Data & Connectivity National Core Study

Sprint 3: 5 December 2020 – 6 January 2021

OUTCOMES

What were our aims for Sprint 3?

1. To respond to National Core Study (NCS) emerging data and connectivity needs:
    - **Vaccine research data infrastructure** – including to define primary care data requirements (all NCS teams)

2. To **engage researchers to use and enrich the Data and Connectivity data infrastructure**:
    - Select and support successful researchers from the joint HDR UK, ONS and UKRI rapid funding call to use data for COVID-19 research

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 2 datasets available:
        o Serology datasets (Pillar 3 and priority Pillar 4 studies)
        o Intensive care data
        o Viral genome data (COG-UK)
        o Symptom Tracker: Zoe App data
    - Build a plan to link **newly identified datasets before the end of March 2021**:
        o Tracing data
        o Shielded people list
        o Wastewater
        o Prescriptions
        o Mental health
        o Emergency care
        o 999 records/ambulance and 111 data
        o Maternity and neonatal
        o 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 3?

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, in preparation for approval and national rollout of a Covid-19 vaccination programme. This group outlined the requirements for a robust UK research data infrastructure to support the COVID-19 vaccination programme. The proposal aims to provision the UK with the linked data infrastructure for research needed to evaluate vaccine safety and effectiveness at a population level in the real world and is essential to support the research needs of **all the National Core Studies.**

   b) In sprint 3, the proposal was reviewed and supported by the Medicines and Healthcare products Regulatory Agency (MHRA), the Public Health Agencies, and MHRA, PHA and the Chief Scientific Advisors. The Data and Connectivity TRE delivery partners are **now identifying current data flows of vaccine data across the 4 nations.**

2. **Engaging with researchers to increase use of data infrastructure**

   a) A rapid funding call was launched in sprint 2 to enable research which **uses and enriches the data within the Data and Connectivity National Core Study.** 48 applications were received, and **12 projects were selected** by an external panel during sprint 3.

   b) The 12 studies cover topics including **palliative care, long-COVID, epidemiology, vaccines, ethnicity, inequalities, and employment.**

   c) The 12 studies will request data via the new Data Access form which is now live on the Gateway. The studies will use data from across the 4 nations and will make their own data accessible for other researchers too.

   d) Our TRE delivery partners have now started to support the data needs of these researchers.

> **"**
> **Welsh SAIL Databank:** We have been able to make good progress with initial discussions surrounding data provisioning to the recently funded National Core Study rapid funding call projects.

3. Making Data Available

Data has been linked and made available during the sprint:

a) Data sharing agreements have been agreed in Northern Ireland and Wales to onboard and link **viral genome data from COG-UK** to routine health data. This data is vital for research now with the presence of a new variant of COVID-19.

b) **Pillar 1 and 2 testing data** has now been linked to the ONS Coronavirus Infection Survey and will be available via the ONS Secure Research Service shortly.

c) Data from the **Covid-19 Clinical Information Network (CO-CIN)** is currently being linked to the 2011 Census data in the ONS and will also be available via the ONS SRS shortly. CO-CIN data is now available in the Scottish Safe Data Haven.

> **Scottish Safe Data Haven:** The COVID-19 Information Network (COCIN) is using the National Safe haven to host their clinical datasets and linked administrative health datasets thus providing a secure environment for International Severe and Acute and Emerging Infection Consortium (ISARIC) researchers to utilise clinical data collected directly via hospitals across the UK on COVID patients.

d) There is ongoing work across all the TRE partners to onboard and link **Pillar 3 and Pillar 4 serology testing data.**

e) The **joint health data asset (ONS and NHS Digital)** is almost complete and will be available to researchers shortly. This is the first time that **routine health and administrative data** has been linked for England.

f) **Death registry data** is now linked to routine health data in Northern Ireland and is available for researchers to access via the Gateway.

g) All the TRE delivery partners have **built a roadmap to ensure the newly identified datasets added in sprint 3 can be onboarded and linked by the end of March 2021**. Many of these datasets are already available via the Gateway.

