase, which involved 32 in-depth interviews with industry representatives

[2] https://www.hdruk.ac.uk/wp-content/uploads/2019/07/Digital-Innovation-Hub-Programme-Prospectus-Appendix-Principles-for-Participation.pdf

We would like to hear from this community about current practices, any challenges associated with accessing and/or sharing data and explore any opportunities for us to improve future use of health data for research and innovation.

Planned activities and proposed timeline

Input from key players, including data custodians, users and patients will directly inform the Phase 2 of the Gateway development with end goal of delivering a Minimum Viable Product at the end of October which includes a functional Data Access management module. Below is our proposed timeline for activities.

Date	Activity
March-April	Engage Hubs and custodians via individual visits (including Scotland, Wales and Northern Ireland). *NB. These plans have been impacted by COVID-19*
2 April	1st workshop RAG sub-group – understand current practice, identify commonalities, pre-validation and standard checks
April-May	Individual interviews with Hubs representatives
29 April	2nd workshop RAG sub-group – draft wireframe and potential workflow for Gateway Alliance Board meeting – will update on what we are doing and who is involved and ask for names of others to contribute
1 May	1st Milestone to inform data access request module development
Mid-May	Data access wider workshop (Alliance + others) – or questionnaire
20 May	2nd Milestone to inform data access request module development (strawman produced)
12 June	3rd Milestone to inform data access request module development (focus on dashboards and reporting)

Progress so far

In addition to individual conversations with representatives from data custodians and the Research Hubs plus a desk-top review of existing data access request forms and processes, we held the first Innovation Gateway Data Access workshop held on 2 April. HDR UK Public Advisory Board members and representatives from HQIP joined a scheduled meeting of the NHS Digital Research Advisory Group (RAG) Streamlining & Ethics workstream, which includes NHS Digital, Public Health England, CPRD, Health Research Authority, the MRC Regulatory Support Centre.

Below are the main areas highlighted at the workshop.

Transparency and the importance of public trust

From a patient and public perspective, it is crucial that the processes are explained well and clearly so that key principles and steps are clear from the start. It is important to communicate clearly who can access the data, what data and what level of data can be shared, for instance if aggregated data or not, and where the responsibilities lie. Making these concepts clear to a lay audience is critical to building public trust. HDR UK could clearly play a key role in communicating key messages in an efficient manner.

It is also important for public to know how data custodians provide access to health data and if they involve patient and the public in decision making. Ideally, all data access committees should have PPIE (Patient and Public Involvement and Engagement) through direct representation. If not, more work could be done in finding people willing to join these committees and provide training to those willing to participate. Involving patients and public in decision making is crucial to demonstrating transparency and gain public trust. As an outcome of this discussion, HDR UK will be producing a document collating information about decision-making processes for each Alliance data custodian.

Pre-submission stage

It was noted that pre-submission stages are a very important step that may be harder to integrate into the Gateway.

Before applications are submitted, data custodians might offer a pre-submission service that includes assessing the feasibility of a project and availability and utility of data, understanding the methodologies of anonymisation, understanding application requirements and identifying potential issues.

Many pre-submission discussions with applicants are about understanding what is needed before submitting an application and it is crucial that applicants know how to articulate the reasons for using the data, the level/format of data required and how the use of that data would demonstrate health and care benefit.

Recognition of Data Security and Protection Toolkit (DSPT) as an accreditation system

All the data custodians involved currently use a 'data release' model primarily. As such assessing 'safe setting' for the data is specific to each application. The use and recognition of Data Security and Protection Toolkit (DSPT) as an accreditation system was highlighted. It was suggested that more work should be done on recognising DSPT as one of the main accreditation systems across data custodians, while also being open to other accreditations in some instances. Of note, if DSPT and further accreditations are required, it will be important to have a mechanism in place to perform these policy checks (ideally in an automated manner).

Validation checks could be integrated into the Gateway

In addition to DSPT, other checks that data custodians apply at various stages of the application process are similar between data custodians and might happen at the same time in the journey. HDR UK could take note of those checks and integrate them in the Innovation Gateway in an automated fashion so that burden on those custodians would be reduced.

