Data Access Management

Alliance Board meeting
29 April 2020

David Seymour, HDR UK Partnership Director

Paola Quattroni, Alliance Delivery Manager

UK Health Data
Research Alliance

The Alliance delivery sub-groups

Data standards and quality
Oversee, coordinate, and produce guidance and policy regarding data quality, standards, dataset publishing, data provenance, ontology / terminology services and use.

Aligning approaches to trusted research environments
Agree on requirements for secure analytic environments where researchers can bring analyses and algorithms to the data.

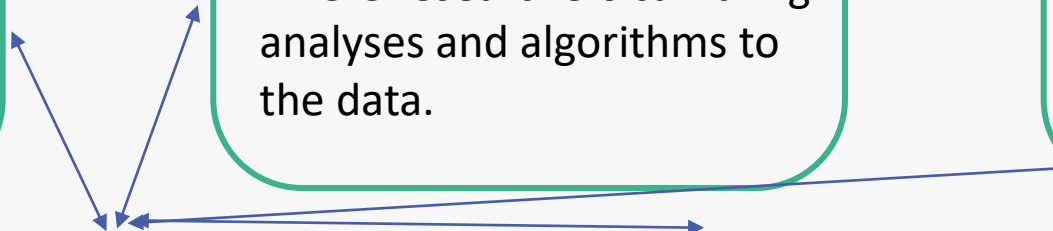**Engaging and involving practitioners, patients and the public**
Understand the interests and values of practitioners, patients and the public in health data.
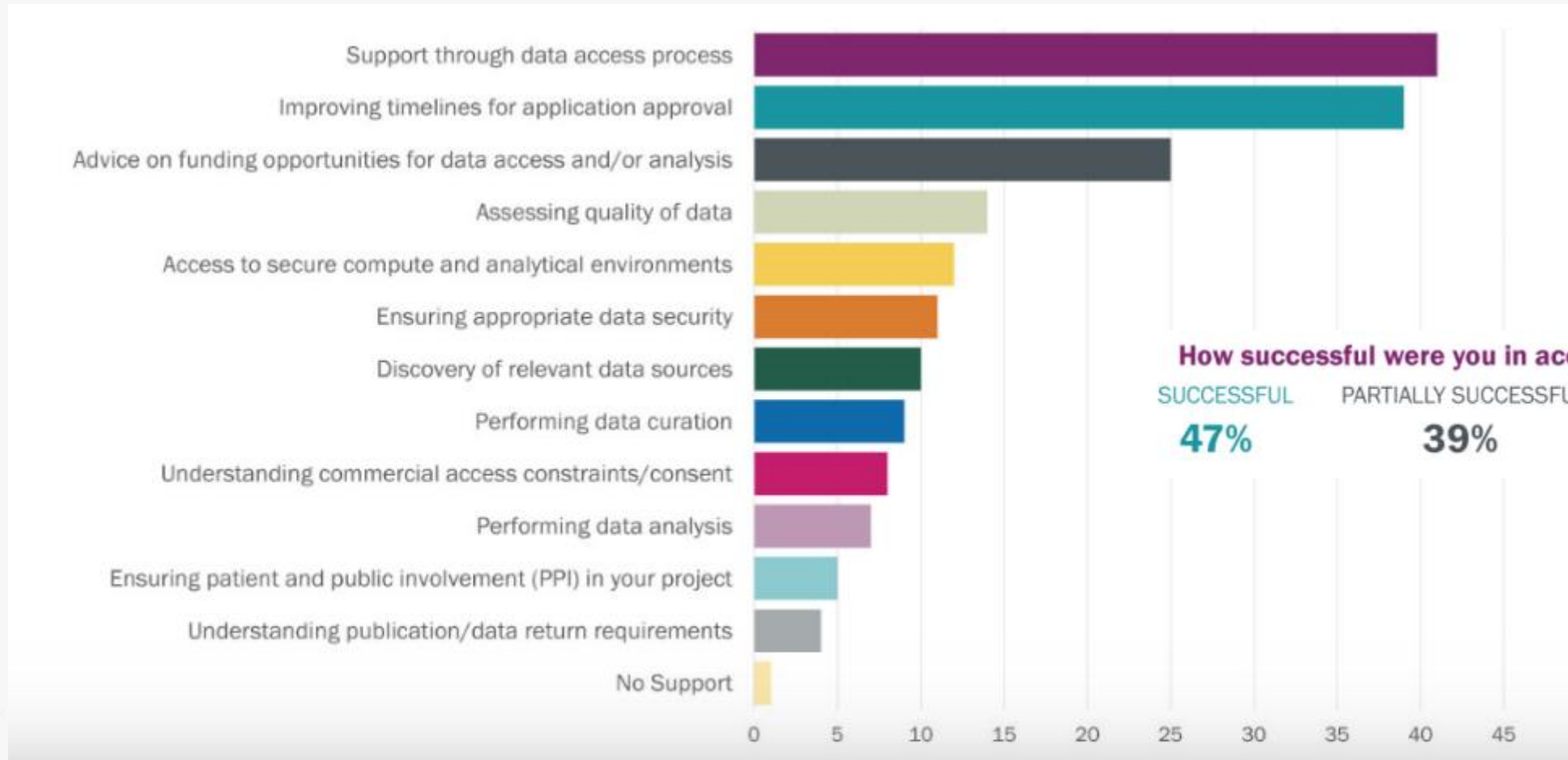
Promoting participation and improving access
Ensure consistency and harmonise principles, definitions, process steps to develop a best practice for assessing requests to access UK health data.

Supporting Health Data Innovation Gateway
A one-stop-shop destination to discover, access, manage and drive the ethical and safe use of data building on existing infrastructure and best practice.

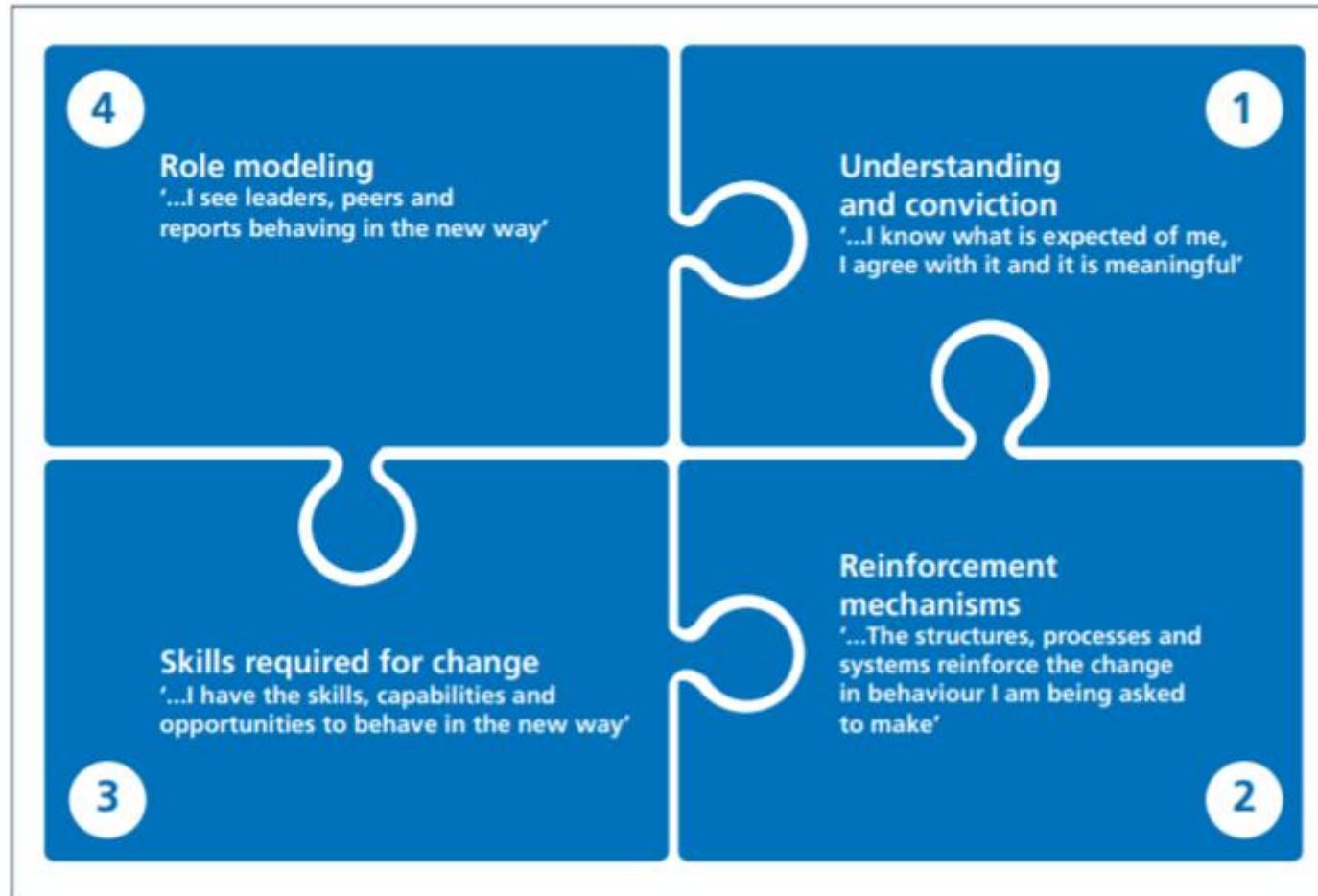# "Considering your needs for accessing health data in future, where do you feel you would need the most support?"

How successful were you in accessing the data?

SUCCESSFUL	PARTIALLY SUCCESSFUL	UNSUCCESSFUL
47%	39%	14%

Improving data access management requires holistic approach to achieve the required change

- RAG Streamlining and Ethics group
- Hubs
- HQIP
- SAIL
- UK Biobank

Communication of benefits

Training Accredited 'safe' User

Innovation Gateway development

4 Role modeling
'...I see leaders, peers and reports behaving in the new way'

1 Understanding and conviction
'...I know what is expected of me, I agree with it and it is meaningful'

3 Skills required for change
'...I have the skills, capabilities and opportunities to behave in the new way'

2 Reinforcement mechanisms
'...The structures, processes and systems reinforce the change in behaviour I am being asked to make'

Keller and Price (2010)

Data Access work will inform development of the Health Data Research Innovation Gateway

- 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**.

- 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**.

Building on existing tools and best practice within UK health data

Health Data Access Toolkit developed by the **MRC Regulatory Support Centre**

- **Health Research Authority** 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'.

And cross-sector and international

- **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.

- 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.

- **Data Access Support Hub (DASH)**: a one-stop shop for requesting access to multi-jurisdictional data across Canada

- **Data Use Oversight System:** Expediting data access for researchers, by facilitating and enhancing data access committees' workflows

Three intended use cases

1. **For data custodians with existing 'in house' web-based workflow solutions** (e.g., NHS Digital, SAIL), the module would provide validated inputs to their approvals processes and receive time stamped progress updates for key process points.

2. **For data custodians with 'MS Word' based documentation and offline workflow** (e.g., HQIP and most other custodians) it would help them harmonise and streamline their approach and provide a scalable, web-based workflow.

3. **For new data custodians** (e.g., some health data research hubs, NHS Trusts, other organisations that have not routinely made data available previously) it would provide the opportunity to move straight to a 'best of breed' web-based access management request solution.

Innovation Gateway Phase 2 with Technology Partner will develop data access management module

High level timeline

From March 2020	HDR UK engaging data custodians, data users, health data research hubs to understand current practice and map end-to-end approval journeys
May 2020	Propose design for an access request form and workflow that could be integrated in the Gateway based on community input
June 2020	Provide our Technology Partner with the proposed requirements resulted from community engagement exercise
June-October 2020	Technology Partners to build Gateway access request module based on community feedback and input received
October 2020	Gateway Phase 2 completed: infrastructure will support workflow management and rules-based pre-validation based on existing best practice management processes

As part of this work we are also engaging the Hubs, Alliance members and potential users.

Main outcomes from first Data Access previous workshop

1. Transparency is key to ensure public trust
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

2. Pre-submission stage and user support are crucial
Understanding what is required, clear communication from data custodians, accreditation, training and guidance are essential to ensure good applications

3. Common project identifiers to link information
Connecting information through different systems (e.g. to pull through relevant approvals) would be extremely helpful to streamline processes

4. Standard Data Sharing agreements and common legal definitions
Uniform data sharing agreements might be helpful to speed up the legal process and negotiations

5. Sharing progress and intelligence across the system
With sufficient commonalities across systems there is scope for the Gateway to connect individual systems in an automated manner and report that progress (including DSPT accredited researchers/breaches)

High level process with Five Safes at its core

Safe projects	Is this use of the data appropriate?
Safe people	Can the users be trusted to use it in an appropriate manner?
Safe settings	Does the access facility limit unauthorised use?
Safe data	Is there a disclosure risk in the data itself?
Safe outputs	Are the statistical results non-disclosive?

