

Data & Connectivity National Core Study

Sprint 2: 5th November – 4th December 2020

OUTCOMES

What were our aims for Sprint 2?

1. To respond to National Core Study (NCS) emerging data and connectivity needs:
 - **Occupational data** - landscape assessment of data sources, terminologies and coverage (Transmission & Environment study requirement)
 - **Vaccine research data infrastructure** – including to define primary care data requirements (all NCS teams)
2. 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:
 - Building on the Sprint 1 testing datasets, newly added datasets for sprint 2:
 - Serology datasets (Pillar 3 and priority Pillar 4 studies)
 - Intensive care data
 - Viral genome data (COG-UK)
 - Symptom Tracker: Zoe App data
3. 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.
4. 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 2?

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

- a) Landscape assessment of available **occupational data** conducted by all delivery partners. To be used by **Transmission and Environment NCS** to inform future data sources and consistent terminologies for data collection
- b) **Designing a robust UK research data infrastructure for the COVID-19 Vaccination Programme.** Concept note developed in collaboration with NCS leads and key system stakeholders to outline the requirements for a robust UK research data infrastructure to support the COVID-19 vaccination programme. This 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 NCSs**. Proposal currently in review by MHRA, PHAs and CSA.

2. Making Data Available

Data has been linked and made available during the sprint:

- a) **Intensive care data** is now available in Wales (ICNARC) and Scotland (SICSAG), linked to routine health data (e.g., primary care, secondary care, C-19 testing, mortality). Progress made to agree permissions for sharing linked ICNARC data in England and Northern Ireland.
- b) **CO-CIN (ISARIC)** data has been linked to the Scottish health data with a formal governance process for rapid access due shortly. Linked data for English participants was provided in [Sprint 1](#). Ongoing discussions to agree data flows for linkage for participants from Wales and Northern Ireland.

“

Scottish Safe Data Haven: “This programme of work has help to make great strides in accessing datasets both within and outwith Public Health Scotland. Our TRE now hosts all the UK ISARIC data linked to Scottish health data for the Scottish component and are in the process of finalising a more formal governance process for rapid access.”

- c) **Pillar 3 serology testing** data (Local NHS lab data) now available in Wales and Northern Ireland. Additional work across all delivery partners to onboard Pillar 3 Thrive testing data flows.
- d) **Pillar 4 priority datasets** identified in response to Immunology NCS needs (**REACT, SIREN, VIVALDI**). Active engagement with data custodians and ongoing work to onboard datasets to delivery partner TREs.
- e) For the first time ever a **joint health data asset (ONS:NHSD)** allows **linkage of routine health and administrative data** for England. Ongoing work to develop a metadata catalogue with data access for researcher from January.



Office for National Statistics and NHS Digital: “The partnership ONS and NHS Digital has been vital, we have had the opportunity to agree an IG approach for the first time due to this project.”

- f) **COG-UK** data – active engagement COG-UK, four nations PHAs and TRE delivery partners. Progressing final approvals to onboard COG-UK data to each national TRE for linkage to routine health data



Welsh SAIL Databank: “SAIL has been able to establish flows and data for the Welsh population relevant to COVID-19 research for over 90% of the in-scope data, with only the sequencing of COG-UK awaiting transfer to SAIL. Several initiatives are already using the priority datasets, such as multi-sector users using the Zoe COVID-19 Symptom Study data.”

3. Making Data Discoverable and Accessible

a) Enhancements to meta-data:

- **38** (+6 from sprint 1) National Core Study priority datasets are now listed on the Gateway with detailed technical metadata for **30** (+8 from Sprint 1)



Northern Ireland Honest Broker Service: “The programme has put us in a much better place in terms of accelerating change of processes, updating meta data and onboarding new COVID related resources.”

b) **Progress with streamlined access management across datasets**

- **NCS Data Access Request form** is now **live on the Gateway**. This form is aligned across all TRE delivery partners meaning researchers can visit the Gateway as a single point of entry to discover and request access to data from multiple data custodians across the UK.

