DECOVID

A highly granular, near real time clinical database and research environment from digitally mature NHS Trusts to answer critical questions and improve patient care during the COVID pandemic

Running Title:	DECOVID
Protocol Version:	2
Date:	5/04/2020
IRAS Reference Number:	**282225**
REC reference:	20/HRA/1689
Management Reference:	(to be added)
Funder:	HDR-UK

CONTACTS

Role	Contact
Data Controller	**University Hospitals Birmingham NHS Foundation Trust (UHB), Mindelsohn Way, Edgbaston, Birmingham, B15 2GW**
Lead Applicant for Database IRAS	**Dr Elizabeth Sapey,** Consultant in Acute Medicine and Director of HDR-UK Hub in acute care: PIONEER. UHB
Information Asset owner, Data Custodian and Information Guardian for DECOVID Research Database	**Dr Clark Crawford** Head of Research Governance at UHB NHS Foundation Trust Email: Clark.Crawford@uhb.nhs.uk
Information Asset Administrator	**Suzy Gallier – Head of Informatics Research & Commercial Development and Research Analytics and Business Lead: PIONEER** **Email: Suzy.Gallier@uhb.nhs.uk**
Clinical Leads at Founding data provision partners 1. **University Hospitals Birmingham NHS Foundation Trust and lead applicant for IRAS/ HRA for purposes of database management at UHB**	Dr Elizabeth Sapey, Consultant in Acute Medicine and Director of HDR-UK Hub in acute care: PIONEER.
2. **University College London Hospital**	Dr Wai Keong Wong. Consultant Haematologist and Chief Research Information Officer | CRIO
Scientific Steering Committee Chair	Professor Bryan Williams, Chair of Medicine, Director of UCL Hospitals NIHR Biomedical Research Centre and Research Lead for UCLH
Alan Turing Institute Lead	Professor Chris Holmes. Professor of Biostatistics at the University of Oxford and Alan Turing Institute

Patient and Public Involvement Co-investigator	**Mr Gary Price** Gary.price@eriks.co.uk
Database Project Leads	**Suzy Gallier -Head of Informatics Research & Commercial Development** **Email: Suzy.Gallier@uhb.nhs.uk** **Stephen Perks - Principal Developer** **Email: Stephen.Perks@uhb.nhs.uk** **John Attwood - Project Manager** **Email: John.Attwood@uhb.nhs.uk**

Protocol Approval

	Signature	Date (dd/mm/yy)
Dr Elizabeth Sapey:		
Dr Wai Keong Wong:		

Table of Contents

Abbreviations

AI	Artificial Intelligence
HDRUK	Health Data Research UK
UHB	University Hospitals Birmingham NHS Foundation Trust
RWE	Real World Evidence
CI	Chief Investigator
UCLH	University College London Hospitals NHS Foundation Trust
UCL	University College London
WM	West Midlands
PICs	Birmingham Systems Prescribing Information and Communications System
PPI/E	Patient and Public Involvement and Engagement
A&E/ ED	Accident and Emergency or Emergency Department
ITU	Intensive Care Unit
HDRH	Health Data Research Hub
QA	Quality Assurance
NDOO	National Data Opt Out
RIS	Radiology Information System
SLA	Service Level Agreement
PAS	Patient Administration System
DB	Database
DSA	Data Sharing Agreement
REC	Research Ethics Committee

SNOMED	Systematised Nomenclature of Medicine
SET	Strategic Executive Team
DMC	DECOVID Management Committee
ML	Machine Learning
COPI	The Health Service (Control of Patient Information) Regulations 2002