Desai, Tanvi; Ritchie, Felix; Welpton, Richard (2016). "Five Safes: designing data access for research". Bristol Business School Working Papers in Economics

Next steps

- Develop a 'wire frame' of the data access request form that will be available through the Gateway that includes:

  - Harmonised set of questions common to all or most data custodians.

  - A customisable component to enable individual data custodians to customise data request forms based on organisation or use case requirements.

- Develop workflows for the two main use cases (linkage to 'in house' process and delivery of 'end to end' process) that includes approach to validation and verification wherever possible

- With our use cases in mind – going forward we would like a group of users from the data custodians community to engage with us and our Technology Partner and accompany us in the process of designing the Gateway module. What level of involvement can you offer?

Paper A: Convenor's Update

14 October 2020

Thank you for your continued support of the UK Health Data Research Alliance. There has been outstanding engagement and leadership from colleagues across the UK, that has built momentum across several priority areas. This update highlights key updates and activities since the July Board meeting, within the context of the Digital Innovation Hub Programme developments.

Announcements

Next UK Health Data Research Alliance Board meeting, 20 January 2021

The next UK Health Data Research Alliance Board meeting is provisionally scheduled for **20 January 2021, 10:00-12:00**. Please save the date and details will be confirmed close to the time.

UK Health Data Research Alliance Symposium, 1st December

Alliance members are invited to attend the second UK Health Data Research Alliance Symposium on 1 December 2020. Please register here: https://www.hdruk.ac.uk/uk-health-data-research-alliance-symposium-december-2020/ . Given the current circumstances the event will be virtual.

The meeting will bring together people from across the UK Health Data research Alliance, Health Data Research Hubs, Innovation Gateway, industry and international players working in health data infrastructure. Attendees will have a chance to learn more about the health data infrastructure in the UK, both the public and the private sectors, and also to hear more about the international landscape and propose new ways to work together. A short update to the Board will be provided by Maria Koufali, Nottingham University Hospitals NHS Trust, and Alliance Symposium Advisory Group Member.

International COVID-19 Alliance Forum, 19 Nov 2020

On 19 November 2020 HDR UK will host a virtual event to showcase some tangible outcomes and achievements of the International COVID-19 Data Research Alliance and Workbench since the launch in June, including plans for driver projects.

Interested UK Alliance members are invited to attend the meeting. A Zoom link and more details will be sent via email in due course.

The International COVID-19 Data Research Alliance is an independent consortium of leading life science, philanthropic and research organisations that was set up to offer a co-ordinated global response to the COVID-19 pandemic. The Workbench is a new data platform that allows scientists around the world to discover, access and analyse multi-dimensional datasets whilst respecting confidentiality and privacy.

For members of the UK Alliance with high value COVID-19 related datasets, we encourage you to consider joining the International COVID-19 Alliance as well.

For more information about the International COVID-19 Alliance and Workbench contact Julia Levy Julia.Levy@hdruk.ac.uk.

New members

We are delighted to welcome two new members of the Alliance this month who signed their letter of intent to join the Alliance. We look forward to their engagement in the Alliance workstreams and to working with them making new data assets discoverable via the Gateway. A public announcement is scheduled in the coming weeks.

- **UK Brain Banks Network –** A coordinated national network of UK brain tissue resources for researchers to use. The Network supplies tissue samples to academic and industry researchers in the UK and abroad and have approval to provide tissue samples to research projects and pilot studies. The banks also work together to agree common standards of operations and harmonise protocols for consent, tissue handling and storage, quality indicators and the application process for access to tissue samples and associated data.
- **Early Detection of Disease Research Project UK –** A major Industrial Strategy Challenge Fund initiative aimed at improving early detection and prevention of major chronic diseases including cancer, metabolic diseases and neurodegenerative diseases. The study plans to recruit a cohort of up to 5 million healthy volunteers linked to routine NHS health records, representative of the whole UK population. The cohort will enable studies developing new tools for disease prevention and diagnosis such as circulating tumour DNA, biomarkers, proteomic panels, metabolomics, imaging tools and digital signatures. It is anticipated that the programme will run over at least the next 20 years as the value of the study will grow as data and samples accrue over time.

National health data research capability to tackle the COVID-19 pandemic

SAGE fortnightly reporting

Reporting to SAGE and the UKRI/DHSC rolling panel takes place on a fortnightly basis in line with SAGE briefings. The latest report can be found here (Paper B.1).

Since last board meeting, SAGE has endorsed five proposed recommendations relating to:
- Coordinated approaches to support and accelerate access to restricted national datasets
- Accelerating insights on long term impacts of health and social care changes during COVID-19 lockdown on major diseases
- Providing insights on understanding high rates of COVID-19 infections in minority groups and care homes.

National Core Studies Programme

Funding for the National Core Studies Programme commissioned by Sir Patrick Vallance to answer the immediate policy and operational needs for the UK response to the pandemic during the winter started on 5 October 2020 and will run for six months to March 2021. It will be formally announced within the next few weeks.

The National Core Studies cover six key areas to enhance our understanding of the SARS-COV-2 virus and the UK's pandemic response at the national level. The purpose of these studies is to accelerate insights essential to support research vital for HMG policy developments and operational decision making.

The National Core Studies include:

- **Epidemiology and Surveillance** led by Professor Ian Diamond (ONS). Collecting and analysing data to inform safe level of restrictions and protection against imminent outbreaks.
- **Clinical Trials Infrastructure** led by Professor Patrick Chinnery (MRC) and Divya Chadha Manek (NIHR). Establishing infrastructure to run large scale trials for Covid-19 drugs and vaccines.
- **Transmission and Environment** led by Professor Andrew Curran (HSE) Understanding and mitigating transmission of the disease in workplace, transport and public places.
- **Immunity** led by Professor Paul Moss (University of Birmingham) Understanding immunity against Covid to inform back-to-work policies.
- **Longitudinal Health** led by Nish Chaturvedi (UCL) Understanding the impact of Covid-19 on long term health to inform the design of mitigating policies.
- **Data and Connectivity** led by Professor Andrew Morris (HDR UK working with ONS) Making UK-wide health and administrative data available for linkage and accessible to catalyse Covid-19 research.

The Data & Connectivity Study underpins all the other five studies and will define the infrastructure, data and services needed to create a robust and trustworthy health and cross-sectoral data capability that will support research in the key areas specified. Our aim is to build upon existing partnerships and capabilities, including the work we have been convening for SAGE. In addition to national bodies from across the four nations of the UK Health Data Research Alliance, the work will also involve the Gateway, as well as Secure Research Service (ONS), and working in partnership with ADR UK.

Paper B.2 provides a more detailed overview of the National Core Studies Programme.

Diversity and Inclusion

Following publication of HDR UK's recent Statement-on-anti-black-discrimination we committed to developing focused actions to improve diversity and inclusion of datasets for research and innovation and of people in the HDR UK community. We are keen to show tangible actions now with two focussed initiatives, the Black Internship programme and the creation of a Data diversity Special Interest Group.

Black internship programme

We propose a programme of work in conjunction with the UK Health Data Research Alliance to create a Black Internship Programme and support growth of health data science leaders. Black data scientists are currently under-represented within the health data science community. We believe we can play a role in providing them with the opportunities and expectations to lead from early in their careers, developing skills needed at key stages and providing tools.

A letter to invite Alliance members to participate in the initiative was sent on 22 September 2020 (Paper C). We received a wealth of positive responses and have now more than 15 participating Alliance members. We look forward to working with members on this exciting initiative.

For more information about the Black Internship Programme, please contact Caroline Cake Caroline.Cake@hdruk.ac.uk.

Data diversity: ethnicity coding

At the last Alliance Board meeting, the Board agreed with the collective ambition to improve the quality of ethnicity coding and use data to help increase the representativeness of research studies. The initial focus on improving understanding of current practices and first questions around collection of representative data and main challenges were discussed at the last Data Officers Group meeting on September 9th.

It was agreed that a Special Interest Group on Data Diversity would be created 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 representativeness of research.

We are now seeking volunteers to join this group, both from the data custodian and user communities.

Innovation Gateway

Overview of progress

The Innovation Gateway development is progressing at pace, with new content onboarded every month. In October, the Gateway reached the 500 datasets milestone, a remarkable achievement from 144 at the beginning of the year. It also has:

- 350 registered users
- 48 Tools
- 517 papers
- 340 Projects, including 190 training courses

Featured collections are also now displayed on the landing page of the Gateway, with each Health Data Research Hub having its own collection, enabling researchers to find all artifacts associated with a specific Hub quickly and easily.

In parallel, the number of searches on the Gateway increases steadily, with over 7,000 searches in the last month.

The Innovation Gateway is nearing the end of the first milestone (31 October 2020), and new functionalities have been delivered, such as:

- **New Data Access Request process** based on the ONS Five Safes. This is currently being tested by some early adopters (both hubs and national Trusted Research Environments as part of the National Core Studies Programme). Their feedback will be taken into account for future versions of the Data Access Request process.
- **Integration with the Phenomics library:** Ability to refine searches using Phenomics where datasets have been tagged.
- **Metadata quality tool:** easy visual representation of the completeness of the metadata information.
- **Technical metadata:** in depth description at variable levels of the dataset.
- **Gap analysis:** highlights searches that do not deliver any results, potentially providing information as to gaps in the market.
- **New dashboards**

In November, the Gateway enters the second part of its delivery plan, Milestone 2, until end of April 2021. During this time, the development team will focus on delivering new functionalities such as

Trusted Research Environment integration, ID federation, production of cohort builder, improvement to the data access request process, and testing all existing and new capabilities with the National Core Studies teams.

Finally, the public and patient engagement strategy is being reviewed and updated building on learnings from Milestone 1.

Access requests workflow

Work is progressing at pace to streamline, improve and accelerate current data access processes and support both researchers and data custodians in facilitating data driven research.

In the past few months we have been working closely with several Alliance members to build a Data Access Request module as part of the Gateway and provide a solution for those data custodians listing datasets in the Gateway and interested in using the Gateway interface to provide access to data for research.

The past month has also seen a speedy increase in engagement with the national Trusted Research Environments/data custodians (NHS Digital, SAIL Databank, Public Health Scotland, Northern Ireland Honest Broker Service and ONS) as part of the National Core Studies Programme. The Programme aims to provide a streamlined access to priority datasets needed to answer questions relating to the five thematic National Core Studies. We have been able to leverage HDR UK's ongoing work to support the National TREs and are now working closely with these organisations to ensure the Gateway module in development is fit for purpose.

Alliance members early adopters and national TREs are committing to adopt the Gateway Data Access module for their processes and are now working closely with us to develop the first iteration of the module.

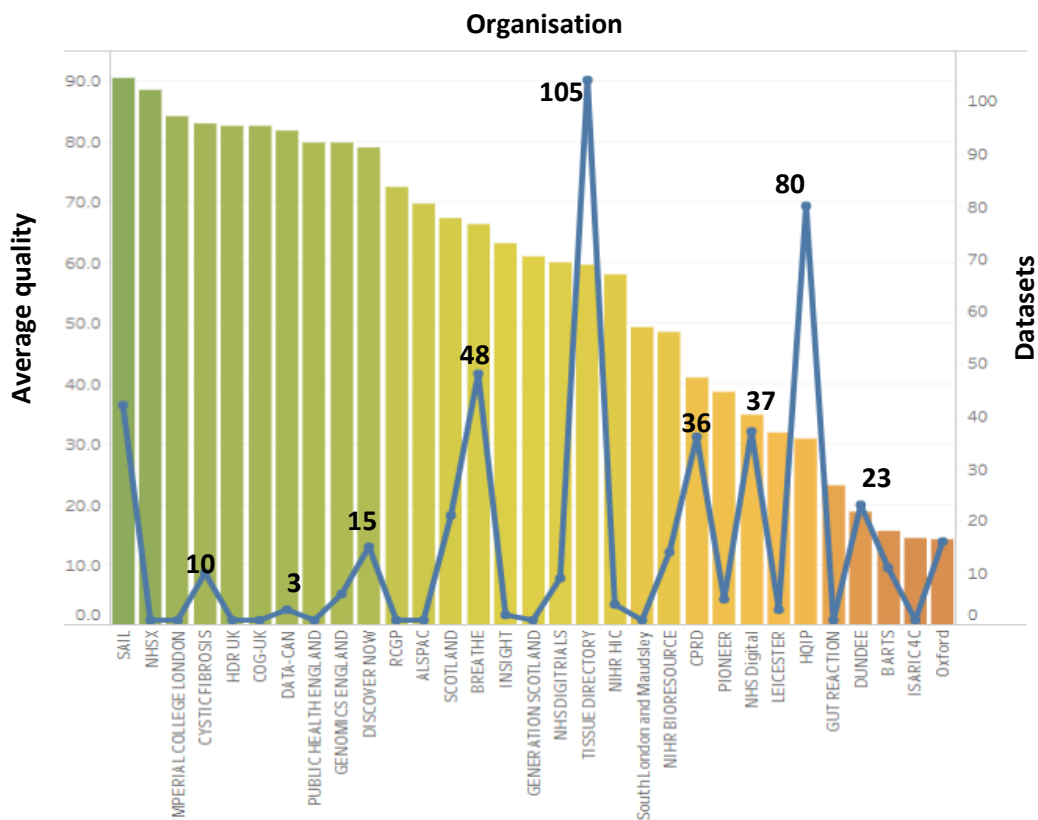To date the following organisations have been involved in this project:

- Healthcare Quality Improvement Partnership, HQIP (Alliance member)
- CPRD (Alliance member – part of initial harmonisation work)
- University Hospitals Birmingham NHS Foundation Trust (Alliance member, INSIGHT and PIONEER)
- Cystic Fibrosis Trust (Alliance member)
- Barts Health NHS Foundation Trusts (Alliance member)
- INSIGHT (Hub)
- Discover-NOW (Hub)
- SAIL Databank (Alliance member, BREATHE hub, National Core Studies TRE)
- NHS Digital (Alliance member, NHS Digitrials and National Core Studies TRE)
- HSCNI (Alliance member, National Core Studies TRE)
- NHS Scotland/Public Health Scotland (Alliance member, National Core Studies TRE)
- Health Research Authority (HRA) – key stakeholders

The first iteration of the data access module for the National Core Studies is expected in early November 2020.

