Data & Connectivity National Core Study

Sprint 4: 6 January – 3 February 2021

OUTCOMES

What were our aims for Sprint 4?

1. To respond to National Core Study (NCS) emerging data and connectivity needs:
   - **Vaccine research data infrastructure** – including defining and communication of an agreed 'Vaccine minimum research dataset' and key linked datasets 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).

2. To **engage researchers to use and enrich the Data and Connectivity data infrastructure**:
   - Supporting successful researchers from the joint HDR UK, ONS and UKRI rapid funding call apply for data access and develop patient and public involvement and engagement (PPIE) plans for their COVID-19 research projects via the HDRUK Innovation Gateway
   - Develop Gateway and data access **user feedback questionnaire** for projects to inform current and future improvements.

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 sprint 3 datasets available:
     - Vaccine data
     - Viral genome data (COG-UK)
     - Serology datasets (Pillar 3 and priority Pillar 4 studies)
     - Intensive care data
     - Symptom Tracker: Zoe App data

   - Continue progress build plan **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 4?

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

Designing a robust UK research data infrastructure for the COVID-19 Vaccination Programme.

a) We coordinated a large group of researchers and organisations with expertise in vaccine surveillance, chaired by Professor Sir Munir Pirmohamed. The group aims are to map vaccine data flows and linkages required for research, identify key research questions and enable data access.
b) In Sprint 4, the group specified a **Minimum Vaccine Dataset** for research, identified **data linkage needs** and set up **two task and finish groups**. The first group will focus on **national vaccine data flows** (to ensure high quality, linked vaccine data is accessible for research in TREs across the 4 nations) and the second group on **acute care rapid access data** (to enable a near real time rapid access flow of acute care data to allow early detection of serious adverse events).
c) **Vaccine data now flowing** to national TRE delivery partners in England and Wales soon to be made available for research, progress made with availability and access to vaccine data in Scotland and Northern Ireland.

2. Engaging with researchers to increase use of data infrastructure

a) The 12 research projects awarded in our rapid funding call enable research which **uses and enriches the data within the Data and Connectivity National Core Study met their initial milestone objectives.**
b) All 12 projects now listed on the HDRUK Innovation Gateway, **submitted Data Access Requests (DAR)** via the Gateway **and developed PPIE plans for their projects.**
c) The projects have submitted **15 DARs** covering **63 data assets.**
d) We are collecting **user feedback** from the researchers and TRE delivery partners to drive forward improvements in the Gateway and DAR process. We will continue to collect user feedback throughout the life cycle of the projects.
e) The studies will use data from across the 4 nations and will make their **own data accessible for other researchers** too

> **Office for National Statistics:** All applications for data access requests for the rapid funding all projects are being given priority status as soon they arrive, we are making good progress and engaging with the research teams.

3. Making Data Available

Data has been linked and made available during the sprint:

a) **Vaccine data** flowing into **ONS Secure Research Service (SRS), NHS Digital and SAIL,** accessible soon for approved researcher use. Progress being made in Scotland and Northern Ireland to enable access.

b) 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.**

c) **The ONS and NHS Digital joint health data asset** approved and will be made available for approved researcher use via the Gateway.

d) **Pillar 3 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.

> **SAIL:** In Sprint 4, we have put most effort into setting up the flow of Welsh Immunisation Service data into the SAIL Databank, which is now live, and we are now beginning to construct documentation and data quality reviews on the data before making it more widely available to NCS and other studies. The data will be linkable to other NHS Wales data routinely within SAIL, such as GP, hospital episode, and COVID-19 testing

e) There is ongoing work across all the TRE partners to onboard and link **COG-UK genomic data**

f) **New ORCHID RCGP Trusted Research Environment** added to 'TRE Collections' on the Gateway

g) All the TRE delivery partners have **built a roadmap to ensure the newly identified datasets added in previous Sprints can be onboarded and linked by the end of March 2021**. Many of these datasets are already available via the Gateway.

> **Scottish National Data Safe Haven:** Data provisioning times are improved for projects. Greater collaboration across data controllers and research community.