Table 1: Database Resource

Title of database:	DECOVID
Rationale:	**Generating a highly granular, near real time clinical database and research environment from digitally mature NHS Trusts to answer critical questions and improve patient care during the COVID pandemic.** **To include but not be limited to details of**: • Patient demographics • The acute care journey and process of care patients undergo • The symptoms of COVID-19 • Acuity data including measures of how unwell people are on presentation and during the acute illness (this includes every physiological measurement) • Previous medical and surgical conditions • Previous medications and treatments • Investigations for the acute presentation including images • Treatments provided to the patients (including every medicine administration event and details of organ support) • Outcomes including escalation of care both within (such as move to intensive care) and outside hospital (such as an increase in social care requirements) • Long-term healthcare consequences of COVID-19 in survivors **The Database will also consider datasets from existing COVID-19 research studies/trials, clinical audits and other sources**

	of data deemed valuable in the context of COVID-19 research.
Establishment responsible for the database:	University Hospitals Birmingham (UHB)
Duration:	5 years provisionally
Resource:	Routine acute care data from healthcare providers
Use:	The database will provide data to support research falling into the broad categories of activity which: Characterises the disease Identifies trends in population groups Suggests effective treatment/management strategies Supports decision making by NHS Bodies and health care providers in placing resources to manage the outbreak Supports further research activity, e.g. clinical trial design
Registration:	All inpatients in participating sites including all confirmed, suspected and populations at risk of COVID-19
Inclusion Criteria:	1. All inpatients in participating hospitals including all confirmed, suspected and populations at risk of COVID-19 2. AND ONLY Upon the expiry of the COVID-19 Notice under Regulation 3(4) of the Health Service Control of Patient Information Regulations 2002: 3. Patient has chosen to not opt-out of UK data collection (currently 30th September 2020).
Exclusion Criteria	1. ONLY Upon the expiry of the COVID-19 Notice under Regulation 3(4) of the Health Service Control of Patient Information Regulations 2002: 2. Patients who have chosen to opt out after the COPI COVID-19 Notice has expired.

1.0 Introduction

1.1 Background

In 2019 a novel coronavirus-induced disease (COVID-19) emerged in Wuhan, China. A month later the Chinese Center for Disease Control and Prevention identified a new beta-coronavirus (SARS coronavirus 2, or SARS-CoV-2) as the aetiological agent (1). The clinical manifestations of COVID-19 range from asymptomatic infection or mild, transient symptoms to severe viral pneumonia with respiratory failure, but also headache, diarrhoea, malaise and myalgia. As many patients do not progress to severe disease the overall case fatality rate per infected individual is low, but hospitals in areas with significant community transmission have experienced a major increase in the number of hospitalized pneumonia patients, and the frequency of severe disease in hospitalised patients can be as high as 30%.(2, 3) The progression from prodrome (usually fever, fatigue and cough) to severe pneumonia requiring oxygen support or mechanical ventilation often takes one to two weeks after the onset of symptoms(3). The kinetics of viral replication in the respiratory tract are not well characterized, and there is significant heterogeneity in how the disease presents in terms of severity, time course and clinical manifestations. Most of what we know of COVID-19 comes from outbreaks in China and Italy, but there may be differences in presentation and course within the multi-ethnic communities of the UK which have not been described thus far. Further, although high level mapping of the outbreak within the UK is being captured by Public Health England and other initiatives, there is a lack of in-depth and granular data capture.

A number of initiatives have been set up to capture a broad footprint of COVID-related data (ICNARC, ISARIC), but they do not include the real-time, in-depth data that can be collected from an advanced electronic health record, which includes:

- Up to quarter-hourly observations (heart rate, Blood pressure, respiratory rate, oxygen saturations, oxygen requirement)
- Ventilator settings and real time physiological responses
- Medications prescribed and given and the physiological response to these interventions immediately, in the short term and long term
- Images (CXR, CT scans), ECGs, Echocardiograms

This in-depth data is needed to inform prognostic prediction tools, enable effective clinical care or model the ongoing and likely future impact of COVID-19.

Further, as the COVID-19 crisis builds, the ability to complete clinical research forms manually will decrease. DECOVID will use data from "the back" of the electronic health records, so data capture can continue irrespective of clinical workload.

1.2 COVID-19 Care Provision

There are currently no approved anti-viral or host-directed treatments for COVID-19. Current data from China and emerging data from the UK suggest that patients who require ventilation will require on going ventilatory support for between one week to 21 days. There are no treatments which have been shown to alleviate or shorten this and little shared data to inform best practice in what "routine care" should include. This has led to heterogeneity in care pathways for patients and little understanding as to what is the best model for patient care.

