Data & Connectivity National Core Study

Sprint 6: 4 March – 31 March 2021

OUTCOMES

What were our aims for Sprint 6?

1. To respond to National Core Study (NCS) emerging data and connectivity needs:
   - **Vaccine research data infrastructure** – national Minimum Vaccine Dataset onboarded and available for approved researcher use within Trusted Research Environments (TREs) across the four nations.
   - **Launch feasibility assessment call for rapid acute data** needed for vaccine research.
   - **Genomic data and other key clinical data assets:** including enabling greater access and linkage to genomic data (e.g. COG-UK) and COVID 19 clinical characterisation data (ISARIC 4C).
   - **Prepare for Phase 1 of Data and Connectivity National Core Study** following successful funding award to extend programme for next 18 months.
   - **Continue to conduct data and connectivity mapping exercise** across **National Core Studies, NIHR Urgent Public Health studies and SAGE** sub-groups such as SPI-M.

2. To **engage researchers to use and enrich the Data and Connectivity data infrastructure**:
   - Supporting our TRE partners to process data access requests via the HDRUK Innovation Gateway from the 12 successful researchers from the joint HDRUK, ONS and UKRI rapid COVID-19 funding call.
   - Supporting these 12 successful research teams to complete their Milestone 2 review.
   - Commence improvements to the data access request process via the Gateway following feedback from TRE partners and researchers, focusing on system integration with TREs.

3. To make **core datasets relevant to COVID response available** to approved researchers in **secure cloud-based Trusted Research Environments (TREs)** in England, Northern Ireland, Scotland, and Wales:
   - Continue to work towards making previous Sprint datasets available:
     - Vaccine data
     - Viral genome data (COG-UK)
     - CO-CIN data for Wales and Northern Ireland
     - Serology datasets (Pillar 3 and priority Pillar 4 studies)
     - Intensive care data

   - Continue progress to build plan to **link secondary datasets before the end of March 2021**:
     - Tracing data
     - Shielded people list

- Wastewater
- Prescriptions
- Mental health
- Emergency care
- 999 records/ambulance and 111 data
- Maternity and neonatal
- Mobility data

4. To make all these priority COVID datasets **discoverable and accessible through a single "shop window", the Health Data Research Innovation Gateway.** This includes having metadata visible for all datasets and creating one aligned data access form for researchers to use to request data.
5. To implement a targeted programme of **communication and engagement** to inform sprint activities, increase researcher engagement and to improve patient and public involvement and understanding.

What has been achieved in Sprint 6?

1. Responding to National Core Study (NCS) data and connectivity needs

Continuing to enable the research data infrastructure for COVID-19 Vaccination research

a) **Vaccine data (Events and Adverse Reactions data) now flowing** to national TRE delivery partners in England and Wales and will be available in Scotland imminently; researchers can now start to work on answering urgent vaccine research questions by requesting access to these critical datasets through the HDR Innovation Gateway Progress continuing with availability and access to vaccine data in Northern Ireland.

b) The HDRUK convened **Vaccine Data Infrastructure Group** (chaired by Professor Sir Munir Pirmohamed) continues to map vaccine data flows and linkages required for research, identify key research questions, and enable data access.

c) **Feasibility assessment call for rapid acute data flow** needed for vaccine research was launched towards the end of sprint 5. Two proposals were received in sprint 6 but no awards were granted. Following this, **HDRUK convened an expert group across the four nations to discuss solutions to the need for rapid acute data flows**. The group agreed to focus on solutions for obtaining a rapid flow of data on strokes to support vaccine research on adverse events and work on this is ongoing.

d) **The task and finish group** set up in Sprint 5 has continued to compare and align the **definitions and analytical methods** being used for C-19 vaccine research and will create an open webpage on the HDRUK Innovation Gateway to share analytical methods and summary of methodological challenges.

e) HDRUK convened discussions between Office for National Statistics, Administrative Data Research UK (ADRUK) and NCS Longitudinal Health and Wellbeing and NCS Transmission and Environment to **discuss solutions to gain wider access and linkage to administrative datasets such as education, employment benefit and occupation for research. A** Working Group to be formed in Sprint 7 to identify key priorities to enable wider access and linkage to administrative data for the NCS and agree next steps.