Training, guidance and accreditation for researchers need to be customised

The group discussed the importance of training and the provision of clear guidance about how to write good applications. Key points raised were:

- Need to consider different audiences e.g., if applicants are from commercial or academic organisations, if the requestors are applying for data access on a one-off basis or are regular requestors. It might not be efficient to train people who request access on every 2-3 years or less (which is the majority of cases for academic applicants). Should funders or universities invest in training professional data mangers rather than researchers to help with applications?
- Guidance about UK-wide research would be welcome, acknowledging differences in the devolved nations.
- Information about researchers, including CVs, could be recorded in a central place. While not all data custodians have these types of checks in place, it might be helpful to be able to check the list of people mentioned on a project automatically, pulling through data from a database.
- The concept of research passport and research visas were mentioned. The Global Alliance for Genomics and Health (GA4GH) already uses these specifications to authenticate a researcher's digital identity and automate their access to requested genomic datasets. A similar approach could be used in the context of health data.

Common Project identifier to help join things up

Participants also talked about the utility of connecting information through different systems (including IRAS). For instance, it would be helpful to pull through the relevant approvals obtained before application, simply by having a project identifier connecting all the relevant documentation digitally. This would be particularly useful to link up ethical approvals and protocols to data provider systems.

Standard Data Sharing Agreements and common legal definitions

Participants highlighted that it is important to think about the type of legislation data custodians are complying with, as consent definitions under GDPR and common law duty of confidence differ. GDPR requirements are at the organisational level, common law duty of confidence applies at individual level.

The level of the data people need is different and data are not always re-identifiable. We need to share clear definitions and provide guidance about what anonymised/de-identified means.

It was suggested that given that all data controllers hold data obtained on NHS patients, a uniform NHS Data Sharing Agreement (DSA) would be very helpful to speed up the legal processes and negotiations.

One suggestion was also to consider what happens from a legal perspective when international researchers or organisations apply for UK data. International users might use different accreditation systems and might need to comply with different policies so careful considerations should be given.

Transparency of progress across systems

One of the potential benefits that the Innovation Gateway could deliver is greater transparency to both the data custodians and data requestors about the application processes. With sufficient commonalities across systems and the main process steps common across the pieces agreed, there is scope for the Gateway to go back and forth with individual systems possibly in an automated manner and to report that progress.

Essential process steps and areas of improvement

It was acknowledged that data sharing processes are not always linear, instead it is common to go back a step and seek clarifications at different stages of the process (for instance if ethical issues are raised). In the last session of the workshop we briefly discussed what the main steps of an ideal data access process should be and started to explore whether we could identify some essential steps and some non-essential steps. We also discussed areas of improvement and ideas that could be potentially taken forward when developing the Innovation Gateway access module.

It was proposed that some of the essential steps could include:

1. Initial assessment (includes pre-submission engagement)
2. Application submitted
3. Seeking independent advice and considering approval
4. Data access request approved and DSA signed
5. Data access granted (via data released or access to TRE)
6. Data destruction/safe outputs checks

Participants also provided specific suggestions to improve current processes or for HDR UK to develop a new efficient data access management process through the Gateway.

Specific suggestions and ideas are shown in the table below.

Process steps	Questions and suggestions
Pre-submission	How do users find out what applications/approvals are needed? How do they find out where datasets are available? How do they identify which datasets they need to answer the research question? Metadata/information about datasets available: one of the gaps identified is the level of information for each dataset available at the data custodian level. The Innovation Gateway with the provision of a rich metadata catalogue provides an opportunity to fill this gap. HDR UK could provide a flowchart for each data custodian member of the Alliance with PPIE in mind so that processes could be very clear for data requestors and members of the public. Infographics might be part of the guidance provided (to explain how data is kept safe/how the application process works). Can HDR UK (a central resource pointing people in the right direction) help navigate all the data custodians who have relevant datasets for a research project (e.g. to get linked datasets)? It might help facilitate complexity of requests. There might be an opportunity to use public approval release registers to help shape examples of successful requests and also provide guidance. Guidance could include information on how and when to apply for grant funding to avoid delays in starting the research due to application/approval processes (this would be relevant to academic users).
Submission	Dashboards: both data providers and data requestors need to track progress of submissions. Notifications and emails back to requestors should all be automated.
Approvals/Documents checks	Linking approval documents to data provider: e.g. HRA approvals and protocols approved could be linked to data providers systems. This would allow data providers to check that application forms matches the CAG/HRA approval. Linking CVs, information on researchers and information on organisation could be recorded on a central place and could be pulled through. Built-in validation opportunity: security. If all data custodians agreed on a minimum level of security (e.g. DSPT) or set a minimum threshold for 'safe people', it would be helpful.
Data Sharing Agreements	Standardised national data sharing agreements for the same set of data could be developed.
Data release	Data transfer or Trusted Research Environments? There is increase demand for TRE use when accessing NHS Digital datasets. How can we switch to TRE use by default?