4. **Communicating and Engaging with partners, researchers, patients, and the public**

- a) Collections launched on the Gateway for four of the National Core Studies providing an **openly accessible overview of the research questions, datasets, people, tools, and publications associated with each study**.
- [Longitudinal Health and Wellbeing Collection](#): 3 datasets, 7 projects, 2 papers
 - [Transmission and Environment Collection](#): 19 datasets, 2 tools, 1 project
 - [Epidemiology and Surveillance Collection](#): 7 projects
 - [Immunity Collection](#): 8 projects
- b) Collections launched for two 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](#)
- c) **Communications plan** developed to 1) promote data assets, 2) demonstrate and share impact, 3) be transparent and accessible to the public and demonstrate trustworthiness and 4) build understanding of data capabilities.
- d) During Sprint 2 we have **published** a [news round-up](#), a [blog written by Margaret Rogers](#) (public representative on the delivery group) and a [video interview with Professor Sir Ian Diamond](#) (discussing the aims of the Data and Connectivity Study and the opportunities it provides for the research community).
- e) Consultation to assess **public views on private sector datasets being available as well as data access and use** (88 public contributors) found it would be an overall welcome addition but there are concerns around the quality and completeness of private sector datasets. In terms of private sectors accessing data there were concerns around benefits private sectors would gain and the need to ensure rigorous access processes for private sector organisations to access data.
- f) Consultation on **public concerns regarding data movement across the four nations** (146 public contributors) generally found people were supportive and emphasised the need to ensure research includes data from across all four nations and not just England. However, there were concerns around if and how different infrastructures would be able to work together to ensure the safeguards protecting data are still in place.

- g) Working group with 12 members of our patient and public network to outline ways **to improve communications about data access** and the **safeguards in place to ensure secure access to data** (addressing concerns identified in Sprint 1 public consultation). This includes, but not limited to, creating an FAQs page, signposting to resources that are currently being developed by other organisations (e.g. HQIP) and creating a group of public champions to help share our key messages.
- h) 12 patient and public contributors have contributed to a template to create consistent and accessible **lay descriptions of the datasets**. 4 public contributors are helping to co-create an **accessible lay description of what a TRE**.

5. Engaging with researchers to increase use of data infrastructure

- a. Rapid funding call launched to enable **research which uses and enriches the data within the Data and Connectivity National Core Study**
- b. 48 applications received, with independent panel currently selecting successful proposals (for project launch from 14/12).

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

1. 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. Possible reasons for limited engagement in Sprint 1 include a **lack of awareness** of the data assets available and **limited project resources**. Both aspects will be addressed as a priority in future sprints.
2. There is value in the D&C team providing a **neutral third party** who can convene stakeholders across the health data landscape and drive progress.
3. **A clear overarching NCS narrative** that articulates the public outcomes and benefits that the NCS are collectively working to deliver by 31 March 2021 **is essential**. This will help to guide **prioritisation of linkages, data flows and use cases**.
4. **Legal barriers to data sharing**, and the timeframe required to resolve these, risks our ability to achieve sprint deliverables at pace. We have made important progress on “proof of concept” activities, enabling data sharing from NHS Digital to ISARIC for linkage to the unconsented CO-CIN cohort however IG and legislative barriers will continue to be monitored and escalated to the Oversight Committee where necessary.
5. 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.

Outline for Sprint 3

December Sprint Goal & Scope:

Completion of Sprint 1 and 2 activities AND future Sprint roadmap (in preparation for launch of next 4 week sprint 7th January)



Deliverables

- Completion of Sprint 1 and 2 activities:
 - CO-CIN routine health linkages for Wales and NI
 - ONS/NHS Digital jointly owned COVID data asset
- The following core datasets available to approved researchers in secure cloud-based TREs with demonstrable linkage:
 - Serology testing data (in collaboration with CO-CONNECT – ONS, SIREN, VIVALDI, PHOSP, REACT II data assets)
 - Viral genome data – COG-UK
 - Intensive care data - e.g. ICNARC and SICSAG
 - Zoe Symptom Study linkage to routine health data across 4 nations and enhanced data collection
- Scoping activities
 - Vaccines data collection requirements
 - Researcher funding call – early engagement and accelerated data access
 - Future roadmap for NCS data asset and technical requirements