> **ACTIONS FOR MEMBERS**: Please contact Paola.Quattroni@hdruk.ac.uk if you want to be involved.

Metadata Onboarding

The focus of this workstream is making datasets discoverable through the Gateway and improving the quality of the metadata available. MetadataWorks are continuously engaging with Alliance members to onboard new datasets and improve the quality of metadata per each dataset listed in the Gateway. To date, 19/33 Alliance members have listed at least one dataset in the Innovation Gateway (410 datasets) and 7/7 Hubs have listed at least one dataset (83 datasets). Of these, all have summary level metadata and 185 have technical metadata. The graph below shows the average of metadata quality per Alliance member. Alliance Board members are encouraged to review how the metadata for their organisation's datasets are represented on the Gateway and provide feedback.

Figure 1. Metadata quality per Alliance member (as of 10.10.20). 'Metadata quality' is the average quality for all datasets listed by each Alliance member.

> **ACTIONS FOR MEMBERS**: A number of Alliance members have not yet had capacity to engage with the process and are encouraged to do so.

Data Utility

The Data Utility project aims to provide a common means of articulating the usefulness of a dataset for a given purpose. This is intended to support user activity through the Gateway, describe the impact of any data improvement activity by Hubs, and see how learnings can be applied across the UK Alliance. It includes the development of Data Utility Evaluation framework and testing open and open-source tools for data profiling. We thank all Alliance members who have been contributing to this work in the past months. The Data Utility Evaluation framework has now been published on the HDR UK website.

Activities relating to Data Utility and two case studies will be discussed during the meeting.

In addition, a paper on the 'Principles of Data Standards' has been developed in consultation with data officers across the HDR UK community (the Data Officers Group). We received feedback from 50 individuals across more than 30 organisations including patients and the public. The vision of the Principles of Data Standards is to encourage behaviours to improve data usefulness and usability through the provision of clear guidance and recommendations. This will greatly increase the possible impact and scale of research across the Hubs and Alliance.

We are now conducting a survey to understand the data standards landscape. The aim is to get a sense of what data models and supported standards are in use across organisations. We invite all Alliance members to contribute to this short survey here.

Practitioner, Public & Patient Involvement and Engagement (PPPIE)

To support the growing and wide range of involvement and engagement activities across the Institute, we launched our new network, HDR UK Voices. With an initial focus on recruiting members of the public to join the network, it offers anyone who is interested an opportunity to influence and shape our work. By joining the network, HDR UK Voices members will be informed of new opportunities as and when they arise, and they will have a variety of options to choose from to suit their needs. For example, they could help us write engaging information about our work, share their views on challenging and complex topics related to health data or help shape research projects.

The benefit to having an umbrella network of HDR UK Voices is that it can be used to enable other groups and populations such as researchers and practitioners to join and also ensure their voice is heard and incorporated as we develop our work. We will shortly be starting work to draw in practitioners that cover a range of professions and specialisms to sit within our network. We see this group acting as a critical friend to the Institute and would ask Alliance members to consider areas of their work in which the practitioner voice is needed and would have a high level of impact.

Current priorities are focused on running involvement and engagement activities around data access and how best to implement the outputs. Working closely with the HDR UK Public Advisory Board we

are developing a set of recommendations on how patients and the public should be involved in the data access decision making process (we aim to publish Nov 2020). In collaboration with useMYdata members, we are working on ways in which to better communicate work around data access to the public with a first priority being developing a lay explanation of Trusted Research Environments. Moving forward, we will build on this dialogue with patients and the public to consider work around data access/release registers and activities related to the Data and Connectivity programme of work. This includes, but is not limited to, public perceptions around private sector data, moving data across the four nations, how we best communicate research to the public from data access -> insight -> outputs and moving beyond the COPI notices. We will work with other organisations who may already be planning work in these areas to not duplicate work.

To gain a better understanding of current PPIE practices and explore areas for further collaboration across the Alliance members, a survey that will act as a mapping exercise has been sent to your PPIE leads. Where we do not have contacts for your PPIE leads we will email you directly for your guidance in finding the right person to connect with in your organisation. We will hold an Alliance Communications and PPIE Leads meeting in early-mid Nov 2020 and invites will be sent shortly.

ACTIONS FOR MEMBER:

Please encourage your PPIE leads to complete the mapping exercise survey by **Friday 30 October 2020**.

Please share the HDR UK Voices Network with your contacts and encourage them to sign up.

Health Data Research Hubs

The development of the Health Data Research Hubs continues at pace while the impact on the Hubs from the pandemic has been somewhat normalised. The opportunities and initiatives developed over this period are being absorbed into Hub operations, Hub services and sustainability planning.

The Hubs have been instrumental in providing data services to key decision makers during the past three months including:

- Working with Hub consortium members HUMA's to use their remote monitoring service. The Discover-Now/Huma team (in collaboration with NHSx) designed 'Community Virtual Wards' where clinicians could rapidly identify deteriorating patients, react quickly to avoid further complications, and better manage system capacity and improve overall survival rates. The team deployed in two health systems in both primary and secondary care and by using patient reported data from the App were able to connect into linked identifiable data which was then deidentified and made available to Discover-NOW researchers. Within two weeks of launch and at the peak of COVID-19 in London, there were over 150+ patients on the Community Virtual Ward with 96% of patients using the app daily as prescribed. The team estimate the service increased secondary care efficiency by 500% based on HCPs making over 150 calls to patients per day which reduced to 30 calls per day using this remote monitoring service. The Discover-NOW and NHSx team led what is believed to be one of the fastest ever implementations of a digital service in the NHS – going live in just four weeks.

- PIONEER facilitated data collection in four regional NHS hospitals within Birmingham by designing and implementing a fully electronic COVID-19 screening and management system and real-time COVID dashboard. This system guided clinical management while facilitating collection of clinical and biomarker data, by capturing details as to where a particular patient lived, presentation severity and their hospital journey from admission to outcome. The dashboards allowed public health, infection control officers and senior clinicians to visualise the distribution of positive, negative, and suspected cases around hospitals in the city.

The Milestone 2 assessment of Hubs is due in March 2021 and work is well underway to sign-off the approach, timescales, deadlines, activities, panel, assessment criteria and dissemination of the results of this crucial milestone to which Hubs must:

- Evidence that the quality of the datasets has been improved (curated) and that the curated data is discoverable through the Gateway
- Provide publishable enhanced service case studies (e.g. from industry, academia, and NHS) that demonstrate impact and value to researchers and innovators from a range of sectors, and to patients and populations
- Provide evidence that the Hub is continuing to engage and involve patients and the public in a meaningful manner.

UK Health Data Research Alliance Board

Wednesday 29th April 2020, 15:00 – 16:30

Immediate actions for Alliance members:

- Agreement from Alliance members to continue work with HDR UK Metadata Onboarding team to prepare data for loading onto the Innovation Gateway over the next months.
- Alliance members to provide contact names and details of their PPIE representatives. We are also seeking input on what a Practitioner Advisory Group should look like ad what types of activities it should drive. Please contact Sinduja Manohar Sinduja.Manohar@hdruk.ac.uk to provide suggestions or if you would like to be involved in this work.
- We have opened a public consultation with patients, the public, researchers and innovators to seek feedback on a draft green paper outlining approaches to health data access via Trusted Research Environments. We invite Alliance members to provide feedback or share the information about the consultation through their networks. The consultation will remain open until 21 May 2020.
- Alliance Board members are all invited to attend the Health Data Research UK's One Institute Annual Conference on Tuesday 16 June 2020. Registrations via Eventbrite.

Immediate actions for Alliance secretariat:

- Send out latest SAGE prioritisation paper and send regular updates about HDR UK's work on COVID-19.
- Send metadata improvement updates to individual Alliance member organisations.

Next steps for Alliance members and secretariat

- Organise a data utility session at the next Alliance Board meeting. If anyone if is interested to join this group, please contact Ben Gordon (Ben.Gordon@hdruk.ac.uk).
- Organise a demo presentation to show new functionalities of the Innovation Gateway (expected in the summer).

Welcome and apologies – Action points from last meeting and questions on Convenor's Update (Paper A)

Caroline Cake (Chair) welcomed everyone to the meeting and set out the agenda. The action points from last meeting were reviewed and an update was given on new Board members joining the Alliance, on recent developments from the Health Data Research Hubs and the recent work by Alliance delivery sub-groups.

Discussion

COVID-19 and prioritisation of research questions

Caroline Cake presented work on the prioritisation process currently being fed into the government's Scientific Advisory Group for emergencies (SAGE) on a weekly basis. This has the objective to support rapid research using health data into healthcare outcomes associated with COVID-19 and into potential interventions to severity of outcomes.

CPRD shared information on their recent work to expedite the data access request process to make data available for research. All approved studies, including those COVID-19 related, are immediately published on the CPRD website: https://www.cprd.com/protocol-list. It is important to share information about approved research so that we can avoid duplication of efforts.

Need to streamline prioritisation of access requests e.g., Public Health England has a number of datasets that can support research and surveillance and there is a need for greater coordination with many people trying to answer the same or similar questions and duplication of effort.

Scope for the Alliance to improve communication about datasets relevant to COVID-19 available from data custodians, including linked datasets, and to share which research projects have already been approved. Researchers might not be aware of what datasets are made available, thus making this clear through the Alliance would be helpful.

In Scotland, research questions are prioritised to meet and inform the requirements of the Scottish government. Data access requests answering different questions are not expedited and go through the normal data access request route.

Finally, it was noted that in Northern Ireland there is an additional need compared to other areas of the UK, which is the coordination with the Republic of Ireland, both in relation to data modelling and research data questions.

Innovation Gateway

Updates on the Phase 2 of the Health Data Research Innovation Gateway and introduction to the Innovation Gateway data access module were presented by Peggy Barthes-Streit and Paola Quattroni.

The technology partner has been appointed. Next phase of development work is to end of October. Key modules include: cohort discovery and data access request. To be demonstrated at the next Alliance Board meeting.

It was noted that international data sharing would require different considerations that have not been covered to date. It would be helpful to discuss how we manage international data sharing, how we provide data access across borders and/or allow linking of UK data to international data. This is of particular

UK Health Data Research Alliance Gibbs Building, 215 Euston Road, London NW1 2BE
enquiries@hdruk.ac.uk | @HDR_UK | ukhealthdata.org

Health Data Research UK is a limited company registered in England and Wales No: 10887014

importance for data coming from rare populations but also international collaborations such as mother/baby data and tackling COVID-19.

The need for standard terminology or legal definitions was also mentioned and how this is approached internationally is also important.

Attendees showed interest in the data utility and data quality improvement work and noted that mapping common data models and looking at how different data users might define 'quality' is of interest. It was proposed that the next Board meeting could cover data quality and utility in a longer session.

Adam Milward (MetadataWorks) presented an update on the metadata onboarding improvements of the past couple of months. He commended the work of those Alliance members who have significantly improved the quality of metadata in the Gateway and have dedicated resources to this work.

Finally, Rebecca Cosgriff from Cystic Fibrosis Trust shared the recently published work on the first 40 cases of cystic fibrosis patients with COVID-19, showing that clinical course of disease in these patients appear similar to the general population.

Next Alliance Board meeting

The next Alliance Board meeting will take place on 23 July 2020 14:00-16:00 via teleconference (Zoom details to be provided). Please save the date, a calendar invite will follow.