2. Engaging with researchers to increase use of data infrastructure

a) **Enabling data enabled trial recruitment via Test and Trace:** we continue to work with AstraZeneca to roll out this approach to support the **TACKLE trial** and are working with the **PRINCIPLE** trial to try and enable this approach across the devolved nations, following it's successful implementation in England.

b) Supporting **data and connectivity needs for the** **Prophylactic Therapy in Care Homes Trial** (PROTECT-CH) from the University of Nottingham. The team are working with Data and Connectivity Delivery partners across the four nations to submit data access requests to enable the data flows and linkages required.

c) The 12 research projects awarded in our rapid funding call, which are listed on the HDRUK Innovation Gateway, submitted **21 Data Access Requests (DAR) covering 84 datasets** via the Gateway**. 10 projects had DARs approved within this Sprint.**

d) **Improvements to the data access process via the Gateway have commenced**, following user feedback collected from the researchers and TRE delivery partners. Improvements commenced in this Sprint include **ability to clone applications on the Gateway** and work with our delivery partner TREs to **ensure further integration of the Five Safes with TRE systems**. The next phase of user feedback will concentrate on the experience of researchers within the TRE environment, and will be gathered this sprint.

e) The studies will use data from across the 4 nations and will make their **own data accessible for others**

3. Making Data Available

Data has been linked and made available during the sprint:

a) **Agreement made between ONS and NHS Digital to add primary care data (GPES) to the ONS-NHSD Joint Health Data Asset (Census-Mortality-HES)**. The census-mortality-HES linked asset is now available to request from the ONS via the Innovation Gateway.

> **Office for National Statistics:**
>
> The joint data asset (linked hospital episode (HES), Mortality and Census data) is now available to researchers to request.

b) **Emergency care datasets and death registry datasets for Northern Ireland** now available via the Gateway.

c) **Vaccine Events and Adverse Reactions data** flowing into **ONS Secure Research Service (SRS), NHS Digital and SAIL.** The data is available to researchers from NHSD and SAIL, and will be made available shortly from the ONS, but researchers can still submit applications. Vaccine data will be **imminently available for Scotland.** Progress being made in Northern Ireland to enable access.

d) **COG-UK viral genomic data** is available in the Scottish TRE will soon start to flow into SAIL databank from Public Health Wales. Progress is being made across the other TRE partners.

> **Scottish National Safe Data Haven:**
>
> Vaccination data and all COG-UK data from website is now available in the TRE in linkable format. Regular weekly engagement with the Scottish Vaccination Programme Data Management team was established. Agreed the minimum dataset to be made available for research and process to be kept up to date with changes or quality issues as they occur. ISARIC data infrastructure development, including PHOSP data preparation, agreements and ingest as well as data access support continues.

e) Data from the **Covid-19 Clinical Information Network (CO-CIN)** available via **Scottish National Data Safe Haven** for research use, which includes a sub-set of **CO-CIN and COG-UK linked data**. CO-CIN data now also being linked with Pillar 1 and 2 testing data in England and work underway to link to **PHOSP-COVID data** and **vaccine data. There is also work ongoing to link CO-CIN data and 2011 census data in the ONS.**

f) **Pillar 4 serology data** available in ONS SRS (and linked to genome testing data for positive C-19 infection survey participants) and in NHS Digital and SAIL.

g) NHS Digital have started work to **provide a more frequent data flow of hospital activity**. They plan to make a monthly extract of hospital activity data available via the SUS (Secondary Uses service) dataset.

> **NHS Digital:**
>
> Work is progressing to make more frequent hospital activity data available within the NHS Digital trusted research environment. Data on community services is also now available within the TRE.

4. Making Data Discoverable and Accessible

a) **Enhancements to meta-data:**

- **75** (+3 from sprint 5) National Core Study priority datasets are now listed on the Gateway with detailed technical metadata for **73** (+4 from Sprint 5)
- **Researchers can now access data from the Northern Ireland Honest Broker Service** remotely using the UK SeRP platform.

> **Northern Ireland Honest Broker Service:**
>
> This Sprint has seen the launch of remote access for researchers to the Honest Broker Service using UK Secure eResearch Platform (UKSeRP). Developed by a team at Swansea University Medical School, this will provide a new mechanism to facilitate Northern Ireland health related research whilst providing a safe, secure and governed environment to ensure that patient data is safeguarded with confidence.
> The HDRUK funded research team from Kings College London will be the first time a research group from outside Northern Ireland has been enabled to remotely deliver their research and access HSCNI data via the SeRP.