1.3 Non-COVID Care provision during the pandemic

The COVID-19 pandemic has caused widespread reconfiguration of hospital services including the deprioritisation of elective procedures and out-patient services, and a lack of face to face primary care contact. This may cause excess mortality over the next year in the population, both because of deaths among those infected, and because people who are not infected are experiencing changes to their usual medical care as well as social and economic upheaval. The net effect of mortality of this emergency on the population is not only a matter of modelling an infectious disease, but of modelling the mortality effects of wider medical and societal changes.

One way of estimating and monitoring excess mortality is to compare observed numbers of deaths with those expected based on the background (pre-COVID-19) risks of death in the population. No current data collection surveys plan to assess the impact of COVID-19 on non-COVID patients coming to hospital. DECOVID will provide just this. By assessing these data, DECOVID will identify if there are particular groups of non-COVID positive patients who are experiencing worse outcomes due to the COVID-19 pandemic, and feed this back to medical teams – so changes in care provisions can be considered.

1.4 Evidence gaps and uncertainties

Current models of the population mortality impact of COVID-19 are based on age-stratified death rates over days in patients infected with COVID-19 and have not incorporated clinical information from NHS health records regarding prevalence of underlying conditions, their differing background (pre-COVID-19) long term mortality risks, or the impact of differing levels of additional risk associated with COVID-19. There have been limited reports so far of the excess deaths beyond those expected in specific high-risk populations(4). The majority of COVID-19 deaths reported so far have occurred in patients with underlying health conditions or at older ages(5); this situation appears to be changing with severe infections being treated in younger COVID-19 patients without underlying conditions. Case fatality rates (CFR) for COVID-19 vary from 0.27% to 10% according to a meta-analysis(6), possibly explained by differing demography, testing strategies and prevalence of underlying conditions. However, population estimates are lacking, since cases need to be tested and most cases go untested. Moreover, since testing is more common in hospital patients, included individuals are sicker.

There is no publicly available dataset or published research using routinely collected and detailed physiological (every vital signs), investigation (every test ordered and the results) and intervention (every medication (such as drugs or fluids) administered or intervention deployed). DECOVID will provide this.

For example, DECOVID can map the changing the position of the ventilated patient to prone and the real time effect this has on vital signs immediately, in the short and the long term. The same data can be provided for every medical intervention, from drugs to fluids to non-invasive ventilation. Only the most sophisticated electronic health records can provide these data, and initial partners and on-boarding data partners within DECOVID have been chosen for this capability – to supplement the more “top level” data being provided by NHS Trusts across the country.

This powerful depth of data within DECOVID can be used to model and predict inpatient deterioration or inform different treatment strategies such as early or delayed invasive ventilation. It can also predict who can be safely discharged following first presentation to hospital.

Existing research studies are also based on data from manually curated case report forms that by necessity leaves out potentially important pieces of information for example the temporal relationship between an intervention and physiological parameter or blood test result. As the pandemic progresses, it is also likely that manual data collection will become even more sparse.

There is a pressing need for in-depth and high-quality data as quickly as possible to improve patient care now, but also for longer term data to understand the impact of COVID-19 on patients with and without COVID-19 infection(7).

2.0 DECOVID

The DECOVID database will be placed within the infrastructure of PIONEER, the HDR-UK Health data research Hub.

PIONEER is an HDR-UK funded data hub which gathers data from across acute care providers. This has been developed to meet required ISOs for data curation, pseudonymisation, anonymisation and distribution and the DECOVID processes will be aligned to PIONEER processes, as described in detail below.

DECOVID, the PIONEER affiliated database, will collect in-depth, longitudinal data from patients with a confirmed, suspected or at risk of a diagnosis of COVID -19 in near real-time. It will comprise electronic health records for patients diagnosed with COVID-19, suspected of having or at risk of COVID-19 across healthcare providers, including data of process, acuity, physiology, prescriptions, investigations, participation in research studies and procedures. These health records will be drawn from the most digitally mature hospitals within the UK. Data will be curated by one of the most advanced hospital digital systems in the world that has been developed in-house by an award-winning team at University Hospitals Birmingham NHS Foundation Trust (UHB). De-identified data will then be interrogated by the UK academic data science community including members of the Alan Turing Institute, National AI Institute and University College London bringing significant computer and data science expertise to generate algorithms, pathways and insight in how to provide the most effective healthcare for patients with this poorly understood viral illness.