Appendix
Attendees via Zoom

Name	Organisation
Brendan O'Brien	NHS National Services Scotland
Peter Bradly	Public Health England
Janet Valentine	CPRD
David Ford	SAIL Databank, Wales
Chris Packham	Nottinghamshire NHS Foundation Trust
Yvonne Silove	HQIP
Marion Bain	NHS Scotland
Emily Jefferson	Health Informatics Centre, University of Dundee
Chris Hollis	Nottinghamshire Healthcare NHS Foundation Trust
Geoff Huggins	NHS Scotland
Neena Modi	NDAU Imperial College
David Crossman	Scottish Government
Tom Denwood	NHS Digital
Alistair Weir	Guy's and St Thomas NHS Foundation Trust
Kinga Varnai	Oxford University Trust
Kate Northstone	ALSPAC
Rebecca Cosgriff	Cystic Fibrosis Trust
Phil Quinlan	UKCRC Tissue Directory and Coordination Centre
Simon Ball	University Hospitals Birmingham NHS Foundation Trust
Tim Skelton	University Hospitals Leicester NHS Trust
Maria Koufali	Nottinghamshire Healthcare NHS Foundation Trust
Ian Young	HSCNI
Tom Fowler	Genomics England
Neil Sebire	GOSH & HDRUK
Mark Howells	Nottinghamshire Healthcare NHS Foundation Trust
Charlie Davies	Data-CAN
Elizabeth Sapey	PIONEER
Alastair Denniston	INISIGHT
Martin Landray	NHS Digitrials
Axel Heitmueller	Discover-NOW
Aziz Sheikh	BREATHE
Gerry Reilly	HDR UK
Amanda White	HDR UK
Peggy Barthes-Streit	HDR UK
David Seymour	HDR UK

Paola Quattroni	HDR UK
Caroline Cake	HDR UK
Andrew Morris	HDR UK
Susheel Varma	HDR UK
Adam Milford	MetadataWorks

UK Health Data Research Alliance Gibbs Building, 215 Euston Road, London NW1 2BE
enquiries@hdruk.ac.uk | @HDR_UK | ukhealthdata.org

Health Data Research UK is a limited company registered in England and Wales No: 10887014

UK Health Data Research Alliance Board

Wednesday 20th January 10:00-12:00

Immediate actions for Alliance members:

- Express your interest to attend the first **Special Interest Group meeting on Data Diversity on 8 February** (13.00-15.00) by emailing Paola.Quattroni@hdruk.ac.uk
- Continue work with HDR UK Metadata Onboarding team to prepare data for loading and continue to improve the quality of metadata in the Innovation Gateway
- Please let Paola.Quattroni@hdruk.ac.uk if there are any changes to contact details for those included in the current mailing list.

Actions for HDR UK secretariat:

- Confirm Alliance members' contact details
- Provide regular update on SAGE reporting, Data & Connectivity Programme of work, vaccination data availability
- Approach all Alliance members to explore opportunities for adoption of the Five Safe data access form and associated workflow for datasets in the Gateway
- Circulate link to 'Data access processes – public involvement, trust and transparency' survey and invite to special interest group meeting on Data Diversity.

Meeting notes:

Welcome, Action points from last meeting and Update

- Andrew Morris (Chair) welcomed everyone to the meeting and set out the agenda.
- Action points from last meeting were reviewed and an update was given on:
  - New Board members joining the Alliance
  - Developments from the Health Data Research Hubs
  - Alliance delivery sub-groups
  - Patient, public and practitioner involvement

Updates from members

As part of the general updates, representatives from the devolved nations provided a summary of recent work.

- Jackie Gray and Heather Pinches provided an update on behalf of NHS Digital.
  - Work is progressing to create linked NHSD-ONS data assets and make them available for the research community.
  - A Data Alliance Partnership has been convened by NHS-X with the aim to improve collection of data from NHS organisations; improve access and sharing of information; provide better access to data that can be made available in an anonymous fashion.

- o Work with ICO in progress on developing a new Article 40 code of conduct for healthcare and provide guidance building on available resources. The Code of Conduct is being progressed through the Health and Care IG Panel.
  - o Work on getting vaccination data into NHSD is progressing.

- Indra Joshi, NHSX highlighted that access to the National COVID-19 Chest Imaging Database (NCCID) has now been extended to hospitals and universities across the country who are using the images to track patterns and markers of illness. In subsequent discussion, the Buyers Guide to AI in Health and Care was also highlighted as an available resource.

- Alex Newberry provided an update on behalf of Public Health Wales.
  - o Work on legacy contracts for accessing data beyond the pandemic
  - o Consultation in Wales about replacing NHS Informatics to improve data processing
  - o Work with the Department of Health on a new research strategy, where the data element is strong.
  - o Onboarding of vaccination data from the Welsh Immunisation Service.

- David Crossman and Phil Couser provided an update on behalf of Public Health Scotland and Scottish government.
  - o Focus on COVID research and importance of bringing data together to make decisions on management of pandemic.
  - o Research Data Scotland has been created and service will be provided by Public Health Scotland and eDRIS.
  - o Programme of the Scottish government in the last year has seen Artificial Intelligence as a major deliverable.
  - o Consideration on linking work on AI and whether HDR UK could have a role in developing an AI stream.
  - o In relation to data for secondary use and with high volumes of data flowing to the government for analyses, consideration should be given how the public is involved and how to maintain public trust.

Black Internship programme
Tammy Palmer provided an update on the Black Internship programme. Twenty-one organisations are taking part on the programme with approximately 3405 internships being offered. Interviews to identify suitable interns are already taking place and successful applicants will be notified of outcomes the first week of February.

Vaccine research
Janet Valentine (CPRD) provided an overview of vaccine data flows.
- Pharmacovigilance Vaccination data are flowing to MHRA supported by CPRD.
- Partnership with NHSD (as processor) on making three datasets available: vaccination dataset, adverse reaction datasets and SUS raw data (already in NHSD).
- CPRD will have around 25% population coverage linked data.
- Data are expected to be available for the research community by end of March.

- When answering research questions around vaccination, it is important to connect groups and avoid duplication of efforts.

David Seymour (HDR UK) provided an update on recent involvement of Data & Connectivity National Core Study in accelerating vaccine data availability for research.

- Data & Connectivity National Core Study is supporting the UK Vaccines Science Coordination group chaired by Professor Wendy Barclay. The group has identified a set of research questions aimed at addressing topics around effectiveness of vaccines, effect of vaccines on virus transmission, immunity and how dosing schedules affects efficacy of vaccines. To support answers to these research questions a sophisticated data infrastructure is needed.
- Working on definition of minimum dataset at point of care to be recorded.
- ONS is receiving data feeds directly by NHS England. Both NHSD and ONS are working towards a joint controllership to enable access to this data for research.
- Major efforts to accelerate vaccination data flows are ongoing in all devolved nations.

COVID-19 National Core Studies – Data & Connectivity

- Work driven by the National Core Study programme and the of Data & Connectivity Study is continuing at pace.
- Twelve project teams have been funded through a rapid funding call to answer priority questions driven by use of priority datasets made available through the Innovation Gateway. All project teams are expected to submit their first data access request through the Gateway by January 31st.
- The Board acknowledged the need of addressing the fragmentation of work leveraging cross-sectoral opportunities. We have an opportunity to unify strands across core studies and propagate the conversation to non-core studies.
- It was noted that studies not currently part of the National programme are still of great importance and it will be interesting to see how these will fit within the National Core Studies work.
- The latest COVID-19 Round-up can be accessed here.
- The latest SAGE report (26 Jan 2021) can be found here.

Innovation Gateway

- Peggy Barthes-Streit presented recent updates on development of the Innovation Gateway and data access work.
- The first version of the Five Safe data access application form is now available for NHS Digital, SAIL, Databank, Public Health Scotland, Health and Social Care Northern Ireland, Office for National Statistics, and INSIGHT. This is currently being used by the data custodians to provide access to COVID-19 related datasets as part of the National Core Study programme as well as to INSIGHT datasets.
- The HDR UK team will be approaching all Alliance members to explore options to replace their current enquiry form with the new Five Safe data access request form.

UK Health Data Research Alliance Gibbs Building, 215 Euston Road, London NW1 2BE
enquiries@hdruk.ac.uk | @HDR_UK | ukhealthdata.org

Health Data Research UK is a limited company registered in England and Wales No: 10887014

Data Access Registers

David Seymour and Sinduja Manohar introduced a project aimed at improving transparency on how health data is used for research and innovation across the UK. A paper outlining 'data access volumes and estimates' was circulated before the meeting.

Alliance members are now requested to:

1. Provide feedback on the estimates outlined in the paper.
2. Contribute to the list of public data access registers where their organisation is not listed.
3. Continue to support the onboarding of metadata for any datasets made available for research by their organisation.
4. Support the continued development of the gateway as a common data access request portal by adopting the Five Safes form and associated workflow.
5. Engage in the work being undertaken on Data Access Registers to increase the number and accessibility of them and improve how the information is presented to a lay audience, including the role of the Gateway in improving transparency of data access

The Board welcomed this work and agreed that improving transparency on how health data is accessed and used is important for the public, researchers and funders.

The Board proposed to focus on
- Best practice and standard way of communicating how data is used
- Linking outputs to application and tracking
- Identifying a minimum set of information available

The Board cautioned that maintenance of registers, responsibilities and resources are aspects to consider when planning this work.

Next Alliance Board meetings

The next Alliance Board meetings will take place via Zoom on:

- 20 April 14.00-16.00
- 14 July 14.00-16.00
- 20 Oct 10.00-12.00

Please save the dates, a calendar invite will follow.

UK Health Data Research Alliance Gibbs Building, 215 Euston Road, London NW1 2BE
enquiries@hdruk.ac.uk | @HDR_UK | ukhealthdata.org

Health Data Research UK is a limited company registered in England and Wales No: 10887014

Appendix
Attendees via Zoom

Name	Organisation
David Jenkison	Brain Tumour Charity
Chris Orton	SAIL Databank
Brendan O'Brien	NHS National Services Scotland
Matt Cooper	Data-CAN
Mark Beggs	NCIMI
Andy Roddam	Our Future Health
David Crossman	Scottish Government
Janet Valentine	CPRD
Colin Smith	UK Brain Bank Network
Mark Howells	Nottinghamshire Healthcare NHS Foundation Trust
Nicola Armstrong	HSCNI
Phil Quinlan	Tissue Directory Coordination Centre
David Ford	SAIL DataBank
Hilary Fanning	University Hospitals Birmingham NHS Foundation Trust
Sneha Anad	University of Oxford, RCGP Research and Surveillance Centre
Neena Modi	NDAU Imperial College
Indra Joshi	NHSX
Suzie Gallier	PIONEER
Julia Hippisley-Cox	QResearch
Stephen Lock	NIHR Clinical Research Network
Tim Skelton	University Hospitals Leicester NHS Foundation Trust
Eleanor Hall	Gut Reaction
Mark Howells	Nottinghamshire Healthcare NHS Foundation Trust
Alex Newberry	Wales Government
Yvonne Silove	HQIP
Heather Pinches	NHS Digital
Steven Casson	Public Health England
Shekha Modhw	UHB
Alastair Denniston	INSIGHT
Jon Smart	Swansea University
Shaun Cochrane	Guy's and St Thomas Hospital
Phillip Courser	Public Health Scotland
Nic Timpson	ALSPAC
David Porteous	Generation Scotland
Jackie Gray	NHS Digital
[Redacted]	[Redacted]

Alastair Denniston	Insight
Mark Caulfield	Genomics England
Joel Backman	HDR UK
Cathie Sudlow	HDR UK
Andrew Morris	HDR UK
Kay Snowley	HDR UK
Caroline Cake	HDR UK
David Seymour	HDR UK
Gerry Reilly	HDR UK
Peggy Barthes-Streit	HDR UK
Ben Gordon	HDR UK
Sinduja Manohar	HDR UK
David Seymour	HDR UK
Paola Quattroni	HDR UK
Tammy Palmer	HDR UK
Clara Fennessy	HDR UK
Richard Edwards	PA Consulting

UK Health Data Research Alliance Gibbs Building, 215 Euston Road, London NW1 2BE
enquiries@hdruk.ac.uk | @HDR_UK | ukhealthdata.org

Health Data Research UK is a limited company registered in England and Wales No: 10887014

Why we need to work on Data Access Registers

- Our Alliance's Letter of Intent highlights transparency of governance and operations as one of its founding principles. It sets out that one way in which members will demonstrate the principle of transparency is by (amongst other things) publishing a register of projects accessing the data under their custodianship.

- There is a big opportunity, through data access registers to:
    - generate insight into data use and data access
    - build public trust and even advocates for the access and use of data

- Publicly available information related to approved data access requests varies significantly. We estimate that around 50% of Alliance members appear to have a public access request register and that these:
    - Are presented in a range of formats (spreadsheets, word documents, embedded on webpage)
    - Have different information on approved requests

- To fulfil our commitment to transparency we need to increase the number, quality and accessibility of public data access registers.