Audits	More information sharing might help streamlining processes. Data providers could coordinate efforts to have a standardised approach to auditing – e.g. standard procurement. For instance: if most of the data controllers rely on DSPT accreditation – can the audit concern go through DSPT?
Post-data release	Tracking applications post-approval: we could have an automated system for researchers to put in a renewal and/or have a certificate destruction triggered automatically. It would be useful to see if researchers have applied through other organisations – and link applications. There might be scope for HDR UK to look at safe outputs – how can publications and outputs be tracked? There should be checks in place that safe outputs are compliant with data sharing terms.

Next steps

A series of engagement events will be carried out throughout April and May. A second workshop with data custodians is scheduled on 29th April 2020. This workshop will focus on the Innovation Gateway and we will ask participants to input more specifically on how we can build a solution that can work for most data custodians, having in mind three use cases:

1. Data custodians with existing 'in house' web-based workflow solutions (e.g. NHS Digital).
2. Data custodians with paper-based documentation and offline workflow.
3. New data custodians or research hubs who require a web-based access management request solution.

Outputs from these workshops, alongside the outputs from engagement exercises with the Health Data Research Hubs and the user community will directly inform development of the access request module in the Innovation Gateway. We are planning to develop an infrastructure that supports access request workflow management for existing data custodians, but that also manages the end-to-end access request process for new data custodians, e.g. Hubs.

Appendix: Background information

Gateway User journey

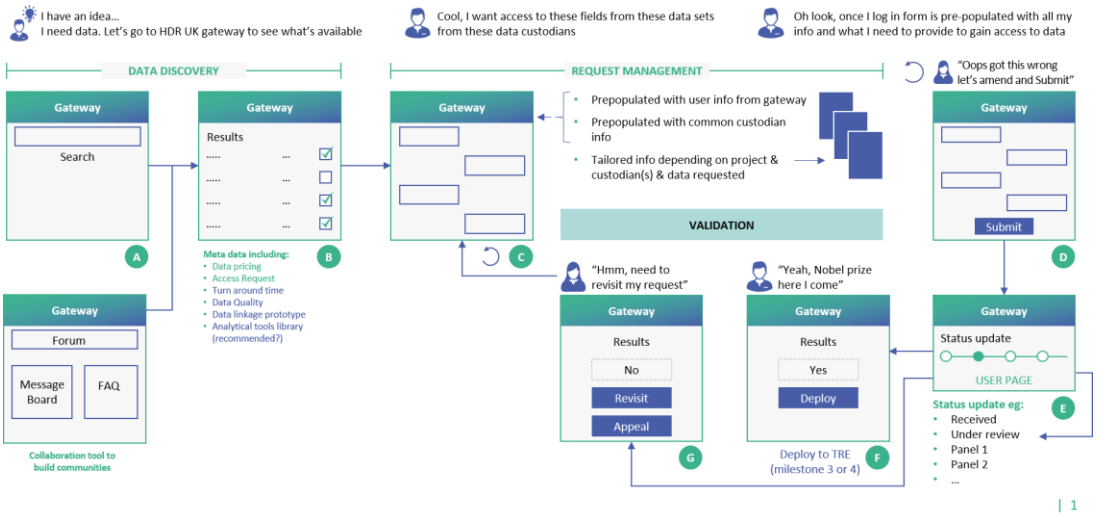


***Figure 1**. High level user journey proposed for the Health Data Research Innovation Gateway.*

Existing tools

Health data access toolkit has been developed by the **MRC Regulatory Support Centre** to help researchers navigate the processes required to gain access to routinely collected NHS health data, signposting the approvals required and exploring the issues that need to be considered during the application process.