4. Making Data Discoverable and Accessible

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

- **62** (+15 from sprint 3) National Core Study priority datasets are now listed on the Gateway with detailed technical metadata for **56** (+19 from Sprint 2)
- **Significant increase in metadata richness scores** in Sprint 4: 36 identified datasets have a score of platinum and vast majority of NHS Digital, SAIL and Scotland datasets have a platinum or gold score. All but 2 datasets have seen an increase in these scores.

b) **Evaluating dataset utility or 'usefulness'**

- **47** datasets for the National Core Study programme have now had their data utility evaluated using a new framework developed and launched in sprint 3. Data utility evaluates the 'usefulness' of the priority datasets for researchers.

> **Northern Ireland Honest Broker Service:** The NCS programme has helped raise the profile of the Honest Broker Service and helped expedite the collaboration with Swansea University and negotiations around the use of the UK Secure e-Research Platform. We are already seeing additional benefits to strengthen this partnership and collaborate further as part of the

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

a) Collections now launched for all TRE delivery partners; providing an **openly accessible summary of the capabilities, software, and compute available in each TRE** and helping to **inform the researcher data access process**

- Scottish National Safe Haven
- ONS Secure Research Service
- NHS Digital Trusted Research environment for England
- Northern Ireland Honest Broker Service TRE
- Secure e-Research platform / SAIL databank

b) During Sprint 4 we continued delivery of the **communications and engagement plan** to provide progress and updates on Data and Connectivity work. This included a video interview with Munir Pirmohamed on COVID-19 vaccine research.

c) Supported the **12 rapid funded projects to develop PPIE plans** for each of their projects and engaged public contributors to provide advice and feedback on these plans.

d) Put together **lay summaries for each of the 12 rapid funded project**s to be reviewed by public contributors and published on the website.

e) Recruited one additional new Lay Member to join the Delivery Group and three new Members to develop a new Advisory Group.

f) Further progressed the development of an **overarching PPIE strategy**, work ongoing to develop a longer-term strategy that will focus on transparency, public trust, accessible communication and focused approach to engaging with the wider public, and establish ways of working in which PPIE can influence decision making.

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

1. There is still much **complexity with vaccine data**. There are complicated issues around data controllership of this data, and across the four nations varying levels of complexity and political sensitivity. There has however been real progress made with access to this data for research and we hope wider access for approved researcher use will be forthcoming in Sprint 5. Working closely with our delivery partners and other key stakeholders however has allowed us to identify the key issues and progress **towards an enabling environment which allows open research** (with appropriate governance for safe researcher access) and ensure transparency and trust despite the sensitivity. There has been demonstrated value within this Sprint of the Data and Connectivity team acting as **a neutral third party to convene stakeholders in this space and drive progress.**
2. The rapid funding call has driven a sharp increase in **Data Access Requests (DARs) for NCS datasets via the Innovation Gateway**. This has identified **potential areas for improvement that** could further **enhance the user experience of the Gateway.** Key to the next Sprint will be collecting user feedback that will drive improvement to the data access request process and **increase engagement with the research community.**
3. There is **demonstrable interest in using and accessing the NCS C-19 data infrastructure** from the research community, as evidenced by the number of applications to the researcher funding call and subsequent increase in DARs via the Gateway. However, to ensure that we continue to focus on **data sets of priority** and continue to **engage the research community**, we will conduct a **data needs mapping exercise again in Sprint 5 and focus on sharing the impact of the Data and Connectivity work** as a priority for future sprints.
4. **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 5

Sprint 5 Goal & Scope: By 3rd March

Onboarding and linkage of vaccines and genomic data into TREs, progress of rapid funded projects, data asset needs mapping

Deliverables

1. **Health data infrastructure for C-19 vaccine research** – a) national Minimum Vaccine Dataset onboarded and available for access b) near-real time acute care data available for access
2. **Health data infrastructure for genomic data and other sovereign data assets**– enable greater access, linkage and availability for access of viral and host genomic data assets (e.g COG-UK, GenOMMIC, ISARIC 4C)
3. **Researcher funding call –** first data access requests approved, user surveys complete and results fed into Gateway.
4. **Data asset mapping exercise conducted**: to obtain data asset and linkage needs of NCS, NIHR UPH studies and SAGE sub groups
5. Progress with publication of delivery partner DPIAs and TRE DEA accreditation
6. **PPIE/Communications** – vaccine research, commercial trials, rapid funding call projects, development of medium-longer term strategy
7. **Completion of Sprint 1, 2, 3 and 4 activities:**
   - Core **priority datasets** available to approved researchers in secure cloud-based TREs with demonstrable linkage:
     - **Serology testing data** (in collaboration with CO-CONNECT)
     - **Intensive care data** - e.g. ICNARC and SICSAG
     - **CO-CIN routine health linkages** for Wales and NI
     - **ONS/NHS Digital jointly owned COVID data asset**
     - **ZOE Symptom app**

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

Openly available here

Research topics with new insights generated in last 2 weeks

Health data research on COVID-19 continues to grow, now reaching 1168 (+5) non peer-reviewed pre-prints & 98 (+5) published papers.