> **Northern Ireland Honest Broker Service:** The Data and Connectivity programme has added impetus to accelerate availability of new data in the Regional Data Warehouse, particularly the COVID related data but also datasets such as ICU data which will have added benefits to researchers

4. Making Data Discoverable and Accessible

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

- **47** (+9 from sprint 2) National Core Study priority datasets are now listed on the Gateway with detailed technical metadata for **37** (+7 from Sprint 2)

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

- **46** 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:** Increasing the accessibility and visibility of Northern Ireland data has been a major benefit, we have been publishing more meta data than previously and the Data and Connectivity programme is helping improve the quality of meta data available.

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 3 we continued delivery of the **communications and engagement plan** to provide progress and updates on Data and Connectivity work. This included the publication of the Sprint 2 report with highlights, a video interview with Sharon Peacock (COG-UK) on viral genome sequencing and a Q&A interview with Heather Rutter, a Senior Clinical Nurse on the PRINCIPLE trial.

c) Consultation to assess **public, patient and practitioner views to prioritise research questions related to COVID-19 vaccines.** A survey ran for two weeks over the Christmas and New Year period and received 828 responses with 304 separate comments and questions in addition to the research question prioritisation. The full report can be found here. A summary of the results shows:

- Patients and the public rank research questions on the safety of the vaccines and length of immunity as the most important

- Practitioners have a stronger affinity to exploring the effectiveness of the vaccines across different populations.

- Whilst it did not show in the overall ranking of research questions, following the news of an additional vaccine being rolled out across the UK, there was a noticeable difference in comments from patients and members of the public. Safety and effectiveness of the vaccine became less important whilst vaccine uptake and population behaviours were prioritised.

d) Developed and disseminated PPIE Overview Form to assist the awarded researchers through the rapid funding to embed patient and public engagement and involvement (PPIE)

e) Recruitment for an additional Lay Member to join the Delivery Group got underway following consultation with existing Lay Member to shape the role profile.

f) Consulted the Public Advisory Board for the development of a PPIE strategy. As well as a commitment to help shape the strategy, there was support for an overarching strategy that outlines key requirements and guidelines to be considered throughout each sprint.

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

1. There is a lot of **complexity with vaccine data**. There are several different stakeholders involved in this space and it is politically sensitive which makes it very important to define roles and remits of the organisations involved. It is also important to **create an enabling environment which allows open research** (with appropriate governance for safe researcher access) and ensure transparency and trust despite the sensitivity. There is value in the Data and Connectivity team acting as **a neutral third party to convene stakeholders in this space and drive progress.**

2. There is **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. A **focus on sharing the impact of the Data and Connectivity work** to date will hopefully increase researcher engagement and will be a priority for future sprints.

3. There are **different ways of working** across delivery partners which impact models for onboarding datasets and timelines. Convening different groups to **share experience** and ways of working is proving invaluable to **improve efficiencies** across all organisations but must also be considered in future sprint planning.

4. **Finalising data sharing agreements,** and the timeframe required to resolve these, risks our ability to achieve sprint deliverables at pace. The TRE delivery partners have made great progress with finalising data sharing agreements in sprint 3 to onboard and link the COG UK data, but the many stakeholders involved, and competing priorities have delayed these activities.

Outline for Sprint 4

Sprint 4 Goal & Scope: By 3rd February

Vaccines minimum dataset into TREs and supporting data access for rapid funding call

Deliverables

1. **Health data infrastructure for C-19 vaccine and genomic research** – System requirements designed and integrated into D&C roadmap

2. **Researcher funding call** – successful projects launched, and accelerated data access supported

3. **Develop scoping plan** to enable greater access and linkage to ISARIC 4C, PHOSP, GENOMMIC sovereign data assets

4. **PPIE** – vaccine research, rapid funding call projects and recruitment of additional D&C advisor