December Sprint Goal & Scope:

Completion of Sprint 1 and 2 activities AND future Sprint roadmap (in preparation for launch of next 4 week sprint 7th January)



Sprint 2 - Measures of success

Data asset	<ol style="list-style-type: none"> 1. Sprint 1 and 2 priority datasets and linkages available in cloud based TREs and with access for identified researchers undertaking priority approved projects 2. Roadmap developed for: <ul style="list-style-type: none"> • Future Sprints – data asset and technical requirements (federation) • Vaccines data collection
Discovery	<ol style="list-style-type: none"> 4. Sprint 1 and 2 priority datasets and linkages discoverable through the Gateway with standardised metadata and Data Utility evaluation 5. Data access request live on the Gateway and integrated into existing delivery partner system, including a harmonised core set of criteria which are used to identify “safe people” by all TREs
Communications and transparency	<ol style="list-style-type: none"> 7. Ongoing programme of public involvement and engagement to inform study activities 8. Transparent reporting on research questions, available datasets, access requests and approvals through the Gateway 9. Onboarding of D&C funding call researchers

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

Openly available [here](#)

Research topics with new insights generated in last 2 weeks

Health data research on COVID-19 continues to grow, now reaching 1044 (+130) non peer-reviewed pre-prints & 68 (+1) published papers.

Topic	Insights from ongoing studies (links provide further details):
Surveillance & Epidemiology	<ul style="list-style-type: none"> A study shows that the detection of certain protein levels can accurately predict COVID-19 outcome weeks later, and distinguish between those who will be asymptomatic vs experiencing severe symptoms. Using data from the COVID Symptom Study app, disease incidence, prevalence and effective reproduction number were estimated and geographically granular estimates used to highlight regions with rapidly increasing case numbers, or hotspots. This demonstrates that self-reported data from mobile applications can provide an agile resource to inform policymakers during a fast-moving pandemic.
Immunity	<ul style="list-style-type: none"> A study using bespoke online surveys found that risk-mitigating behaviours in people with immune-mediated inflammatory diseases receiving targeted systemic therapies, were responsible, in part, for fewer adverse COVID-19 outcomes compared to patients receiving no systemic treatments. Analysis of data from SARS-CoV-2 PCR-positive results in seropositive and seronegative healthcare workers attending asymptomatic and symptomatic staff testing at Oxford University Hospitals, found that prior SARS-CoV-2 infection that generated antibody responses offered protection from reinfection for most people in the six months following infection. Further work is required to determine the long-term duration and correlates of post-infection immunity. Antibody levels to SARS-CoV-2 wane within months, and faster in younger adults and those without symptoms. Ongoing longitudinal studies are required to track the long-term duration of antibody levels and their association with immunity to SARS-CoV-2 reinfection.
Longitudinal health & wellbeing	<ul style="list-style-type: none"> Using routine health data from c10m patients in CPRD, it was found that interactions with primary care dropped dramatically after first national lockdown, and as of July 2020, engagement with primary care had not recovered to pre-lockdown levels (except for unstable angina and acute alcohol-related events). The largest reductions were for diabetic emergencies, depression, and self-harm, which will have substantial ramifications for these vulnerable patients. The Invasive Respiratory Infection Surveillance Initiative reveals significant reductions in invasive bacterial infections during the COVID-19 pandemic, likely owing to introduction of COVID-19 restrictions and associated public information campaigns, leading to a significant global reduction in life-threatening invasive diseases. The Clinical Record Interactive Search tool was used to examine routine health data of people receiving mental healthcare from a NHS Foundation Trust and found no evidence of changes in psychiatric prescribing, as a result of the marked increase in remote consultation since the epidemic began, particularly among younger patients.
Transmission & Environment	<ul style="list-style-type: none"> A prospective sero-epidemiological study of healthcare workers at a UK teaching hospital using a SARS-CoV-2 immunoassay, found that risk of infection amongst healthcare workers was found to be influenced by COVID-19 working location, role, age and ethnicity. Analysis of data from the OPENSafely cohort found that for c10m adults living with children, there is no evidence of increased risk of severe COVID-19 outcomes – as compared to adults not living with children. This does not support the hypothesis that close contact with children increases risk of SARS-CoV-2 infection.
Clinical Trials	<ul style="list-style-type: none"> A RECOVERY trial participant has become the first person in England to receive convalescent plasma treatment and then donate convalescent plasma. Convalescent plasma is being tested as part of the RECOVERY trial to explore its efficacy in helping people recover than the usual standard of hospital care which all patients receive.