5. Communicating and Engaging with partners, researchers, patients, and the public

a) Continued delivery of the **communications and engagement plan** to provide progress and updates on Data and Connectivity work.

b) Announced the **National Core Study Symposium event on June 24th** which will bring together the six National Core Study teams and wider community to share impact on UK COVID science and policy. Registration is now open, and the programme will be announced in coming weeks.

c) Announced 'Open Door Discussion – Our Response to COVID-19 ' held in the following sprint for members of the public to be part of an open discussion about the Data and Connectivity National Core Study, and have any questions answered.

d) **Newly designed webpages for Data and Connectivity now live.** These were designed with input from our patient and public advisory board.

e) **11 participants attended workshops to gather patient and public feedback on proposed accredited researcher passports**, which would allow accredited researchers to access multiple TREs. There was broad support, and participants raised important questions about security, who could access the data and how long the passports lasted.

f) **Blog published** sharing insights on the **impact of patient and public involvement in the Data and Connectivity National Core Study** so far

g) **A new public member was appointed** to join the weekly Delivery Group meetings to provide public perspective and challenge into the programme.

What we have learnt for next Sprints and key risks to delivery

1. The **scope of work for each sprint in Phase 0 (last 6 months) has been wide**, which has made it **difficult for our delivery partners to identify priorities**. As we move into **Phase 1,** which covers the next 18 months, the scope of work for each sprint moving forward will have a tighter specification to identify the key priorities for that Sprint, but still maintaining momentum and agility across our partners.
2. The user feedback from the 12 funded project research teams and from our TRE partners have **identified issues with delays in accessing data for urgent COVID research**, **complexity of data access request paperwork**, and **additional requirements** which need to be enabled (researcher accreditation, remote access IT requirements). Key areas for **improvement to the Gateway** on finding and requesting data are now being taken forward and we will work with our delivery partner TREs to help to further streamline the process.
3. There is still an urgent **need for a more rapid feed of hospital admission data to monitor roll out of the vaccination programme**. Current national data feeds are incomplete, lack sufficient detail or do not provide fast enough data access to researchers. The launched funding call to identify a team to run a feasibility study to establish a rapid data flow was unsuccessful and no awards were granted, and so this remains a critical need. **HDRUK therefore convened an expert group across the four nations to discuss solutions.** The group agreed to focus on solutions for obtaining a rapid flow of data on strokes to support vaccine research on adverse events and work on this is ongoing.
4. As we move to the 'Recovery' phase of the COVID response there is a **demonstrable** need for wider **access and linkage to administrative datasets such as education, employment benefit and occupation for research.** Our mapping exercise with the **NCS Longitudinal Health and Wellbeing and NCS Transmission and Environment** have highlighted this and we will work closely with them, ADRUK and ONS to deliver wider access to these data sets in future Sprints.
5. Enabling **rapid recruitment to urgent COVID clinical trials via data enabled approaches** has again shown to be highly efficient and deliver rapid results; especially when **directly informed by public feedback on acceptability of this approach.** We will continue to work closely with NCS Clinical trials Infrastructure & support - Therapeutics and Vaccines to enable this support when required.
6. **Finalising data sharing agreements,** and the timeframe required to resolve these remains a risk to our ability to achieve sprint deliverables at pace. We will continue to work closely with our TRE delivery partners and stakeholders to help facilitate and support progress.

Outline for Sprint 7

Sprint 7 Goal & Scope: By 30th April 2021

Deliver Phase 0 key priority linked and accessible data assets, develop roadmap for key prioritised data assets, continual user feedback improvement cycle, enrich and enhance assets available on Gateway, public engagement and transparency.