DECOVID will include data from a number of acute care hospitals, starting with a core group but designed with scalability to provide national insight. Understanding the journey of COVID positive patients and those at risk of COVID-19, their disease burden and outcomes will provide critical insights into where we can change practice through research using health data. DECOVID will also aim to link data with manually curated data from case report forms collected as part of other studies with the agreement of the data controllers and sponsor of those studies to further enrich the dataset.

2.1 DECOVID Rationale

The very scale of the COVID-19 pandemic could also provide the means to quickly understand the effects of this virus across the population. By understanding both the acute care experience and long-term health needs of COVID-19 positive patients, and those admitted to hospital during the pandemic, there is an opportunity to identify critical points in delivery pathways where new approaches, treatments and devices might revolutionise care. This would offer significant benefits to the care of individuals, especially in those COVID-19 positive patients with complex care needs and multiple health conditions, where disease outcome and course are harder to predict and outcomes following ventilation generally poor. Providing the Alan Turing Institute access to routinely collected and curated health data will facilitate researchers to understand and access COVID-19 care requirements in depth for the first time. This access will drive innovation.

The potential utility of the DECOVID dataset is vast, with potential benefits including:

1. Better prognostication markers for patients on first presentation and during the course of illness while an in-patient (admitted to hospital)
2. Rapid in-silico assessment of signals emerging/detected from NHS trust sites either within or without the founding partner sites.
3. Pathway innovation to tackle diagnostic delay
4. Modelling of the impact of age, multi-morbidity and poly-pharmacy within the ethnically diverse UK population
5. Identifying specific populations at risk of poorer outcomes (both in patients who are COVID-19 positive and negative) and those most likely to respond to new therapies
6. Assess outcomes of various treatment strategies employed in a new disease providing knowledge gained from a natural 'experiment' outside the confines of a formal clinical trial

And to the wider community

1. Creation of learning health system to enable increased sophistication of resilience/response modelling/planning
2. Up skill the workforce in health data
3. Enable NHS data to solve our own healthcare challenges
4. Have first access to health innovation across providers

With support from patients and the public, using health records for the population and allowing these anonymised data (and in specific and individually justified circumstances, pseudonymised data – see section 2.11.8.1) to be used to understand COVID-19 processes and outcomes and then model and test new approaches could significantly improve COVID-related health care for the nation and help maintain a sustainable and resilient NHS response. This approach benefits from 'big data' – and the 'wider and deeper' the data footprint, the greater the opportunity.

DECOVID will be led by the NHS sites (UHB and UCLH) to ensure robust health data governance at all times with the data controller for the database (UHB) providing final approval for data release. But all decisions will be made in consultation with the founding partners of DECOVID: UHB, University of Birmingham (UoB), UCLH, UCL and the Alan Turing Institute.

2.1.1 University Hospitals Birmingham NHS Foundation Trust (UHB) in DECOVID

UHB is one of the largest NHS Trusts in England, providing direct acute services and specialist care to a Birmingham population of 1.2m, across four hospital sites, in addition to specialist regional and supra-regional services. The Trust has over two million patient visits per annum.

UHB serves the population of the West Midlands (WM), and will become a "receiving centre" for COVID-19 patients from the WM as regional hospitals' ITU facilities become full. With a population of just under 6 million, the WM has one of Europe's largest, most diverse and non-transient populations. WM has the highest birth rate in England and is one of the youngest cities with 40% of the population being under 25. The region faces significant health challenges that impact on regional productivity. Life span is reduced by 1.4 years in females and 1.9 years in males when compared to the South East of the UK. Health span (years spent in good health) is even further reduced with 66.8% of the WM population being obese or overweight, and citizens experiencing high burden of cardiovascular disease, cancer and type 2 diabetes, often at an earlier age than the general population. Poor health drives low socioeconomic status with a high percentage gap in employment rates between those with chronic illness compared to the general population. The West Midlands is also a COVID-19 "hot spot".