- Public and Patient Involvement and Engagement has the potential to shape best practice for data access registers.

- The Gateway offers a unique opportunity to bring together information about data access that is initiated via the Gateway and requests made directly to the custodian.

- This work picks up from where we left off just before the pandemic (January 2020 board meeting)

What do patients and the public want to see?

There were a number of requirements that they believe were necessary to ensure public accountability and transparency.

What research is using their data and how to find out more about the research

Accessible format

Easy to find and share with others

Consistent approach

Confidence, reliability, trust in the data/information accuracy and authenticity

Next steps to move this forward…

1. Alliance members will be sent a simple survey to gather information on existing practices – survey will be sent w/c 25 Jan 2021

2. Alliance members to share any existing publicly available data access registers (provide links via the survey) – we propose to link through to them from Alliance member collections on the Gateway (where established) – **March 2021**

3. Run a series of workshops with data custodians and researchers to understand their requirements and any concerns – **beginning of March 2021**

4. Use workshops to determine how we can work together to 'close the loop' with Data Access Registers and role of Gateway. Ideas include:
   - Co-creating a standard for Data Access Registers
   - Exploring ways to increase transparency through the Gateway

Paper A: Convenor's Update

29 April 2020

Thank you for your continued support of the UK Health Data Research Alliance during these extraordinary and difficult circumstances. There has been outstanding engagement and leadership from colleagues across the UK, that has built momentum across several priority areas. This update highlights key activities and achievements since the January board meeting, within the context of the Digital Innovation Hub Programme developments. The last six weeks have inevitably been dominated by the COVID-19 pandemic and the activity across Alliance has been pivoted to meet this global emergency.

National health data research capability to support COVID-19 research questions

On 14 April the Scientific Group for Emergencies (SAGE) endorsed all the requirements of an approach to establishing a legally compliant approach to link data and provide access to secure analytical environments for researchers to answer rapid COVID-10 related research questions, across the four nations, using cardiovascular disease as an exemplar.

The approach was developed by NHS Digital (NHSD), Health Data Research UK, National Institute for Cardiovascular Outcomes Research (NICOR) and the British Heart Foundation (BHF) and is generalisable to other conditions. It will be subject to review after the COVID-19 emergency, or after six months, whichever comes first.

The specific requirements were: i) this approach should be operational over the next 1-2 weeks ii) government should ensure that key policy and research questions can be addressed via this approach iii) data custodians need to make their data rapidly accessible through this route iv) funders must facilitate the scale-up and operation of this approach.

There are three parts to the approach: A. Research Question Funnel, B. Linked data via Trusted Research Environments, C. Information Governance and Access. Full details are available in Paper B.

From 20 April the prioritisation process is being fed into the government's Scientific Advisory Group for Emergencies (SAGE) on a weekly basis, with the objective to support rapid research using health data into healthcare outcomes associated with COVID-19 and into potential interventions to reduce the severity of those outcomes. The latest report to SAGE can be accessed here (Paper C).

Further details of the approach, including how to submit research questions or volunteer capabilities to support research priorities can be found on the HDR UK COVID-19 webpage https://www.hdruk.ac.uk/covid-19/. The latest developments will be covered as the main agenda item.

Health Data Research Hubs

In the last six weeks, the Health Data Research Hubs shifted most of their priorities towards COVID-19 related activities. Below we provide a summary of activities for each Hub:

- **BREATHE** have pivoted as much effort as possible to addressing the considerable national and global challenge posed by COVID-19 and have prioritised all work that aims to contribute to management and mitigation of the spread and impact of COVID-19.

- **PIONEER** are heavily involved in DECOVID, a national COVID-19 dataset, with partners including University Hospitals Birmingham NHS Foundation Trust (UHB), University College London Hospitals NHS Foundation Trust (UCLH), Kings College Hospital NHS Foundation Trust (KCH) and The Alan Turing Institute (ATI). Once the dataset has been curated, this will be used both clinically and academically to gain insights into the pandemic and support clinical and government decision making.

- **Digi-Trials** are pivoting to support COVID-19 clinical trials. This includes support for making Outcome data available to the large number of requests now being received. As an example, Digi-Trials are assisting The Recovery Trial being run by the University of Oxford by providing patient outcomes data from the Secondary Uses Service (SUS+) and from other source for the cohort of patients in the trial.

- **Gut Reaction**'s immediate focus is on linkage of NIHR BioResource (phenotype & genomics), PHE and HES data to answer COVID-19 related research questions. This is particularly important for the IBD BioResource and Gut Reaction Hub, where the medications participants are taking may influence the course of COVID-19 infection.

- **DATA-CAN** have developed core services and are in various negotiations to understand how they could be utilised for COVID-19 activities. They are focusing on leading programmes of work to look at the impact of COVID-19 on cancer diagnosis, treatment and outcomes. It is anticipated that DATA-CAN may become a key disease specialist by supporting activities in the SAGE prioritisation process.

- **Discover-NOW** is keen to support the research challenges presented by the COVID-19 crisis, and are highlighting the opportunity for supporting retrospective, feasibility and prospective studies through various channels with and with partners.

- **INSIGHT** can contribute through health data modelling to rebuilding NHS services and supporting patient care in the post-COVID-19 landscape. In conjunction with Hub partners and the resource agreed and drawn from Roche and AAAMD, INSIGHT are in the early stages of exploring the potential of real-time data to inform advancement of Oculomics research in understanding COVID-19, and in parallel to serve future NHS operational planning to aide getting services back up and running in the period after the immediate crisis.

Health Data Research Innovation Gateway

Minimum Viable Product (MVP)

The Innovation Gateway MVP continues to be accessed by users around the world, who are discovering the metadata from over 400 datasets hosted on the Gateway. This metadata continues to improve as a result of engagement with the data custodians and publishers of the metadata. Further work is taking place across both the Alliance and Hubs to increase the number of datasets discoverable through the Gateway. The functionality of the MVP continues to improve as the development team have been responding to user feedback. Work in the coming weeks will focus on improving the filtering and sorting of search results, as well as enabling additional means of logging in.

To date, 412 datasets are currently listed in the Gateway, including the COVID19 Symptom Tracker App data, the first COVID-19 relevant dataset to be added. In the coming weeks we aim to onboard more metadata from COVID-19 relevant datasets.

Technology Partnership

Phase 2 of the Innovation Gateway is being developed through a technology partnership. Thank you to colleagues from NHS Digital, NHSX, Public Health England, NHS Scotland, the Office of the National Data Guardian and the HDR UK Public Advisory Board who helped with the selection. Three potential partners undertook an 8-week rapid development task (RDT) as the final stage of the procurement process. The RDT exercise started on 27th January and completed on 2nd April with a final panel interview.

A Technology Partner has been now chosen and will be announced publicly in early May. The Phase 2 specification is currently being developed and will be publicly available once finalised in conjunction with the technology partner.

The components that need to be delivered by the first milestone of the second phase of the Gateway delivery (31 Oct 2020) include:

- A data access request workflow and dashboard
- Integration of user collaboration tooling
- MVP of cohort discovery

UK Health Data Research Alliance

New members

We were delighted to announce publicly three new members of the Alliance at the beginning of April who had signed their letter of intent to join the Alliance. We look forward to their engagement in the Alliance workstreams and to working with them on onboarding metadata for their data assets.

- **Neonatal Data Analysis Unit (NDAU, Imperial College London)**

The NDAU at Imperial College holds the National Neonatal Research Database, a unique resource containing data from over 600,000 babies and 7 million days of care for these babies across all the neonatal units in England and Wales. The NNRD is a very valuable international resource and access to this data has the potential to facilitate research to improve prevention, treatment, and delivery of health services for mothers and babies.

- **Health Informatics Centre (HIC), University of Dundee**

The HIC was the first centre in Scotland to offer a Safe Haven and has been operational for over 10 years. The Safe Haven allows secure collaborative research using sensitive electronic health records. HIC maintains a clinical data repository covering approximately 20 % of the Scottish population, combining routinely collected datasets for the Tayside and Fife population

- **Oxford University Hospitals (OUH) NHS Foundation Trust**

OUH is a renowned centre of clinical excellence and one of the largest NHS teaching trusts in the UK, bringing a wide variety of datasets to the Gateway, and covering areas ranging from emergency and routine care to microbiology and pathology.

The following organisation has also signed their letter of intent but not yet been publicly announced:

- NIHR Clinical Research Network

Alliance Workstreams

Promoting participation and improving access

Work is underway on the priority of improving access to health data by addressing some of the difficulties that exist in the UK.

As well supporting both researchers and data custodians in facilitating access and streamlining current processes, we are working to provide a solution for some data custodians through the development of the Health Innovation Gateway.

As part of Phase 2 of the Gateway development, an Access Management module will be developed that enables users to request access to datasets, submit required information and track the progress of applications directly. For data custodians with existing 'in house' solutions, the Access Management module will 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.

We have started by building on the work to date by the NHS Digital Research Advisory Group (RAG) Streamlining and Ethics sub-group. Initial members of this workstream have provided crucial input to inform development of this module and define technology requirements. Thank you to those who already participated in the first two workshops on 2nd and 29th April. The main highlights from the first workshop can be found in this Report.

We are still working with the health data research community including data users, data custodians, health data research hubs, regulators and public and patients, to understand current processes and identify areas of improvement. We encourage all Alliance members to engage with us and the Gateway Technology Partner and provide input at each stage of the Data Access Management module development.

A general overview of the project can be found in paper D and this will be covered as an agenda item.

Data Standards & Quality

This workstream is currently focused on the Data Utility project which aims to provide a common means of articulating the usefulness of a dataset for a given purpose, in order to support user activity through the Gateway, describe the impact of any curation activity by Hubs and to support effective targeting of any future investment around "improving the data".

There is an engaged community to drive this work through the Hub Data Officers Group, which is chaired by Monica Jones (DATA-CAN Chief Data Officer). This group supports several concurrent workstreams, including evaluating data utility, data standards and metadata. It is worth noting that across the central team ad each of the Hubs and Alliance members, work responding to COVID has impacted the ability to devote time to data utility activities. We provide an outline of this project in paper E.

Practitioner, Public & Patient Involvement and Engagement (PPPIE)

Since the last meeting, we have created an aligned strategy and delivery plan for communications, engagement and involvement for HDR UK, which incorporates the key areas of focus for this Alliance workstream. Recent developments include the establishment of a new COVID-19 patient/public advisory group, with over 60 participants, who are available and interested in providing advice and input on data research for coronavirus. HDR UK can provide a co-ordinating role to any Alliance members that wish to consult with this group and we aim to do this via our Slack channel.

Other work has included supporting the Hubs in recruiting and involving lay representatives. We have seen some developments in the Hub PPIE strategies, with particular reference to the One London citizen engagement work, which provides insights about the trade-offs that the public are willing to take relating to health data use. Recommendations from this are yet to be published, but will inform the work of Discover-NOW and other Hubs including INSIGHT.

Through desk research and working with various members of the HDR UK community, we are looking to identify the key questions and grey areas in which PPPIE can have an impact in addressing.

At HDR UK we are boosting input from patients, the public and practitioners through the following routes:

- Recruitment of three new Public Advisory Board members, representing Scotland and Northern Ireland

- Recruitment of four new lay members to join our two new delivery groups
- Early stages of developing a Practitioner Advisory Board as we feel practitioners are a particularly important group to engage in the discussions.

Questions and input from Alliance members on this workstream are:

- To support the research question prioritisation process that we are reporting to SAGE, we require input from Alliance members' established PPPIE groups. We are working with your communications leads to get this started, but require the help of Board members to mobilise these teams so that we can ensure patients and the public are shaping the way that data is made available in the pandemic response
- We will be seeking input from Alliance members on what the Practitioner Advisory Group should look like and what type of activities it should drive.

Metadata Onboarding *(formally 'Supporting Health Data Research Innovation Gateway development and launch' workstream)*

The focus of this workstream is making datasets discoverable through the Gateway, and improving the quality of the metadata available in the Gateway. An MVP metadata standard has been developed and used for this work, and this is being revised based on the outputs of the first phase of work where the metadata of the 400+ datasets provided a valuable picture of the current state of metadata across the Alliance. An improvement plan at attribute level is being implemented by MetadataWorks and Parity and feedback is continuously gathered from all those participating in the work to date. Many of the Alliance members who engaged in the metadata onboarding process took advantage of the financial investment provided in the past months.

Alliance members who have not yet had capacity to engage with the process are encouraged to do so, with a particular focus on COVID-19 relevant datasets. Alliance Board members are encouraged to review how the metadata for their organisation's datasets are represented on the Gateway and provide feedback.

Aligning approach to Trusted Research Environments

On 12th March 2020 we hosted a workshop to discuss approaches to Trusted Research Environments (TREs) and brought together a community of experts interested in improving the use of data for research in a safe and ethical way. Representatives from the data custodian community, TRE providers, HDR UK Public Advisory Board, and funders attended the workshop either in person or via Zoom videoconferencing.