Topic	Insights from ongoing studies (links provide further details):
Surveillance & Epidemiology	• Four devices used for COVID-19 antibody tests were compared against each other by running the same c4500 samples through them. All devices displayed a trade-off between specificity and sensitivity - such as reduced sensitivity at lower antibody levels – and this could correspond to 5% false positive rates on the most specific device. • A novel algorithm to predict COVID-19 deaths in England (QCovid) developed using linked datasets from general practice, death registry and hospital episode data, has been externally validated in a population-based cohort study. The algorithm, which showed high levels of discrimination between men and women, has potential to support public policy and clinical risk assessments.
Immunity & Vaccines	• According to data collected using the ZOE app from nearly 40,000 people who received the Pfizer/BioNTech vaccine in December, people who have previously had COVID-19 are almost twice as likely to experience mild systemic after-effects such as fatigue and headache compared to people who have not had COVID-19. • Analyses of linked patient-level data on the OpenSafely platform indicate lower vaccination coverage in people over 80 with severe mental illness, dementia, and learning disabilities – as well as lower coverage in people from ethnic backgrounds and people from the most deprived communities. • Real-world data from 503,875 individuals in Israel suggest the 1st dose of the Pfizer/BioNTech is 51% effective 13-24 days after immunisation. • Recent results from the ONS Opinions and Lifestyle Survey found that 49% of people surveyed from the Black or Black British ethic group would be very or fairly likely to have the COVID-19 vaccine if offered. This is the lowest proportion amongst ethnic groups surveyed including White (85%), Mixed (80%), and Asian or Asian British (72%).
Longitudinal health & wellbeing	• A population-based study drawing on data from all 52 English NHS radiotherapy providers identified changes to radiotherapy services during the first wave of COVID-19 including: treatment delays where safe to do so, shorter radiotherapy courses, and increased radiotherapy to compensate for reduction in surgeries. The overall number of patients receiving radiotherapy fell significantly and remained reduced in June pointing to reductions in cancer referrals and diagnoses.
Transmission & Environment	• Early data from the VirusWatch Household Study found that people living in the most deprived areas of England and Wales were more likely to leave their house for work or school, use public transport and visit an essential shop compared to people living in the least deprived areas – suggesting that differences in daily essential activities may be contributing to the disproportionate infection rates in deprived areas. • A representative, household survey of SARS-CoV-2 infection amongst a UK strictly Orthodox Jewish population revealed an infection rate of 64% (5 times national estimates). This study adds to the growing body of evidence demonstrating a disproportionate impact of COVID-19 on minority groups. • Using a combination of crowd-sourced app data and existing geo-statistical techniques within a safe data environment, researchers are now able to predict COVID-19 hotspots at fine geographic scales in the UK. This is particularly relevant given that emerging risks such as deprivation may vary over small distances.
Clinical Trials	• Interim data from the PRINCIPLE trial (Platform Randomised trial of INterventions against COVID-19 In older peoPLE) show that two commonly used antibiotics, azithromycin and doxycycline, have no beneficial effect in patients over the age of 50 in the early stages of COVD-19.

Data & Connectivity National Core Study: COVID-19 dataset availability–9 February

Vaccine data now flowing into the ONS SRS and the Welsh SAIL databank, to be made available to researchers shortly.