Health Research Agency provides a number of decision support tools:

- Ethics: http://www.hra-decisiontools.org.uk/ethics/
- Confidential Advisory Group (CAG): http://www.hra-decisiontools.org.uk/cag/
- What is research: http://www.hra-decisiontools.org.uk/research/
- Managing consent: http://www.hra-decisiontools.org.uk/consent/

NHS Digital has produced guidance on the information expected in the various sections of a DARS application, named 'Standards'. For each 'Standard' there is written guidance covering important items to consider when writing your application and, where available, an accompanying video.
https://digital.nhs.uk/services/data-access-request-service-dars/dars-guidance

HDRUK
Health Data Research UK

Outline of the end to end data access and use process

In scope

1. **Study proposal development:** requiring identifying likely data sources, their applicability to the study topic, how best to manage/analyse, developing project plan and indicative budget.
2. **Study funding proposal:** usually undertaken prior to seeking approvals, but needs to reflect likely data acquisition costs/fees as well as realistic timescales.
3. **Application development & submission** (for each data source and possible CAG/REC as well): based on study proposal, but detail may need to be tailored to each data provider/approver; providing supporting materials (Data Flow Diagrams, legal approvals/bases, security assurance/documentation, etc.). May include a registration process for new organisations / users and may involve more than one data custodian.
4. **Pre-vetting:** a process carried out by the approval secretariat (or equivalent) that forms have been correctly completed, necessary documentation provided, legal & governance requirements met – this may require the application to be amended/corrected and re-submitted
5. **Actual Approval review:** this may require scheduling for next available committee meeting (and hence depend on overall workload); some processes allow for 'precedent' consideration by the secretariat and/or Chair; decision may be to reject/approve/require amendment usually with reasons and/or conditions to be met, so application may loop back to any of the earlier steps

Out of scope for initial Data Access Management Module development

6. Contract development and signing – standard set of terms for each data provider, potential for legal negotiations. This may give rise to further questions about legal bases, controllers/processors, and even require a return to the very beginning of re-designing the study protocol.
7. Data release: resources available at each data provider to generate the required dataset / provide access to the research environment.
8. Data Analysis: this can show up data quality problems and/or be delayed through lack of resources at the study centre.
9. Data Destruction: Usually a contractual requirement if data release model. Extensions may be allowed but may require access request process again.
10. Publication: may depend on contractual requirements around publication, audit of safe outputs may be required and acknowledgement of data sources in publication.
11. Audit/Monitoring: checking that security is maintained, data held safely and processed appropriately, including deletion as noted above

Cross-sector national approaches

The Office for National Statistics Secure Research Service (SRS) gives accredited or approved researchers secure access to de-identified, unpublished data in order to work on research projects for the public good.

https://www.ons.gov.uk/aboutus/whatwedo/statistics/requestingstatistics/approvedresearcherscheme

The METADAC (Managing Ethico-social, Technical and Administrative issues in Data ACcess) is a multi-agency multi-study data access structure that services several of the UK's major cohort studies and provides a scalable mechanism to incorporate additional cohorts in the future.
https://www.metadac.ac.uk/

International approaches

Data Access Support Hub (DASH): a one-stop shop for requesting access to multi-jurisdictional data across Canada https://dash.hdrn.ca/

Data Use Oversight System: Expediting data access for researchers, by facilitating and enhancing data access committees' workflows https://duos.broadinstitute.org/

From: [REDACTED] On Behalf Of Cabinet Secretary for Health and Sport
Sent: 23 April 2021 09:41
To: Public Engagement Unit
Cc: Cabinet Secretary for Health and Sport
Subject: FW: Progress in UK health data research - April 2021

OR

From: Andrew Morris and Caroline Cake
Sent: 22 April 2021 16:22
To: Freeman J (Jeane), MSP
Subject: Progress in UK health data research - April 2021

Dear Jeane,

Health data research is providing enormous value to people across the UK, made possible by recent advances in data and infrastructure, including those set out in this letter. However, there continue to be significant challenges to overcome, including issues around data linkage and access, gaps in acute hospitalisation data flows and capacity of the system to manage the pace of change. To tackle these challenges and create a legacy from the pandemic response we seek your support to embed collaborative and trustworthy health data science across the NHS, academia, industry, charities and government.

Profound public benefit

The tremendous benefit of health data science through the pandemic is indisputable. HDR UK and our partners have provided 28 detailed reports on health data science to SAGE, and a summary of the research impacts include:

- The NHS DigiTrials Health Data Research Hub and HDR UK led work with the PRINCIPLE trial to accelerate recruitment of participants within 24 hours of a positive SARS CoV-2 result in the community, which enabled the addition of the budesonide treatment arm. This has since become the first widely available, inexpensive drug found to shorten recovery times in the community.
- The world's first real world data on single dose vaccine effectiveness: The BREATHE Health Data Research Hub and the EAVE II project produced a national (5.6m people) data infrastructure which enabled the first whole

country estimates of Astra-Zeneca and Pfizer-BioNTech vaccine effectiveness by vaccine type in different age groups. This was announced by the Prime Minister in February 2021 and influenced national regulatory strategy in the UK, Canada, Denmark, France and Germany (Lancet 2021).