UHB is a Global Digital Exemplar site, and has nationally recognised strength in design and deployment of novel digital solutions to support direct clinical care and health informatics capability. UHB's electronic health care record, **P**rescribing **I**nformation and **C**ommunications **S**ystem (Birmingham Systems), developed over the last twenty years, is highly regarded, and offers unprecedented flexibility for rapid response to the changing clinical imperatives in the current clinical setting. **PICS** is

able to capture real time physiological, drug prescribing and administration, investigations and laboratory data, integrated with care processes and patient pathways. UHB also has significant expertise in implementing regional and supra-regional health data systems as part of clinical care (for example the 100K genome programmes' Genie platform deployed across the WM, Great Ormond Street and Devon and Exeter Genomic Medicine Centres, and NORSE, a tertiary referral platform), within a robust data governance framework. UHB houses PIONEER, the HDR-UK Hub in acute care, in which DECOVID will reside, and will use its expertise to rapidly deploy DECOVID.

2.1.2 University College London Hospitals NHS Foundation Trust (UCLH) in DECOVID.

UCLH is one England's leading NHS Foundation Trusts providing first class specialist and acute services in six hospitals in central London. The Trust has over one million patient visits every year across its hospital sites. In partnership with University College London (UCL) UCLH is one of the country's five original comprehensive National Institute for Health Research biomedical research centres with world class research structured into 9 disease-based themes. Although UCLH is a major national and international referral centre, approximately one-third of UCLH clinical activity is from the London Boroughs of Camden and Islington, 2 of the most culturally diverse boroughs in the United Kingdom. 34% of Camden residents and 32% of Islington residents are from black or minority ethnic groups. Both boroughs have particularly large Bangladeshi, Black African, Chinese and Indian communities. Both boroughs have high levels of deprivation. Over 20% the Islington population are living in income deprived households. Islington is the 24th most deprived borough in England and Camden is among the 69 most deprived districts. In common with other inner London boroughs, both boroughs have small but growing communities of migrants who are refugees or seeking asylum, as well as migrants. Given the global city status of London and the diverse and transient nature of its population, London is clearly a Covid-19 "hot spot" and is several weeks ahead of the rest of the UK in the outbreak.

In April 2019 UCLH implemented the Epic electronic health system which has transformed the Trust into one of the country's most digitally mature NHS organisations. Alongside Epic the UCLH Biomedical Research Centre has invested significantly in data science expertise and infrastructure to enable the power of electronic health records to be leveraged for research and clinical care. The BRC's Clinical Research Informatics Unit (CRIU) was established 5 years ago to house the digital and data expertise and develop the platform technologies such as the Experimental Medicine Application Platform (EMAP) that has established a safe and secure analytical environment for data. The BRC has also enabled a strong partnership between UCLH and UCL's world leading Computer Science Department which has significant capabilities in data science and in particular artificial intelligence,

focused around the new UCL AI Centre. UCLH has established links with the Alan Turing Institute. UCL/UCLH will host the Scientific Steering Committee.

2.1.3 Kings College Hospital NHS Foundation Trust (KCH) in DECOVID

The next hospital to join DECOVID will be KCH. KCH is a large NHS Trust spanning two acute sites in South East London (Kings College Hospital, Denmark Hill and Princess Royal University Hospital, Bromley) providing tertiary specialist services throughout south east London and Kent. It hosts a Major Trauma Centre, two Hyper-acute Stroke Units, haematological and liver malignancy specialist units. It is part of the Kings Health Partners (KHP) Academic Health Sciences Centre, together with Guys & St Thomas' Hospital NHS Foundation Trust (GSTT) and South London and Maudsley Hospital NHS Foundation Trust (SLaM).

National Covid monitoring shows that Southwark and Lambeth are regularly in the top 3 London boroughs affected, making it a rich source of data of ethnically mixed heterogenous populations. The local boroughs served by KCH is from Lambeth, Southwark and Bromley, making KCH an ideal site for COVID research.

In 2016, KCH deployed a cutting-edge data pooling and search engine called Cogstack in partnership with SLaM BRC in Bioinformatics, and then progressed to deploying Natural Language Processing (NLP) tools to handle the large volumes of unstructured clinical data in the electronic health record. It has constituted a patient-led committee for governing real-world data research (KERRI), and KCH as partner of the InnovateUK London Medical Imaging AI for Value-based Healthcare (AI4VBH), has on-site high-end computing for machine learning. Due to the informatics technologies on-site, KCH has near real-time analytical pipelines for data research.