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	Currently limited use only	Acting as data processor	Data to be transferred soon	Welsh Immunisation Service (WIS data now in SAIL	Data flow and IG to be agreed
COG-UK viral genome		Finalising governance	Subset linked to COCIN data a/w IG sign off for remainder	Governance agreed, to be linked shortly	Governance agreed, to be linked shortly
Pillar 1 COVID-19 Testing Data					
Pillar 2 Testing data (UK Gov)					Missing results prior to 26 Apr Data quality issue
Primary Care	GPES linked to census, mortality an hospital data for internal access onl	GPES extract – 98% practice coverage, large subset of codes (4bn items) Community Prescribing	AlbasoftESCRO GP Extraction* Prescribing Information System	80%+ coverage of full longitudina record, with 100% coverage for COVID codes	Enhanced Prescribing Database as proxy
Secondary Care	HES – available for internal access only 100% coverage	100% coverage- HES SUS via DARS extract only	100% coverage	100% coverage	
Personal Demographic Service	100% coverage	100% coverage (via DARS extract only)	100% coverage	100% coverage	
Death registry	Provisional Monthly Extract &Inked Census and death occurrence	100% coverage	100% coverage	100% coverage	
C-19 Infection Survey (CIS)		N/A	Awaiting DEA accreditation	Awaiting decision on data access	Awaiting decision on data acce
COVID-19 Clinical Information Network (CO -CIN)	Being linked to 2011 census	English COCIN data available in Scottish Nation Data Safe Haven	Limited metadata	Awaiting decision on data access	
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	Reviewing IG to share ICNARC. HES Critical care to be onboarded to TRE	SICSAG (updated weekly)	ICNARC COVID weekly, ICNARC quarterly all admissions and critic care routine data (CCDS) monthly	3 datasets being explored
Pillar 3 Testing data (NHS labs)	Available in CIS data	Internal use only			
Pillar 3 Testing data (iELISA)	N/A	Available via DARS extract only			Data to be validated
Other Pillar 4 Testing data	VIVALDI, REACT II				
ZOE Symptom Study App Data				UK wide (unlinked) Wales (linked)	

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 **Sprint 3 report**.

New Trusted Research Environment collection added to Gateway collections.

| 4

Status of COVID-19 projects using the data – 9 February 2021

39 additions to active research takes total over 290. Most new projects enabled by the SAIL Databank in Wales. 15 data access requests covering 63 data assets now submitted via the Gateway for the 12 rapid funded projects.

# 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	*11 (-6)*	32 (+1)	40 (+4)	90 (-)	6 (-)	179 (-1)
- a/w researcher	*11 (+1)*	*25(+6)*	*Not available*	*55 (-1)*	*6 (-)*	*N/A*
- a/w data custodian	*0 (-3)*	*7(-5)*	*Not available*	*35 (+1)*	*0 (-)*	*N/A*
Submitted for Information Governance approval	*4(+1)*	7 (-1)	5(+1)	0 (-1)	0 (-)	16 (-)
Approved but not yet active	2 (-4)	3 (+2)	9 (-)	3 (+1)	1(-)	18 (-1)
Active research taking place	**15 (+4)**	**87 (+3)**	**69 (+2)**	**121 (+30)**	**1(-)**	**293 (+39)**
Active Number of Researchers	*Not yet available*	*Not yet available*	*Not yet available*	205 (+3)	*Not yet available*	
Average time from application to active research	*Not yet available*	*Not yet available*	*Not yet available*	3 days	*Not yet available*	

Participation in key UK wide studies:

- PRINCIPLE: 4,053 participants (+8% in last 2 weeks, with continuing data flow of Pillar 2 COVID +ve test results to support recruitment)
- RECOVERY: 35,672 participants across 178 active sites (+11% in last 2 weeks, and +1 active site)
- CO-CIN (ISARIC 4C)
  - 176,413 Tier 0 (case report) (+9% in last 2 weeks)
  - 650 Tier 1 (single sample)
  - 1,658 Tier 2 (serial sampling)
- GENOMICC: 10358 participants (+12% in last 2 weeks) across 212 ICUs with a total of 5091 intensive care beds
- COVID-19 ZOE symptom study: 4,582,772
- COG-UK: 250,673 viral genomes sequenced (+17% in last 2 weeks)

Data & Connectivity National Core Study projects now underway

- 15 data access requests covering 63 data assets from 12 studies now submitted via **HDRUK's Innovation Gateway.** Examples include:
- Is exposure to airborne fine and ultrafine particulate matter a determining factor in COVID-19 infection and outcome within the UK? led by Professor Kevin Wyche, University of Brighton
- Enhancing the Utilisation of COVID-19 Testing in Schools Studies: The Joint Analysis of the ONS COVID-19 School Infection Survey and COVID-19 Mapping and Mitigation in Schools (CoMMinS) Study led by Dr Rachel Denholm, University of Bristol
- What are the relative contributions of different exposures and settings to COVID-19 community transmission? Analysis of community cohort studies linked to national testing data led by Professor Andrew Hayward, UCL

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.