Data & Connectivity National Core Study: COVID-19 dataset availability – 24 November 2020

Significant data custodian engagement across the sprint 2 priority datasets and continued improvement of sprint 1 metadata, but no new additions yet available for access.

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	To be linked				
Pillar 2 Testing data (UK Gov)	To be linked				Missing results prior to 26 Apr – Data quality issue
Primary Care	GPES Onboarding underway	GPES extract – 98% practice coverage, large subset of codes (4bn items) Community Prescribing	Albasoft ESCRO GP 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)					
COVID-19 Clinical Information Network (CO-CIN)	To be linked to 2011 census	Data for linkage in Scottish Data Safe Haven agreed	Limited metadata	Awaiting decision on data access	
Census 2011	Household structure				
Covid Opinions Survey					
Business Impact of Covid Survey	c. 5,000 businesses				
Labour Force Survey	40,000 households, 100,000 individuals				
Intensive Care data		ICNARC – internal use only	SICSAG (updated weekly)	ICNARC – internal use only	
Pillar 3 Testing data (NHS labs)		Internal use only			
Pillar 3 Testing data (Thirva)		Internal use only			
Other Pillar 4 Testing data	VIVALDI, REACT II				
ZOE Symptom Study App Data				UK wide (unlinked) Wales (linked)	
COG-UK Viral Genome				Finalising governance	

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

Data & Connectivity Sprint 2 linkage priorities

Status of COVID-19 projects using the data – 24 November 2020

23 additions to active research taking place takes total over 200. Most new projects enabled by Scottish Data Safe Haven plus ONS projects enabled for first time. Some increase in bottle neck at IG approval and not yet active.

Further information about **Data & Connectivity** can be found [here](#). Including end of **Sprint 1 report**.

Data Readiness: Lessons from an Emergency recently published by **The Delve Initiative**

# 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	13	23 (-4)	42 (+1)	84 (-)	6 (-)	168 (+10)
- <i>a/w researcher</i>	11	16(-4)	Not available	53 (-)	6 (-)	N/A
- <i>a/w data custodian</i>	2	7 (-)	Not available	31 (-)	0 (-)	N/A
Submitted for Information Governance approval	5	10 (+1)	9(+2)	0 (-)	0 (-)	24 (+8)
Approved but not yet active	1	2 (-1)	6 (+6)	2 (-2)	2 (-)	13 (+4)
Active research taking place	6	68 (+1)	60 (+14)	80 (+2)	0 (-)	214 (+23)
Active Number of Researchers	Not yet available	Not yet available	Not yet available	178	Not yet available	
Average time from application to active research	Not yet available	Not yet available	Not yet available		Not yet available	

Participation in key UK wide studies:

- **PRINCIPLE:** 2,068 participants, an increase of around 25% in last 2 weeks following establishment of data flow of Pillar 2 COVID positive test results to support recruitment
- **RECOVERY:** 18,570 participants across 176 active sites
- **CO-CIN (ISARIC 4C)**
 - 105,144 Tier 0 (case report)
 - 724 Tier 1 (single sample)
 - 1,833 T2 (serial sampling)
- **GENOMICC:** 5,989 participants across 210 ICUs
- **COVID-19 ZOE symptom study:** 4,453,193
- **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.