DECOVID will gather routinely collected patient data from at risk, confirmed and suspected COVID-19 patients from across these acute care providers and more – to build the first granular and holistic data-record combining routine acute care provision with unparalleled detail and data granularity, centred on patient benefit. The data will be anonymised for research and innovation purposes.

This positions DECOVID as an exemplar for making the NHS 'AI-ready' in an area and time of critical clinical challenge. DECOVID, utilising the skills and expertise of the Alan Turing Institute and other data science institutes, will support innovation, making existing but inaccessible datasets discoverable,

and bringing scale and efficiency to dataset aggregation and curation of anonymised routinely collected COVID-19 patient data, bringing insight and innovation into clinical practice.

2.2 Aims

The aim of this database is to curate routinely collected data from patients admitted during the COVID-19 pandemic and use this in an anonymised form (but in a pseudonymised form in specific, individually justified and individually considered projects as described in section 2.11.8.1) to innovate health care provision in this over-burdened and challenged healthcare sector, to answer specific research questions which have been prioritised by the UK government and are of relevance to patient care.

2.3 Objectives

DECOVID will support the following objectives:

1. To develop a research database to understand and inform healthcare processes for patients admitted to hospital during the COVID pandemic which can inform current and future patient care and understand the healthcare processes needed during and after this time of clinical need.
2. Work with healthcare providers, patients, public and other stakeholders to ensure that the design, development and governance of data access through DECOVID are in the public interest, and that these principles are communicated effectively to researchers, patients and the public.
3. Bring scale and efficiency to dataset aggregation and curation of anonymised (and in some, individually considered and specifically justified circumstances, pseudonymised) routinely collected health data relevant to COVID-19.
4. Make these and existing inaccessible datasets discoverable and appropriately accessible to the Alan Turing Institute and other data scientists to conduct research which will lead to direct patient benefit (research and innovation) for future care for both COVID and non-COVID positive patients.
5. Provide a forum for cross-sector collaboration with strong relationships between NHS/academic consortium members to support research, development and innovation in COVID-19 care.

2.4 DECOVID Design

DECOVID is the name of the database which will collect and link data from national acute care providers.

Initially, the DECOVID Research Database will hold health data from 2 NHS trusts:

- University Hospitals Birmingham NHS Foundation Trust;
- University College London Hospitals NHS Foundation Trust

Then, DECOVID will then identify additional datasets and data collection centres to provide greater value of the database for COVID-19 based research and ultimately patient benefit. All direct patient data collection centres will operate within the same mechanism as described below. Kings College Hospital is the first partner hospital to be included.

Patient's data will be collected as part of their routine care when seeking unplanned medical assistance during the COVID-19 pandemic. No additional data will be collected. However, data collected as part of other existing and relevant studies may be included with permission from their respective data controllers and sponsors and where consent has been given by the studies participant (as needed). Initially, any acute care contact during the COVID pandemic from UHB or UCH will be the initial trigger for DECOVID data collection. From that time point, the acute care journey of those patients are mapped retrospectively and prospectively to provide a clear and detailed picture of preceding symptoms and health care problems, and prospectively over time, to determine changes in healthcare utilisation after an acute care presentation.

Ultimately, acute care contacts from other health data providers during this pandemic will also trigger data curation, making an ever more complete dataset of acute care provision nationally. These healthcare contacts are unpredictable, and so no minimal or maximal timelines for data acquisition will be set.

The DECOVID team (as described in this document) within UHB will lead the design and construction of the database (construction configuration, implementation, QA testing) and ensure secure web hosting.

The DECOVID team within UHB will facilitate data harvesting, data curation, pseudonymisation and anonymisation and de-identified data provision to named and approved personnel approved personnel within the Alan Turing Institute and other researchers using a rules-based approach as

described below (see section 6). The UHB and UCLH Clinical Research Informatics team will provide expertise for clinical data standardisation and alignment with the OMOP standard.