Sprint 7 - Measures of success	
Data asset	**1.** **Progress with COG UK, ICNARC and Vaccine datasets** and linkages available in cloud based TREs and with access for identified researchers undertaking priority approved projects 2. Roadmap developed for wider access and linkage to **sovereign data assets; (e.g ISARIC 4C, PHOSP)** **3.** **Continue to identify** urgent data and connectivity NCS priorities and enable solutions **4.** **Admin Data Linkage Working Group** co-ordinated with ADRUK to scope priority admin data infrastructure needs of NCS 5. Continue to provide support to NCS Clinical Trials and Studies – e.g PROTECT CH, PRINCIPLE
Data discovery	**4.** **Enrich and enhance the assets available to the research community on the Gateway** e.g. additional data sets, linkages, legacy resources (open datasets, code or tools) – build up NCS Collections on Gateway 5. Gateway improvement development work underway - **user informed improvements** **6.** **Data Access Request API integration: further scoping meetings with TREs to progress** integration into existing delivery partner system
Productive and safe settings	**6.** **Cohort discovery –** a) technical profiling, OMOP mapping of datasets in NCS TREs, test and deploy software in at least 3 TREs b) ICO led workshop **7.** **Federated identity**: a) Gateway to provide passport and Visa brokerage service b) TRE workshop to provide visa issuance service **8.** **Metadata catalogues –** test and deploy in at least 3 TREs
Communications and transparency	7. Targeted programme of **public engagement - open door, lay summaries, FAQ** **8.** **Development of Communications and Impact, PPIE strategies for Phase 1** **9.** **Impact case studies developed for website** **10.** **Phase 0 'Key impacts and lessons learnt'** published

Annex: Current Status of Health Data Research (SAGE Report 9th March 2021)

Openly available here

Research topics with new insights generated in last 2 weeks

Health data research outputs on COVID-19 continues to grow, now reaching 1,204 (+11) non-peer-reviewed pre-prints & 131 (+6) published papers.

Topic	Insights from ongoing studies (links provide further details):
Surveillance & Epidemiology	• A population-based cohort study involving >15,000 participants (COVIDENCE UK) found that people from Asian/Asian British ethnicities and people with raised body mass index have an increased risk of developing COVID-19 – whilst people with allergic diseases such as eczema and hayfever, have a decreased risk. • Using data from nearly 7,000 participants in UK Biobank, researchers have shown that older *biological* age is associated with a greater risk of COVID-19 hospitalisation and death. Biological age may be older or younger than a person's chronological age, and was predicted using telomere length in this instance. • A model,developed using health records from 2,815 COVID-19 inpatients from Wuhan, predicts patient outcomes at the point of COVID-19 hospitalisation, helped clinicians identify appropriate treatments at admission, with applications for long COVID. The model will benefit from being tested in larger datasets and different cohorts. • A study using laboratory and clinical data from hospitals in London and Oslo to validate published COVID-19 prediction models for hospitalised patients, found varying performance between the 2 sites – underscoring the need for local/regional/national recalibration of models and validation across multiple locations to improve accuracy.
Immunity & Vaccines	• Analyses of antibody tests linked to national SARS-CoV-2 testing programme data & vaccination records from >10,000 residents of 310 long term care facilities in England as part of the VIVALDI study, indicate that a 1st dose of the Pfizer-BioNTech or Oxford-AstraZeneca vaccine substantially reduces the risk of COVID-19 infection in older adults. Closely related, an observational study of >14,000 vaccinated older care home residents in Wales, combining electronic health records and administrative data from the SAIL databank, found a small number of residents were infected with COVID-19 following vaccination; these infections were associated with frailty and usually occurred within 28 days of vaccination. These findings are relevant for policymakers prioritising 2nd vaccine doses or revisiting the control measures in long term care facilities. • Recent research combining Canadian demographic, epidemiological, economic and Prospective Urban Rural Epidemiology (PURE) cohort data, estimates that those with obesity, diabetes or hypertension are at higher risk of death compared to individuals without these risk factors, therefore prioritising individuals with obesity, diabetes, or hypertension may be an efficient way to prevent deaths. • Survey response data from >20,000 adults as part of the VirusWatch household study indicate that 86% of participants who were reluctant or intending to refuse a COVID-19 vaccine in December were planning on accepting (or already had accepted) a vaccine in February – and these findings were consistent across ethnic and social groups. This shift in attitude highlights the need to offer vaccines repeatedly as people change their minds over time.
Longitudinal health & wellbeing	• A social media survey of >2,500 people, co-produced with patients, found that long COVID affects people's daily life, including their mental health, ability to work and do domestic chores. For the majority, symptoms fluctuate or relapse, and common triggers include physical activity, stress, sleep disturbance and cognitive activity. • Patient reported outcomes from >1,000 patients discharged from hospital following treatment for COVID-19 as part of the UK multicentre PHOSP-COVID study, revealed most survivors are not fully recovered five months after discharge; ~1 in 5 people developed a new disability, stopped working or changed jobs due to their health, and/or experienced symptoms of anxiety or depression. • Analyses of data from a smaller group of participants using International Severe Acute Respiratory and emerging Infections Consortium (ISARIC) WHO Clinical Characterisation Protocol (CCP-UK), indicate that COVID-19 survivors experience long-term symptoms, new disability, increased breathlessness, and reduced quality of life. • Individuals discharged from hospital after COVID-19 had increased rates of multiorgan dysfunction compared with the expected risk in the general population. The increase in risk was not confined to the elderly and was not uniform across ethnicities. The diagnosis, treatment, and prevention of post-covid syndrome requires integrated rather than organ or disease specific approaches, and urgent research is needed to establish the risk factors.
Transmission & Environment	• Ongoing progress.
Clinical Trials	• Ongoing progress.