During interactive breakout sessions, participants discussed three main topic areas:

- Ideal requirements for TREs
- Approaches to federation across TREs
- TRE accreditation

The intention was to stimulate discussions on the importance of TREs and to establish a working group to collaborate on the development of the Alliance TRE workstream.

The outputs from the workshop have been incorporated in a draft green paper which we are sharing widely for consultation and is available here (Paper F). We invite the Alliance Board to provide input directly in the Google Doc shared.

A workshop report can also be accessed via this link.

Alliance Symposium Tuesday 4th February 2020

On February 4th we held the first UK Health Data Research Alliance Symposium which brought together more than 150 people across the UK Health Data Research Alliance, Health Data research Hubs, Innovation Gateway, sandbox and others working closely with HDR UK in infrastructure. We had the opportunity to showcase achievements to date and ongoing work across the Digital Innovation Hubs programme. Attendees had a chance to learn more about elements of HDR UK strategy, share feedback on challenges and opportunities and propose new ways to work together.

The keynote speaker was Jeni Tennison, the Chief Executive of the Open Data Institute, who shared a cross-sector perspective on trustworthy access to data.

For those who were unable to attend, we livestreamed the plenary sessions and collated key outputs from discussions during the breakouts and World Café sessions. Slides and outputs from the event can be found here.

Upcoming: Health Data Research UK's 'One Institute' Annual Conference, 16th June

Alliance members are invited to attend Health Data Research UK's 'One Institute' Annual Conference on Tuesday 16 June 2020. Given the current coronavirus (COVID-19) situation, the event will be virtual. We shall also use it as an opportunity to focus on the COVID-19 Taskforce, and the research that is being led by colleagues across the HDR UK family.

The meeting is for the whole HDR UK community and is our annual opportunity to come together and to celebrate our achievements. The event will bring together all staff from across the HDR UK regions: our research teams, Fellows, Alliance members, Hub teams and Gateway delivery partners, along with our Public Advisory Board and public and patient representatives, and colleagues from our key partnership organisations.

This year, we will also be joined for the whole day by our distinguished HDR UK International Advisory Board, including the Chair Dr Amy Abernethy, Deputy Commissioner of the FDA in the US. Amy will be participating in a fireside chat, focusing on how data and technology will influence the future of clinical and regulatory science.

Register via Eventbrite.

Paper B:

A national health data research capability to support COVID-19 research questions

15 April 2020

There is considerable need for COVID-19 research questions to be rapidly answered to guide national (and international) decision making. The four nations of the UK need an approach to enable rapid research at scale and utilising existing research assets, talents and expertise across academia, NHS, government, charities and industry.

This paper proposes a legally compliant approach to link data and provide access to secure analytical environments for researchers to answer rapid COVID-10 related research questions, across the four nations, using cardiovascular disease as an exemplar. This has been developed by NHS Digital (NHSD), Health Data Research UK, National Institute for Cardiovascular Outcomes Research (NICOR) and the British Heart Foundation (BHF) and is generalisable to other conditions. It will be subject to review after the COVID-19 emergency, or after six months, whichever comes first.

For this to make the greatest contribution:

    i)   **this approach should be operational over the next 1-2 weeks**

    ii)   **government ensure that key policy and research questions can be addressed via this approach**

    iii)   **data custodians need to make their data rapidly accessible through this route**

    iv)   **funders must facilitate the scale-up and operation of this approach**

SAGE discussed this on 14 April 2020, and endorsed all these requirements.

Objectives and measures of success

In this unique environment, there is a pressing need to address both service delivery issues and research to improve understanding and treatment. This approach aims to support rapid research using health data into healthcare outcomes associated with COVID-19 and into potential interventions to reduce the severity of those outcomes. It aims to be scalable, to represent all four nations of the UK, to provide safe and trustworthy access, and to avoid distracting the operational activities of the NHS.

Rapid research using health data requires:

- Good research questions, whose answers will rapidly benefit current clinical care and the public's health
- Skills in the rapid design, analysis and interpretation of data-driven studies (including data scientists, data stewards, research software engineers and systems engineers)
- Large scale data, ideally from all four nations, to enable statistically well-powered studies (and to identify regional differences)
- Pseudonymised datasets linked at an individual level to analyse patterns and test hypotheses[1]
- An analysis environment with appropriate tools for data management and analysis, including efficient ability to rapidly bring in new software representing new analysis modalities, and also including the full suite of modern data science tools including a wide variety of R and Python libraries, within which researchers can address high priority questions effectively and efficiently
- A high throughput computing environment with the scale and power to bring modern AI and Data Science methodologies to bear, including GPU computing
- Ability, through open standards and open APIs, to link with international efforts of open research and data sharing, that harness data science to combat COVID-19
- An access and authentication process, supported by logging, that allows large numbers[2] of approved researchers to access data safely (safe data, safe projects, safe people, safe settings, safe outputs) [3]
- Furthermore, the information governance will need to be in place to ensure transparency and fair processing that complies with legal gateways, duty of confidence and current best practice (including the application of opt-out where required), research ethics and data minimisation.

The success of this approach will be judged by the speed and quality of the answers to priority research questions. The scalability of the approach in terms of number of researchers and research questions that can be concurrently addressed, along with public perception of trustworthiness will be additional measures of success.

Cardiovascular disease example

People with cardiovascular disease are one of the groups most likely to be directly and indirectly adversely affected by COVID-19. For example, data from hospitals across England show that the number of people seen in hospital with a suspected heart attack has halved since the beginning of March, raising concerns that people may be at greater risk of suffering long term heart damage, needing intensive care, or even

---

[1] Although this data will be pseudonymised, it will remain identifiable and sensitive. As such it will need information governance appropriate to identifiable data.

[2] 100s to 1000s

[3] Aligned with UK Health Data Research Alliance Principles for Participation for data custodians

dying as a result.[4] In the light of this, the clinical cardiovascular community, through Professional Societies and NICOR, proposed collaboration with NHSD/X and PHE to create a linked dataset without IG barriers.

In the last few weeks, highly effective collaboration has been established with a first imperative to focus on Acute Coronary Syndromes based on SUS+/HES, PDS/ONS, and NICOR datasets to describe presentation and deaths. The British Heart Foundation have brought together the BHF Data Science Centre, HDR UK, other academic stakeholders, to develop further the key research questions and to ensure the best access to these combined NHS datasets for research. These research questions are provided in Appendix 1.

Proposed approach

The proposed approach is condition agnostic. It uses cardiovascular disease as an exemplar condition, but it is fully generalisable to other data sets and conditions such as cancer, diabetes and mental health.

There are three parts to the approach, A. Research Question Funnel, B. Linked data via Trusted Research Environments, C. Information Governance and Access.

A. Research Question Funnel

The best research questions might come from clinicians observing what's happening at the front line, life science companies working on treatments for similar viruses, epidemiologists observing implications for patients with cardiovascular disease, or others. We have created a Darwinian question funnel that helps the 'fittest' questions to be answered fastest:

- Priority being given to those commissioned by CSA/CMO, and through SAGE
- Supported by a simple, open route for anyone to pose research questions HDR UK COVID-19 Knowledge + Skills Matchmaker
- A weekly, objective and transparent prioritisation process to identify which research questions need to be answered quickest: the NIHR/HDR UK prioritisation process with panel members drawn from across the four nations.

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

By linking de-identified data together within each jurisdiction and by making it accessible in trusted research environments, many questions can be answered by different researchers, using the same datasets. Note the proposal is not to link datasets across jurisdictions.

To respond rapidly we propose utilising existing trusted research environments, which have established information governance processes. These are:

Location*	Trusted Third Party	Trusted Research Environment
England	NHSD	Data Processing Services (DPS)

---

[4] https://www.bhf.org.uk/what-we-do/news-from-the-bhf/news-archive/2020/april/drop-in-heart-attack-patients-amidst-coronavirus-outbreak
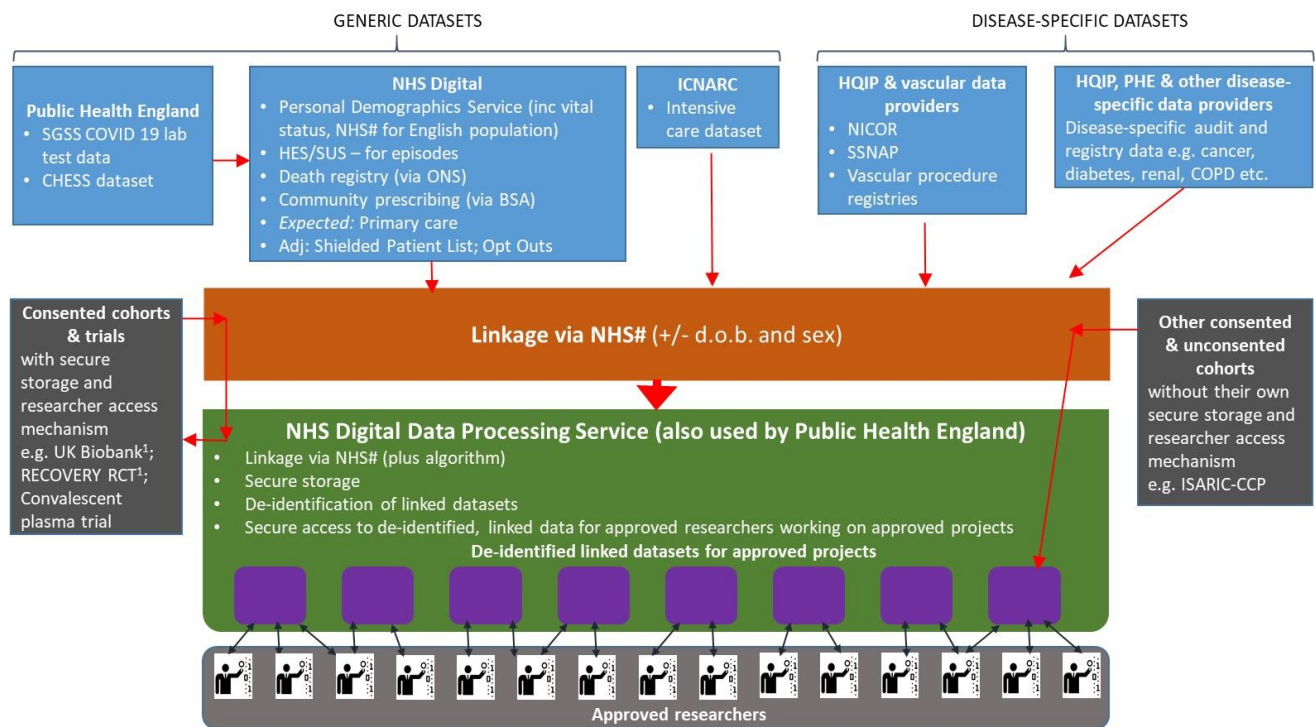
Wales & NI	NHS Wales Informatics Service (NWIS)	Secure Anonymised Information Linkage (SAIL) databank
Scotland	Edinburgh Parallel Computing Centre (EPCC) on behalf of Public Health Scotland	Scottish National Data Safe Haven
Northern Ireland	Honest Broker Service	Business Services Organisation (BSO)

This is where the TRE is located; it does not mean that the TRE can only host data from that geography

These Trusted Research Environments would host and link datasets (within their jurisdiction) and provide access to de-identified linked data to approved researchers. The following figure sets out the generic datasets and the disease specific datasets in the England TRE:

Figure 1. Linkage and secure access for approved research to nationally-collated, generic and disease-specific datasets (cardiovascular and other) in England[5]

Similar arrangements developed for linkage of key datasets in Scotland, Wales & Northern Ireland are described in Appendix 2.