2.5 Patient and public engagement and involvement in data sharing

The theme of the use of health data within Acute Care services was developed in workshops including 168 members of the public, patients and healthcare providers. We held three separate workshops; one for patients with chronic illness who were frequently healthcare "users" and their carers; one for members of the public who had not accessed secondary healthcare frequently and one for NHS hospital staff and GPs. They were asked to consider which parts of healthcare provision needed most improvement.

People identified:

1. Unplanned healthcare contacts as the most negative experience within the NHS, noting the lack of new approaches and delays in acute care due to overstretched front door services.
2. Research needed to be more inclusive to address acute health concerns of our ageing, multi-morbid patients.
3. Research needed to be more inclusive across national sites, to understand and improve acute healthcare in geographical areas of greatest need.
4. Research should benefit all ages
5. Improvements in acute care was the main priority for health innovation (including new ways of accessing healthcare, admission avoidance, hospital care at home, ambulatory care, tracking their own health and new therapeutic approaches).

We asked the same 168 people about their thoughts on health data use. After discussing real world examples of how health data had improved aspects of care, 99% of participants were happy for their pseudonymised or de-identified health data to be used in research for patient benefit by non-NHS and academic institutions. After discussing real world examples of how health data had improved non-healthcare services (public transport or local services), 96% of participants were happy for their health data to be used in non-health related research for public benefit. After discussions about the type of researchers who may request access to health data, the principles of GDPR; identifiable data, pseudonymised data and de-identified data and principles of appropriate data sharing, 100% of participants were happy to have their de-identified health data

to be accessed by NHS staff not directly involved in their care; 98% by academic researchers not involved in the NHS and 96% by industry, if the data would improve health or care for other patients or members of the population.

Since this, UHB has discussed specifically the use of de-identified data without explicit consent, with >300 members of the population, including > 40 children aged between 13 and 17 (as the national data opt out includes children aged 13 and over).

The results of this consultation are that the following percentage of patients would be happy for their de-identified health data to be used, without their explicit consent in the following circumstances:

- 99% for research which improves NHS services
- 98% for research undertaken by healthcare staff
- 97% for research undertaken by academic staff not connected to the NHS

These initial consultations have informed the design for DECOVID and provided a structure for meaningful PPI/E within DECOVID (see section 8).

2.6 Transparency in DECOVID operations

DECOVID will provide data in the public domain regarding its operation and purpose. We aim to publish this protocol once finalised, as evidence of this.

2.6.1 Privacy notices provided by the data controllers

The data controller and data providers will provide information through their research privacy notices.

The controller's privacy notices may be found at:

https://www.uhb.nhs.uk/privacy-notice

https://www.uclh.nhs.uk/aboutus/Pages/Cookiepolicy.aspx

and KCH privacy notice can be found at:

https://www.kch.nhs.uk/Doc/mi%20-%20222.1%20-%20privacy%20notice.pdf

2.6.2 Lay summaries and updates

All data requests will be expedited to face this pressing health crisis. Researchers will be asked to provide a lay summary of data outputs with the assistance of clinical teams within DECOVID. These lay summaries will be published on the HDR-UK website after scrutiny by the DECOVID team and a PPIE representative to ensure that it is a readily understandable and accurate representation of how data is being used to help patient care.

There will also be transparency in the process for evaluating which data is to be released. This includes the standard criteria by which applications are assessed; including public good, the "5 safes" and open access policies. See below in section 5.0 and 6.0.

2.6.3 Record of data requests to DECOVID

A list of all data requests to DECOVID will be available on request, updated on a six-monthly basis. The summary for each data request will include the lay summary.

2.7. DECOVID Population

Patients who have been admitted acutely to hospital during the COVID-19 pandemic within a health data partner. Since there is a critical need for acute health care innovation which is ageless in approach, there will be no upper or lower age limit for data inclusion.

2.8 Main Inclusion Criteria

1. Acute admission to hospital during the COVID-19 pandemic within DECOVID health data partners.
2. Given Department of Health and Social Care COPI COVID Directive – until September 2020 or until stated by the Department of Health – all patient data irrespective of opt out
3. Following the end of the stated directive of the Department of Health and Social Care COPI COVID-19 Directive (planned for end of September 2020 or as per extended directives), patients will only be included if they have chosen not opt-out of the use or disclosure of their data for research and planning.
4. Aged 18 years of age or older (no upper age limit)

2.9 Main Exclusion Criteria

1. From September 2020 or when the stated directive of the Department of Health and Social Care COPI COVID-19 Directive ends - Patients will be excluded if they have chosen to opt out of the use or disclosure of their data for research and planning.