Advances in data and infrastructure

The UK's largest linked health data research asset: the national data infrastructure, enabled by the Data and Connectivity National Core Study is supporting 885 researchers, working on 323 research projects across five national trusted research environments. As part of this and by working in partnership with NHS Digital and the BHF Data Science Centre, we have created a new linked health data resource covering 54.4 million people – over 96% of the English population (BMJ 2021). This is led by the CVD-COVID-UK consortium and is available to UK researchers to collaborate in NHS Digital's new secure research environment. It is the largest research resource in the UK and is being extended to 67 million people through federated working. It is already being actively used by over 50 researchers, with all code fully open and accessible via Github.

UK Health Data Research Hubs: In their first 18 months, the Hubs have made 157 datasets discoverable on the Health Data Research Innovation Gateway, have delivered 300 multi-sector projects, with over 20,000 meaningful patient and public interactions, and 2,300 training activities. Our latest report Improving UK Health Data: Impacts from the Health Data Research Hubs shows how the Hubs have informed UK policy decisions on the effectiveness of COVID-19 vaccines, created tools to improve clinical decision-making in the management of patients with vascular disease, and supported research in cancer, heart disease and hospital care pathways by linking routinely-collected data. We are excited to announce that two new Hubs will join our network next month to support data research in pain and mental health, funded by the MRC.

HDR UK Data Utility Framework: This new framework shows the usefulness of data for research, it has over 100 datasets evaluated against it and is now integrated into the Gateway. Other organisations in the UK, such as NICE and NHSX, are seeking to

adopt the framework, and there is considerable interest internationally from colleagues in the USA, Singapore and Israel. Key papers from the community are in development and consultation, including the data standards green paper and Trusted Research Environment white paper, which will be published in May.

Gateway: We achieved a major step forward in achieving the federated data ecosystem vision, with the launch of cohort discovery research within the Gateway. This enables, for the first time, researchers to search across datasets to find cohorts of patients with specific, defined characteristics; opening up a huge potential for increased discovery, while maintaining safety and security, helping researchers get to impact faster.

This month we reached the milestone of 1,000 researchers registered on the Gateway. With over 11,500 searches each month from around the world and 640 datasets now discoverable, the Gateway is becoming the "go to" place for researchers to discover and request access to UK health datasets with the added benefit of giving much-needed transparency to the UK public on what data is available, how they are used and why.

As highlighted by the National Data Guardian and partners, the public expect more transparency over how health and care data are used and how decisions are made, at a systemic level. Public and patient involvement and engagement, open science and open code are at the heart of HDR UK's developments, including:

- Our consultation with the public on vaccine research prioritisation elicited over 800 responses that demonstrated vaccine safety was the primary public concern and has been used to inform the work of the National Core Studies.
- Over 150 resources available on Github. We are accelerating reproducible science by bringing together repositories of open standards, data and source code, tackling some of the most important challenges in wrangling multi-model data and generating replicable insights.

Next generation of health data scientists

Our Black internship programme, in partnership with the UK Health Data Research Alliance, has 54 interns (60% women) this summer. Not only will this bring diverse

perspectives and skills to our health and science community, it provides an opportunity for young Black people to flourish in STEM careers.

Our ambition to train over 10,000 health data scientists is moving forward at pace. Later in the Spring, we will be unveiling the next stage of this strategy with the launch of an open, virtual learning environment with hundreds of courses to build skills in data science.

We have events scheduled throughout 2021 to engage with the wider health data research community, both in the UK and internationally. These include our joint event with the Government Office for Science on 24 June the National Core Studies for COVID-19 Symposium and our annual Scientific Conference – Data Insights in a Pandemic - on 23 June. You are warmly welcomed to both events.

We greatly value your support in these exciting developments and look forward to our ongoing collaboration and partnership.

Director, HDR UK Chief Executive Officer, HDR UK

HDR UK Stakeholder

Our mailing address is:

Health Data Research UK

Gibbs Building

215 Euston Road

London, London NW1 2BF

United Kingdom