This approach is open to other trusted research environments, and we would expect others to participate once this approach is established, for example 'enhanced cohorts' with genomic and other biological data with particular requirements that may not be met by the national TREs.
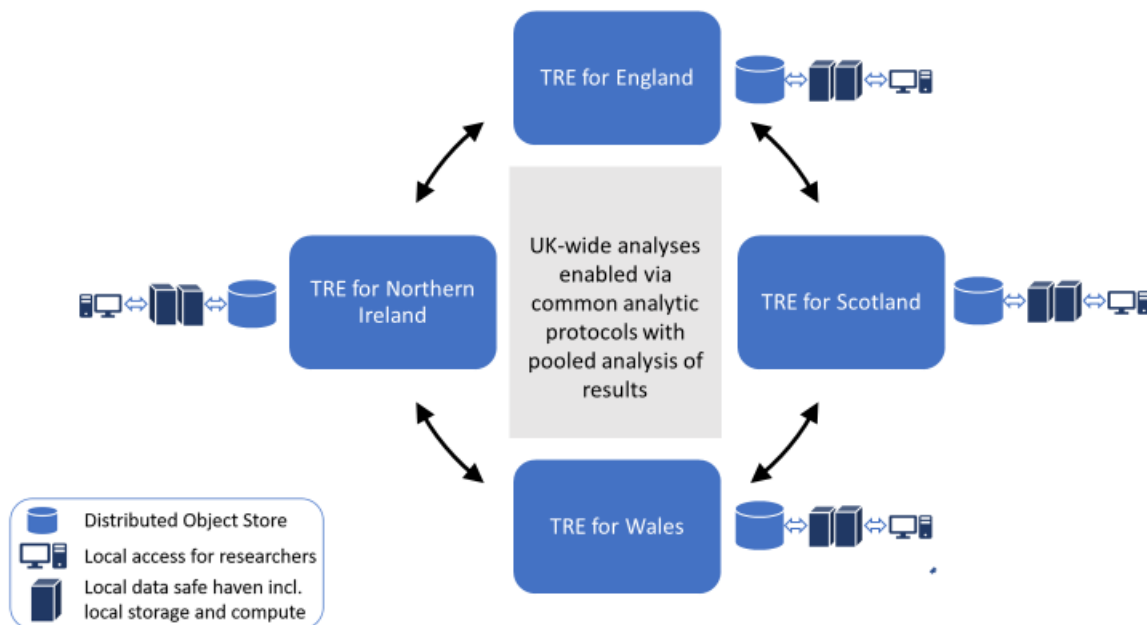
---

[5] SGSS: Second Generation Surveillance System; CHESS: COVID-19 Hospitalisation in England Surveillance System (CHESS); HES: Hospital Episode Statistics; SUS: Secondary Uses Service; BSA: Business Services Authority; PDS: Personal Demographic Service; ICNARC: Intensive Care National Audit and Research Centre; HQIP: Health Quality Improvement Programme; NICOR: National Institute for Cardiac Outcomes Research (datasets include Myocardial Infarction National Audit Project; Percutaneous Coronary Intervention audit; Cardiac Surgery audit; Heart Failure audit; Cardiac Rhythm audit; Congenital Heart Disease audit; Left Atrial Appendage Occlusion audit; Percutaneous Mitral Valve Leaflet Repair audit; Transcutaneous Aortic Valve Implantation audit; Patent Foramen Ovale closure audit); SSNAP: Sentinel Stroke National Audit Programme); COPD: chronic obstructive pulmonary disease; ISARIC-CCP: International Severe Acute Respiratory and emerging Infection Consortium – Clinical Characterisation Protocol

NB: UK Biobank is currently set up to receive data feeds of these datasets separately from PHE, NHSD and ICNARC, and receives linked primary care data for participants in England direct from the primary care computer system suppliers (TPP and EMIS). RECOVERY (and others) are currently pursuing a similar route for linkage to data from PHE, NHSD and ICNARC

Linkage within jurisdictions would be conducted by the trusted third party, with linked data for the whole population in that jurisdiction hosted in the TRE. Linkage would be achieved using the NHS (or CHI in Scotland, or Health and Care number in NI) number.

UK-wide analyses would be achieved through a federated network of TRE's across the four nations of the UK, by sharing of analytical outputs, as shown below:

Figure 2. Federated Network of Trusted Research Environments

C.   Information Governance and access

The discovery, access, authentication and analysis process has the following components:

- **Understanding of datasets:** The Innovation Gateway, will be the open source way of researchers understanding the availability of datasets and how to access them.
- **Prioritisation:** Research questions that have been prioritised through the Research Question Funnel will be prioritised for rapid access.
- **Application:** Each TRE and Trusted Third Party will provide a clear data access process, capable of operating at the scale and pace required for rapid response to prioritised research questions. Each request will need to be clear on its legal basis and purpose, and how it meets the policy and legal framework for each country. For example, within GDPR, COPI notices, Common Law on dissemination, and for England, whether National Data Opt-Out applies. Each will be required to have the relevant Transparency Notices.

- **Access:** Where access granted, each of the TREs will provide remote access for accredited academic, NHS, and government researchers to support multi-disciplinary team science. Each TRE will provide a practical approach for researcher accreditation and make this transparent on the Gateway so that the requirements are clear. The access journey for prioritised research questions will be tracked and transparent.
- **User Management:** Expedited Trusted access management processes will be used to assess and authorise remote access to data in each TRE.
- **Audit trails:** for all queries and activities and the resources to analyse and publish on the use of the data
- **Public transparency:** including publication of data access requests granted, stating the datasets, the research question, the legal purpose and the name of the organisation. In addition, patient groups included; and how subject access requests and other data subject rights, opt-out and ethics will be handled.
- **Data minimisation**: ensuing the fields and duration of data is minimised

The proposal will require further detailed development for each of these Information Governance and access arrangements.

Funding

Utilising existing assets and expertise, enables a fast and efficient approach. Additional funding is required to enhance and accelerate:

- The scale of the Access Management Teams in each TRE to enable speed of response
- Establishing and maintaining new data flows and linkage with frequent updates (e.g., fortnightly)
- Platform processing capacity, needed for the application of modern methodologies at scale.
- Software to accommodate specific analytical requirements, including, where necessary, engineering support for the installation of advanced modern analytical toolsets
- A small team to operate the question funnel and end to end performance metrics
- Teams of digital research professionals to support domain expert research teams

In addition, the researcher costs will need to be funded by re-purposing existing grants or through new funding sources.

Roles & Responsibilities

All **data custodians** who are members of the **UK Health Data Research Alliance**, supported by HDR UK, will be asked to make their data accessible for linkage to support rapid response to COVID-19 research questions through each national proposed route.

Where data is required to answer the prioritised questions, but custodians are not yet members of the Alliance, these custodians will be asked to join the Alliance and in parallel will be asked to make their data accessible through this route.

The UK Health Data Research Alliance network of **patient and public advisory groups** will be asked to involve patient and public representatives in the design of this approach and the access decision making processes to ensure they are trustworthy.

NHSD, SAIL, the Scottish National Data Safe Haven and the BSO Honest Broker Service will be responsible for ingesting data flows, data linkage, operating the TRE in line with the Five Safes model, and managing the access management process in a secure and rapid way to respond to the priority questions.

NICOR will continue to support NHSD, to deliver to SAGE in an expeditated fashion the key cardiovascular indicators relevant to service delivery on an ongoing basis. Together with the professional societies and relevant clinical colleagues, it will collect, curate, analyse and make available the data to all relevant stakeholders through the proposed research platform. NICOR will work with HDR UK, NHSD to ensure that the process is compliant with the information Governance framework adopted.

The **BHF Data Science Centre**, in partnership with NICOR and the NIHR-BHF CV Partnership (which brings together the combined BHF and NIHR-supported cardiovascular research community), will provide the expertise in priority, data-driven, cardiovascular questions, as well as in study design, analysis and interpretation to ensure good quality research questions are being fed into the question funnel, and that they are appropriately resourced with the appropriate, accredited, analytical expertise to answer the questions once they have access to the data and environment.

HDR UK will manage the research question funnel and operate the end to end performance tracking information and provide summary updates to SAGE, Trusted Third Parties, data custodians and PIs for the research questions so that any delays and obstacles are visible and quickly addressed.

Expected scale up to other conditions

The approaches to link multiple sources of clinical data within a trusted environment and the research approach that we propose are generic, and will not only meet the immediate cardiovascular ask, but will also enable rapid answers to a wide range of questions without disease-specific dataset linkages. For example, vaccination research questions, such as whether the MMR vaccine might provide protection against COVID-19 can be tested with the linked primary and secondary care data.

New disease-specific data to link into the environment will be prioritised based on the importance of the questions, whether a disease-specific dataset(s) exists, and what its added value is alongside the generic datasets. HDR UK will facilitate a datasets for linkage prioritisation exercise to help focus linkage efforts.

Indicative timescales

By
- 21 April 2020:

  – For the cardiovascular example presented here we would have the initial generic and disease specific datasets accessible and access granted to first research applicants for each of the Trusted Research Environments.
  – First set of key questions relevant to SAGE CV COVID policy for the short, medium and long term available to drive next data sets to be ingested

- 28 April 2020 (and weekly thereafter)

  – Summary of prioritised questions
  – Report on speed and scale of access
  – Summary of emerging research findings

Relationships with other health data initiatives

- **NHSX COVID DataCell -** This approach would be complementary and supportive of the operational and clinical focus of the NHSX team. That team's primary focus is providing national and regional decision-makers with reliable and timely data to support decision-making[6] on the supply and demand of healthcare, equipment and personnel during the crisis. It is also supporting other specific operational and policy decisions, but does not have a primary focus on the wider research community. The processes will be closely aligned, in particular access management and Information Governance, and some datasets will overlap, creating an imperative for collaboration to ensure a "single version of the truth".
- **Innovation Gateway** - will make metadata for datasets involved in this approach, discoverable and clearly signpost the access approach so that researchers understand what data is available and how to access it
- **Clinical Trials** – for example, NHS DigiTrials – is running in parallel, and using the same source datasets in NHS Digital, to support e.g. the RECOVERY trial on follow-up and outcome; and to provide a cohort for NHS Blood and Transplant to contact donors into the Convalescent Blood Plasma trial.
- **Health Data Research Hubs** – the other hubs, including DATA-CAN, Discover NOW, BREATHE, Pioneer (including DeCOVID), Gut Reaction will support the roll-out of this approach to other conditions (akin to the role the BHF Data Science Centre is undertaking in the cardiovascular example and NHS DigiTrials for clinical trials)
- **Genome sequence alliance (COG-UK) -** Sequencing of viral genomes, of those hospitalised, and also NHS workers to create family trees of virus genotypes, and different viral types, linking records of individuals with viral types-> understand specific genicity. Using the same source data from PHE and NHS Digital (in England), and using the MRC Cloud Infrastructure for Microbial Bioinformatics (www.climb.ac.uk)

---

[6] https://healthtech.blog.gov.uk/2020/03/28/the-power-of-data-in-a-pandemic/

- **Health Data Research UK Multi-omics Consortium –** This is bringing together about a dozen cohorts, collectively comprising some 400,000 participants. Taken together with UK Biobank, the cohorts provide a platform of ~1 million well-characterised research participants to study COVID-19 and CVDs (and other conditions), as the pandemic unfolds in the UK. The proposed approach in this paper would provide an environment to include analysis of these datasets.
- **NIHR-BHF Cardiovascular Partnership -** This existing Partnership which brings together the combined BHF (Research Excellence and Accelerator Award Centres, BHF Chairs) and NIHR (CV themes within BRCs) cardiovascular research infrastructure has created an operational framework to implement and resource coordinated strategy to (i) identify and prioritize Covid-19 research questions that respond to the Government's need, maximize impact and patient benefit and avoid duplication (ii) garner the full potential of the UK cardiovascular clinical and research community in a rapid, collaborative and transparent fashion (iii) align with other national and international initiatives. The Partnership will consider applications weekly and those that are approved and are data-driven will be channeled through the BHF Data Science Centre under this proposal.

Risks of this approach and proposed mitigation

The following risks have been considered in the development of this approach and suggested mitigation:

Risk	Mitigation
TREs are unable to support type of analysis that is needed (e.g. each TRE will be limited by the existing data within it, or that can be linked within the CV-19 time period).	Ensure each is supported in meeting a common set of user needs/stories across each nation. Each TRE data access environment has been designed to support multi-tenancy. Provide additional toolsets onto each TRE, as required (e.g. the NHS Digital Data Processing Service supports ACL, Python and SQL; and has set up a separate PHE environment to support R. This can be done for other CV19 projects).
Data access requests processes are not responsive enough for current emergency	Focus on the outputs of the prioritisation process, to ensure each request is addressed appropriately. Additional resource to support each TREs IG access process, while balancing the need to retain public trust. HDRUK to coordinate between each TRE.
Confusion over data request routes with NHSX Data Store	Make the distinction that the NHSX is focused on health technology policy, infrastructure and commissioning, within which the NHSX COVID DataCell is focused on supporting operational and policy decisions in the NHS Exec and central government.

Public / patient privacy concerns undermine proposal	Each Trusted third party to lead the transparency and fair processing of the data in each environment with explicit leadership from their respective public and patient advisory groups.
	Provide transparent communications on the proposed approach, responsibilities and a route for the public to raise concerns.
	Transparent and clear articulation of the 'five safes' for each TRE. For example NHS Digital has safe data (Privacy Enhancing Technologies to de-identify, use of derivations and restrictions), safe projects (only those with IG approval), safe people (staff with DBS check and above), safe settings (encrypted and cyber-tested platform), safe outputs (related to scope of project; researcher responsible for statistical Disclosure Control)
Legal basis for accessing some data ceases with COVID-19 emergency powers ceasing (or being extended) in Sep 2020, leading to time limited impact of research database.	DHSC/NHSX carefully consider the requirements for the research database at point of cease/extension
Linkage capability is insufficient to meet research requirements	In England, capability exists today via the Master Patient Service (MPS) component of Data Processing Service. MPS is an authoritative secondary use list of patients in England, fed daily from the direct care Personal Demographics Service (56m population).