Data & Connectivity National Core Study: COVID-19 dataset availability – 6 April

Vaccine data now available for Wales and England and will soon be available for Scotland. Viral genome data also soon to be available for Wales.

Core COVID-19 Datasets available for linkage	Office for National Statistics Secure Research Service	England (NHS Digital Data Processing Service)	Scotland (National Data Safe Haven)	Wales (SAIL Databank)	Northern Ireland (Honest Broker Service)
C-19 vaccine data collection	To be made available shortly. Accepting applications now	Vaccines Events & Adverse Reactions	Available in TRE, available via Gateway shortly.	COVID Vaccination Dataset	Vaccine Management System Operational – Data to be transferred to Data warehouse
COG-UK viral genome	Awaiting data flow from Public Health England	Awaiting data flow from Public Health England	Subset linked to CO-CIN data Remaining data being transferred, available shortly.	Data flow started from Public Health Wales	Governance agreed, automation of data flow to PHA in progress
Pillar 1 COVID-19 Testing Data	To be linked to Test and Trace data	COVID-19 Second Generation Surveillance System (SGSS)	Electronic Communication of Surveillance in Scotland (ECOSS)	COVID-19 Test Results	COVID antigen testing - Pillar 1
Pillar 2 Testing data (UK Gov)	To be linked to Test and Trace data	COVID-19 UK Non-hospital Antigen Testing Results (Pillar 2)	Electronic Communication of Surveillance in Scotland (ECOSS)	COVID-19 Test Results	Missing results prior to 26 Apr – Data quality issue
Primary Care	GPES linked to census, mortality and hospital data for internal access only	GPES extract – 98% practice coverage, large subset of codes (4bn items) Community Prescribing	Albasoft ESCRO GP Extraction* Prescribing Information System	80%+ coverage of full longitudinal record, with 100% coverage for COVID codes	Enhanced Prescribing Database as proxy
Secondary Care	Census-Mortality-HES linked data asset now available (ONS/NHSD)	100% coverage – HES. SUS via DARS extract only, available in TRE soon	100% coverage	100% coverage	Admissions & Discharges
Personal Demographic Service	Internal use only	100% coverage (via DARS extract only)	100% coverage	100% coverage	
Death registry	Provisional Monthly Extract & Linked Census and death occurrence	100% coverage Civil Registrations - Deaths	100% coverage	100% coverage	
C-19 Infection Survey (CIS)	Linked to Test and Trace data	N/A	Awaiting DEA accreditation	Awaiting decision on data access	Awaiting decision on data access
COVID-19 Clinical Information Network (CO-CIN)	Being linked to 2011 census	Data for English CO-CIN participants available in Scottish Nation Data Safe Haven	Limited metadata. Includes English linked data, and COGUK/CO-CIN data asset	Awaiting decision on data access	Discussions ongoing to collect data in NI
Census 2011	Household structure	N/A			N/A
Covid Opinions Survey		N/A	N/A	Awaiting decision on data access	N/A
Business Impact of Covid Survey	c. 5,000 businesses	N/A	N/A	Awaiting decision on data access	N/A
Labour Force Survey	40,000 households, 100,000 individuals	N/A	N/A	Awaiting decision on data access	N/A
Intensive Care data	HES Critical Care, ICNARC	HES Critical Care Reviewing IG to share ICNARC.	SICSAG (updated weekly)	ICNARC COVID weekly, ICNARC quarterly all admissions and critical care routine data (CCDS) monthly	Critical Care Minimum dataset to be acquired
Pillar 3 Testing data (NHS labs)	Captured within Test and Trace data	N/A		COVID-19 Test Results	
Pillar 3 Testing data (iELISA)	N/A	COVID-19 UK Non-hospital Antibody Testing Results (Pillar 3)			Data to be validated
Other Pillar 4 Testing data	VIVALDI, REACT II				
ZOE Symptom Study App Data	Finalising data sharing agreement			UK wide (unlinked) Wales (linked)	