5. **Completion of Sprint 1, 2 and 3 activities:**

   - Core datasets available to approved researchers in secure cloud-based TREs with demonstrable linkage:
     - **Serology testing data** (in collaboration with CO-CONNECT)
     - **Viral genome data** – COG-UK
     - **Intensive care data** - _e.g._ ICNARC and SICSAG
     - **Zoe Symptom tracker app**
     - **CO-CIN routine health linkages** for Wales and NI
     - **ONS/NHS Digital jointly owned COVID data asset**

Annex: Current Status of Health Data Research (SAGE Report 24th November 2020)

Openly available here

Research topics with new insights generated in last 4 weeks

HDR UK
Health Data Research UK

Office for National Statistics

Health data research on COVID-19 continues to grow, now reaching 1153 (+31) non peer-reviewed pre-prints & 87 (+12) published papers.

Topic	Insights from ongoing studies (links provide further details):
Surveillance & Epidemiology	• Analyses of genome sequencing data from the COVID-19 Genomics UK (COG-UK) of over 1000 SARS-CoV-2 samples from the UK found that the majority of mutations are due to host attempts to edit RNA, rather than viral errors in replication. Of note, 4% of cases showed co-infection with more than one variant at a time with more infections yielding more mutations. • New research investigating the role of a human enzyme, furin, in protection against SARS-CoV-2, identified a key antiviral protein (IFITM2), which prevents SARS-CoV-2 entry - the mode of action is related to the furin-cleavage site in the virus's spike protein. The data suggest that therapeutics acting on furin-mediated cleavage of SARS-CoV-2 S may reduce viral replication. • An international data sharing collaborative, The Consortium for Clinical Characterization of COVID-19 by EHR (4CE), involving over 36,000 patients from 342 hospitals in the US and Europe, revealed that laboratory test values at admission can be used to predict severity in patients with COVID-19. This addresses the largely unmet need to predict severity across hospitals/regions, and given the multinational cohort, it is likely that these results will generalise well. • Use of the OpenSAFELY platform to analyse routinely collected electronic primary care data linked to national death registrations has revealed that people with HIV in the UK are at increased risk of COVID-19 mortality. As the pandemic response evolves, targeted policies could be considered to address this increased risk.
Immunity & Vaccines	• Results from the ONS Opinions and Lifestyle Survey over Christmas and New Year show a marginal increase in public willingness to be vaccinated, whilst support for mass testing remained the same. Specifically, 85% of adults would be either *very likely* or *fairly likely* to have the COVID-19 vaccine if offered and 81% of adults *strongly supported* or *tended to support* mass testing for COVID-19. • The ZOE COVID Symptom study has now pivoted to track vaccine side effects, with the most recent app update including improved COVID-19 vaccine questions. This data will provide important insight into the vaccines people are having and how they are responding.
Longitudinal health & wellbeing	• Longitudinal analyses of the CovidLife surveys, including over 11,000 participants, revealed that wearing a face mask is associated with better rather than poorer mental health and wellbeing (even after controlling for behavioural, social, and psychological confounders). These findings may spur policymakers to introduce further, or more stringent, guidelines on face coverings.
Transmission & Environment	• Two independent analyses of the new SARS-CoV-2 variant (VOC 202012/01) have found it to be more transmissible than other variants by over 50% (and here) presenting a major challenge for on-going control of COVID-19. Further research to understand the nuances of the transmissibility changes as well as any changes in severity is crucial. • A systematic review combining global data found that migrants in high income countries are at high risk of COVID-19 infection and over-represented in confirmed COVID-19 cases including deaths. This work highlights the urgent need for engagement with migrants, particularly surrounding barriers and facilitators to vaccination.
Clinical Trials	• The latest trial data from A Randomised, Embedded, Multifactorial, Adaptive Platform Trial for Community-Acquired Pneumonia (REMAP-CAP) showed that two medicines (immune modulators, tocilizumab and sarilumab), improved survival in critically ill patients with Covid-19. Updated guidance from the government and NHS has been issued to Trusts across the UK, effective immediately, encouraging the use of tocilizumab in the treatment of COVID-19 patients who are admitted to intensive care units.

Data & Connectivity National Core Study: COVID-19 dataset availability – 12 January 2021

Pillar 1 and 2 data is now linked in the ONS Secure Research Service, and death registry data now is available in Northern Ireland's Honest Broker Service.