Authors:

Caroline Cake, Health Data Research UK

Tom Denwood, NHS Digital

Cathie Sudlow, BHF Data Science Centre

David Seymour, UK Health Data Research Alliance

Nilesh Samani, British Heart Foundation

John Deanfield, NICOR

14 April 2020

Appendix 1: Cardiovascular Disease Research priorities to inform clinical and public health policy response to COVID-19

A. Key questions of relevance for UK cardiovascular disease data science community in the COVID 19 pandemic

(1) What is the impact of pre-existing cardiovascular disease,its risk factors (e.g. hypertension, diabetes, smoking. Ethnicity, gender) and cardiovascular medication on outcome in COVID 19 infection?

(2) What are the cardiovascular complications of COVID 19 infection?

(3) What is the impact of the NHS, public health and population response to the COVID 19 crisis on non-COVID diseases, in particular on the presentation, management and outcomes of other cardiovascular diseases?

(4) What are the molecular/multi-omic determinants and consequences of COVID 19 infection and what is the role and mechanism of cardiovascular disease in susceptibility to infection?

(5) What specific interventions might improve outcomes from COVID 19 infection in patients with pre-existing cardiovascular disease or cardiovascular complications of COVID 19?

B. Data-enabled routes to address these questions

(1) How does cardiovascular disease (and its treatment) affect outcome in COVID 19 infection?

This requires the study of cohort(s) of people with COVID 19 disease, characterised according to their cardiovascular disease history, medication (e.g. ACE inhibitors) and current status, with information on potential confounders (sex, age, socioeconomic status, health behaviours, other co-morbidities, drug treatments etc) and with follow-up information on their outcomes, both short term (e.g. within 30 days of admission to hospital, - death, requiring admission to ITU, requiring mechanical ventilation) and long term (e.g. death by cause at 6 months, one year and beyond post admission to hospital; recurrent stroke, MI and revascularisation events at 6 months, one year and beyond post admission to hospital)

Potential datasets that could address this question:

- *ISARIC CCP* (ISARIC CCP, as of 7 April 2020, n=9,000, increasing daily) – UK-wide NIHR-prioritised clinical characterisation of patients hospitalised with lab-proven or suspected COVID 19. Approval has been obtained for recruitment without consent.

  Substantially enhanced by linkages to primary care, secondary care (HES or devolved nation equivalents), mortality, ITU (ICNARC or SICSAG in Scotland) +/- community and hospital prescribing data.

  Additional linkages to NICOR (cardiac audits), SSNAP (stroke audit) and vascular registry (vascular surgical procedures audit) data would increase depth of cardiovascular disease characterisation.

- *ISARIC CCP with additional cardiovascular characterisation data (CAPACITY COVID,* CAPACITY*)*

- *All COVID 19 test +ve patients* from PHE (or devolved nation equivalent) lab testing data (as of 13 April 2020, n=69,329 in England, 6,067 in Scotland, 5,610 in Wales, 1,882 in N Ireland) linked to primary care, secondary care (HES or devolved nation equivalent), death registration, ITU (ICNARC/SICSAG) +/- community and hospital prescribing data.

- Additional linkages to NICOR, SSNAP and vascular registry data would increase depth of cardiovascular disease characterisation. (NB in theory all COVID 19 test +ve patients should be included in the ISARIC cohort but this is not yet the case.)

- *UK Biobank / other large multi-omic cohorts* linked to COVID 19 lab testing data, primary care, secondary care (HES), death registration, ITU (ICNARC) +/- community and hospital prescribing data. Additional linkage to NICOR, SSNAP and vascular registry data would increase depth of cardiovascular disease characterisation.

(2) What are the cardiovascular complications of COVID 19 infection?

This requires the study of cohorts of individuals with and without COVID 19 infection and/or individuals with different severities of COVID 19 disease with information on or follow-up for cardiovascular disease outcomes, both 'classical' CV outcomes (e.g. MI, stroke etc) and rarer, specific CV outcomes that may be a direct consequence/complication of COVID 19 infection (e.g. acute cardiac injury; myocarditis).

Potential datasets that could address this question:

- *ISARIC CCP* with linkages as above

- *ISARIC CCP with additional cardiovascular characterisation data (CAPACITY COVID)* with linkages as above

- *HIC-Cardiovascular COVID 19 initiative* (https://hic.nihr.ac.uk/)

- *Population wide linked data from England, Wales, Scotland and N Ireland:* COVID 19 lab testing data (to identify COVID 19 +ves) linked to primary care, secondary care (HES or devolved nation equivalents), death registration, ITU (ICNARC/SICSAG) +/- community and hospital prescribing data. Additional linkage to NICOR, SSNAP and vascular registry data would increase depth of cardiovascular disease characterisation.

- *UK Biobank / other large multi-omic cohorts* with linkages as above

(3) What is the impact of the NHS, public health and population response to the COVID 19 crisis on non-COVID diseases, in particular on the presentation, management and outcomes of cardiovascular diseases such as acute myocardial infarction and stroke?

During the current pandemic, people will continue to develop acute cardiovascular conditions (e.g., heart attacks, strokes, acute disturbance of cardiac rhythm, acute exacerbations of heart failure etc). Yet, emerging evidence from the UK and other countries is that hospital admissions of patients with acute

coronary syndromes have declined substantially in recent weeks, while patients are seeking medical help much later (or not at all) ([BHF on COVID 19](#), [ESC on COVID 19](#)).

Datasets that could inform on time trends analysis to assess population-level monthly incidence and cause specific mortality before during and after the epidemic of ischaemic heart disease, stroke, heart failure, cardiac surgery, cardiovascular interventional procedures, vascular surgery etc… include:

- *Hospital admissions with diagnostic and OPCS procedural coding (HES and devolved nation equivalents)*

- *NICOR datasets*

- *SSNAP*

- *Vascular registry datasets*

- *Mortality data from death registries*

- *Primary care data*

These would not need to be linked to enable informative analyses (although linkage would enable a greater range of informative analyses).

(4) ***What are the molecular/multi-omic determinants and consequences of COVID 19 infection and what is the role and mechanism of cardiovascular disease in susceptibility to infection?***

Potential datasets that could address these questions include:

- *UK Biobank / other large multi-omic cohorts* with linkages as above

(6) ***What specific interventions might improve outcomes from COVID 19 infection in patients with pre-existing cardiovascular disease or cardiovascular complications of COVID 19?***

Addressing this question would require streamlined data-enabled RCTs of promising interventions.

Appendix 2: Linkage and secure access for approved research to nationally-collated datasets in Scotland, Wales and Northern Ireland

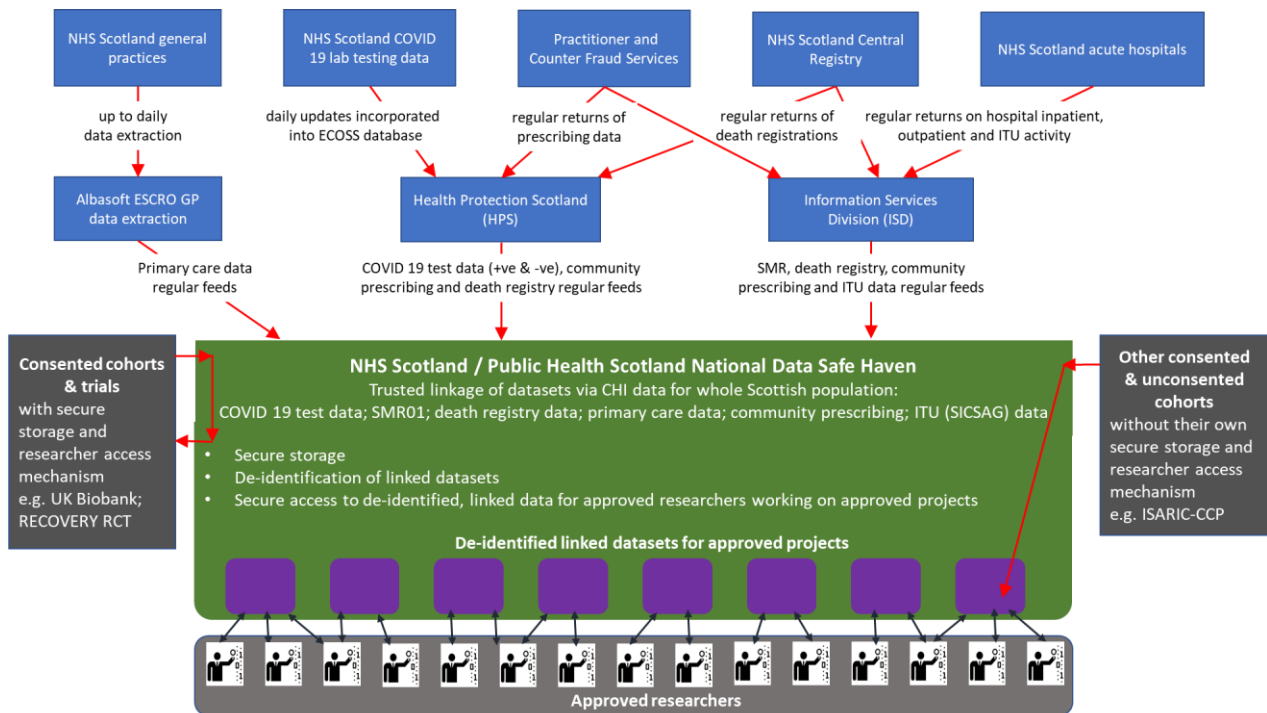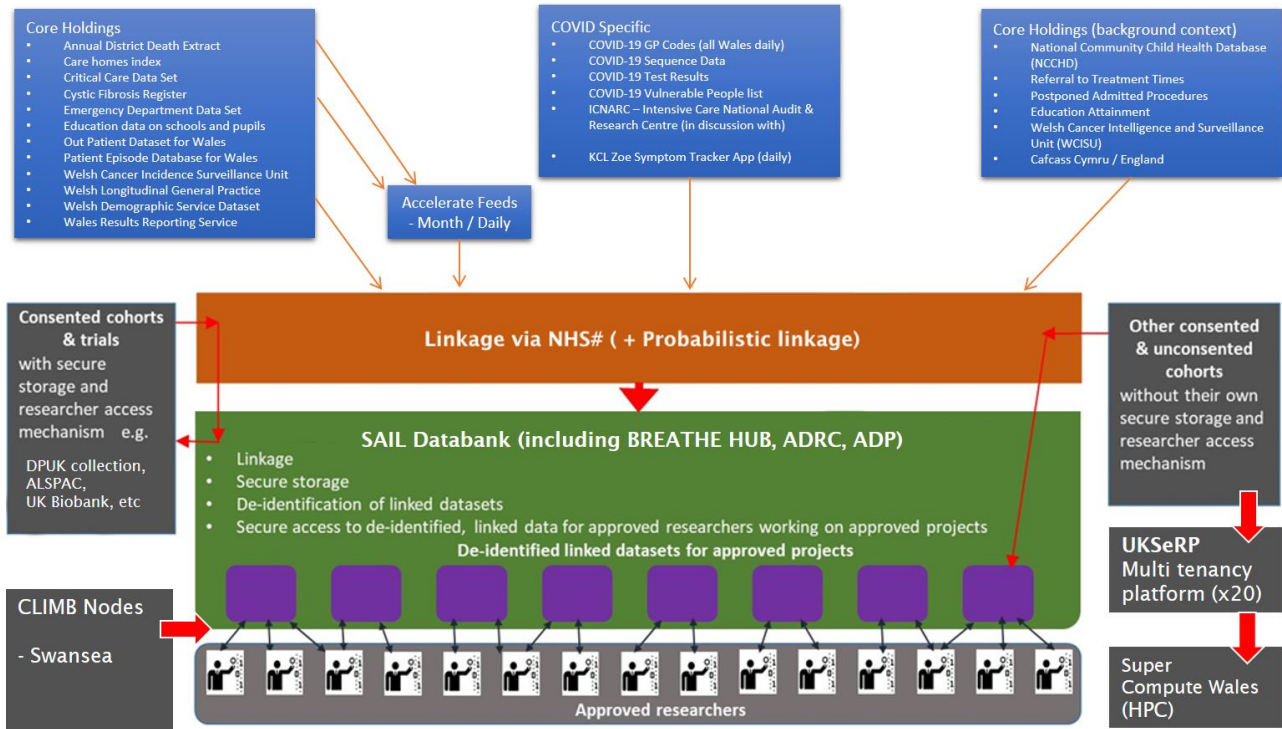*Figure 3: Linkage and secure access in Scotland*

Figure 4: Linkage and secure access in Wales

UKSeRP = UK Secure eResearch Platform; ADRC = Administrative Data Research Centre; ADP = MQ Adolescent Data Platform

Northern Ireland: the Honest Broker Service - http://www.hscbusiness.hscni.net/services/2454.htm