HDRUK Health Data Research UK

Office for National Statistics

KEY

1. Custodian engagement
2. Dataset available in secure Trusted Research Environment
3. Linkages established to other priority datasets (within TRE)
4. Datasets available for COVID-19 research via Gateway

Further information about **Data & Connectivity** can be found here. Including the recently published Sprint 5 report.

Data and Connectivity National Core Study webpages and dashboard now live

| 4

Status of COVID-19 projects using the data – 6 April

17 new research projects now active, taking the total number of active research projects over 320. 10 of 12 rapid funded project now with at least one data request approved (increase of 6 projects since last report).

# of COVID-19 Projects by stage (change from previous report)	Office for National Statistics Secure Research Service	England (NHS Digital Data Processing Service)	Scotland (National Data Safe Haven)	Wales (SAIL Databank)	Northern Ireland (Honest Broker Service)	Total
In development	*6 (-)*	40 (+2)	35 (-1)	16 (-4)	3 (-1)	100 (-3)
Submitted for Information Governance approval	*1 (-5)*	10 (-2)	10 (+3)	1 (+1)	0 (-1)	22 (+2)
Approved but not yet active	1 (-5)	4 (-)	8 (-)	1 (-1)	3 (+1)	17 (-5)
Active research taking place	**22 (+6)**	**100 (+3)**	**75 (+4)**	**129 (+4)**	**1 (-)**	**327 (+17)**
Completed projects	*Coming soon*	*Coming soon*	0	0	*Coming soon*	
Active Number of Researchers	312 (+39)	56 (+4)	232 (+32)	254 (+5)	1	885 (+80)
Average time from application to active research	102 days	*Not yet available*	10 days	3 days	*Not yet available*	

Participation in key UK wide studies:

- PRINCIPLE: 4,717 participants (+2% in last 2 weeks, with continuing data flow of Pillar 2 COVID +ve test results to support recruitment)
- RECOVERY: 39,546 participants across 181 active sites (+0.5% in last 2 weeks, and +1 active site)
- CO-CIN (ISARIC 4C)
  - 200,287 Tier 0 (case report) (+3% in last 2 weeks)
  - 645 Tier 1 (single sample)
  - 1,665 Tier 2 (serial sampling)
- GENOMICC: 12,208 participants (+2% in last 2 weeks) across 212 ICUs with a total of 5091 intensive care beds
- COVID-19 ZOE symptom study: 4,641,484
- COG-UK: 413,687 viral genomes sequenced (+15% in last 2 weeks)

Data & Connectivity National Core Study projects now underway. Examples include:

- **Can phenotypes developed from enhanced remote primary care assessment of COVID-19 be used to identify a cohort of community cases, and enable comparison of recovered and long COVID** led by Professor Greenhalgh, University of Oxford
- **How is COVID-19 impacting women and men's working lives in the UK?** led by Professor Tracey Warren, University of Nottingham
- **Uptake and comparative safety of new COVID-19 vaccines by age, sex, region, ethnicity, comorbidities, medication, deprivation, risk level and evidence of prior COVID infection** led by Professor Hippisley-Cox, University of Oxford

HDRUK Health Data Research UK

Office for National Statistics

Data Access Registers

For more information on the active projects:

- **ONS Secure Research Service:** List of accredited researchers and research projects under the Research Strand of the Digital Economy Act)
- **NHS Digital:** Register of approved data releases (includes all access)
- **Scotland:** Public Benefit and Privacy Panel approvals
- **SAIL Databank:** COVID-19 projects listed on gateway
- **NI Honest Broker Service:** Projects currently being carried out.