HDR UK
Health Data Research UK

Office for National Statistics

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)
Pillar 1 COVID-19 Testing Data	Initial linkage completed				
Pillar 2 Testing data (UK Gov)	Initial linkage completed				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	HES – available for internal access only 100% coverage	100% coverage	100% coverage	100% coverage	
Personal Demographic Service	100% coverage	100% coverage	100% coverage	100% coverage	
Death registry	Provisional Monthly Extract & Linked Census and death occurrence	100% coverage	100% coverage	100% coverage	
C-19 Infection Survey (Pillar 4)		N/A			
COVID-19 Clinical Information Network (CO-CIN)	Being linked to 2011 census	Data for linkage in Scottish Data Safe Haven agreed	Limited metadata	Awaiting decision on data access	
Census 2011	Household structure				N/A
Covid Opinions Survey		N/A	N/A		N/A
Business Impact of Covid Survey	c. 5,000 businesses	N/A	N/A		N/A
Labour Force Survey	40,000 households, 100,000 individuals	N/A	N/A	N/A	N/A
Intensive Care data	HES critical care	ICNARC – internal use only	SICSAG (updated weekly)	ICNARC: COVID-specific weekly extract	
Pillar 3 Testing data (NHS labs)		Internal use only			
Pillar 3 Testing data (iELISA)		Internal use only			
Other Pillar 4 Testing data	VIVALDI, REACT II	VIVALDI			
ZOE Symptom Study App Data				UK wide (unlinked) Wales (linked)	
COG-UK Viral Genome		Finalising governance	Finalising governance	Governance agreed, to be linked shortly	Governance agreed, to be linked shortly

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 2 report**.

HDRUK asked the public, patients and practitioners to **prioritise vaccine research questions** put forward by National Core Studies. **Read the feedback here**

Status of COVID-19 projects using the data – 12 January 2021

14 additions to active research taking place takes total over 240. Most new projects enabled by NHS Digital or the Scottish National Data Safe Haven. There is a decrease in the bottle neck at projects submitted for IG approval.

# 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	18 (-)	32 (+3)	30 (-)	90 (-)	6 (-)	176 (+3)
- a/w researcher	10 (-)	24 (+2)	Not available	56 (-)	6 (-)	N/A
- a/w data custodian	8 (-)	8 (+1)	Not available	34 (-)	0 (-)	N/A
Submitted for Information Governance approval	9(-)	3 (-3)	4(-2)	0 (-)	0 (-)	17 (-4)
Approved but not yet active	7 (-)	4 (+1)	9 (+2)	2 (-1)	1(-)	23 (+2)
Active research taking place	6 (-)	82 (+7)	65 (+6)	89 (+2)	1(-)	242 (+14)
Active Number of Researchers	Not yet available	Not yet available	Not yet available	180 (-)	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: 3,218 participants, an increase of around 25% in last 4 weeks following establishment of data flow of Pillar 2 COVID positive test results to support recruitment
- RECOVERY: 26,555 participants across 176 active sites (an increase of 40% in last 4 weeks)
- CO-CIN (ISARIC 4C)
  - 140,089 Tier 0 (case report)
  - 661 Tier 1 (single sample)
  - 1,717 Tier 2 (serial sampling)
- GENOMICC: 7,896 participants across 210 ICUs with a total of 5075 intensive care beds
- COVID-19 ZOE symptom study: 4,533,041
- COG-UK: 117,375 viral genomes sequenced

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.

12 new COVID-19 research studies funded

- Selected following rapid funding call from HDRUK, ONS and UKRI to use and enrich the National Core Study data.
- Studies have been chosen for their potential to impact near real time Covid-19 responses.
- 12 studies cover topics including: palliative care, long COVID, epidemiology, transmission, vaccines, ethnicity, inequalities, and employment.
- 12 studies will use data from across the four nations and will make data accessible for other researchers.
- Researchers will request data for their studies via the new Data Access Form on The Innovation Gateway