2.10 Identifying Potential Participants

Patients will be identified by each recruiting hospital from their acute care records.

2.11 DECOVID Patient Process

All steps referred to within 2.11 are presented in Figure 1 – the DECOVID Dataflow Process where examples of 3 data flows are given. For each example, the steps refer to the numbers in yellow circles within the data flow diagram.

2.11.1 Processing patient identifiable data without explicit written consent

Section 251 approval is currently not being applied for but the DECOVID database will curate health data without obtaining explicit written consent.

The rationale for not seeking informed consent is that we wish to:

- Include as many people as possible. We need to include all patients who have had a COVID-19 related acute care contact across the UK to understand the burden of this pandemic and its long-term effects. However, it is recognised that the COVID-19 pandemic will impact on non-COVID related health outcomes as well, as resources are diverted from standard clinical care. There is a need to understand and map both impacts. There is also great interest, in time, in including international datasets, so we can benchmark our services and outcomes against the best and worst performing sites internationally, to learn where our services can be improved. Including these numbers is vital to allow an in-depth study of the impact of COVID-19 on health care across the nation, which can provide national and international insight into COVID-19-associated care challenges.
- Include a population that is fully representative of the patient population, which cannot be achieved from usual research cohorts.
- Include data from patients who have died. Mortality for admitted patients with COVID-19 is high. We wish to include data from people who have died following COVID-19 infections.

Mortality is also expected to go up from non-COVID-19 related health complaints, and so data from non-COVID-19 patients will also be critical to understanding the impact of COVID-19.

- Include people who may not have capacity to consent so that the COVID-related health journeys of more vulnerable adults also have the potential to benefit from innovation. Delirium appears common in older people with COVID-19 infection. Currently, visitors are restricted in NHS hospitals, so there will be no ability to consent relatives, and the distressing nature of COVID-19 infections places an undue burden on relatives if assent was sought for study participation.

The scale of these data and the inclusion of data from people who have died prevent the gaining of informed consent for data use, as would be the usual standard. No additional data to that collected as part of standard of care is requested and all data will be accessed to fulfill the research request in a format where patient identification is highly unlikely by the researcher (see sections 5 and 6 of the protocol).

The rationale for not seeking CAG approval at the current time is that on 23rd March 2020 the Secretary of State for Health and Social Care has given legal notice to individual healthcare organisations to support the processing and sharing of information to help the COVID-19 response. This is to ensure that confidential patient information can be used and shared appropriately and lawfully for purposes related to the COVID-19 response. Organisations are only required to process such confidential patient information where the confidential patient information to be processed is required for a Covid-19 Purpose and will be processed solely for that COVID-19 Purpose in accordance with Regulation 7 of COPI and from the date of this Notice until 30th September 2020.

A COVID-19 Purpose has been defined in writing to include but is not limited to the following:

● understanding COVID-19 and risks to public health, trends in COVID-19 and such risks, and controlling and preventing the spread of COVID-19 and such risks;

● identifying and understanding information about patients or potential patients with or at risk of COVID-19, information about incidents of patient exposure to COVID-19 and the management of patients with or at risk of COVID-19 including: locating, contacting, screening, flagging and monitoring such patients and collecting information about and providing services in relation to testing, diagnosis, self-isolation, fitness to work, treatment, medical and social interventions and recovery from COVID-19;

● understanding information about patient access to health services and adult social care services and the need for wider care of patients and vulnerable groups as a direct or indirect result of COVID-19 and the availability and capacity of those services or that care;
● monitoring and managing the response to COVID-19 by health and social care bodies and the Government including providing information to the public about COVID-19 and its effectiveness and information about capacity, medicines, equipment, supplies, services and the workforce within the health services and adult social care services;
● delivering services to patients, clinicians, the health services and adult social care services workforce and the public about and in connection with COVID-19, including the provision of information, fit notes and the provision of health care and adult social care services; and
● research and planning in relation to COVID-19.

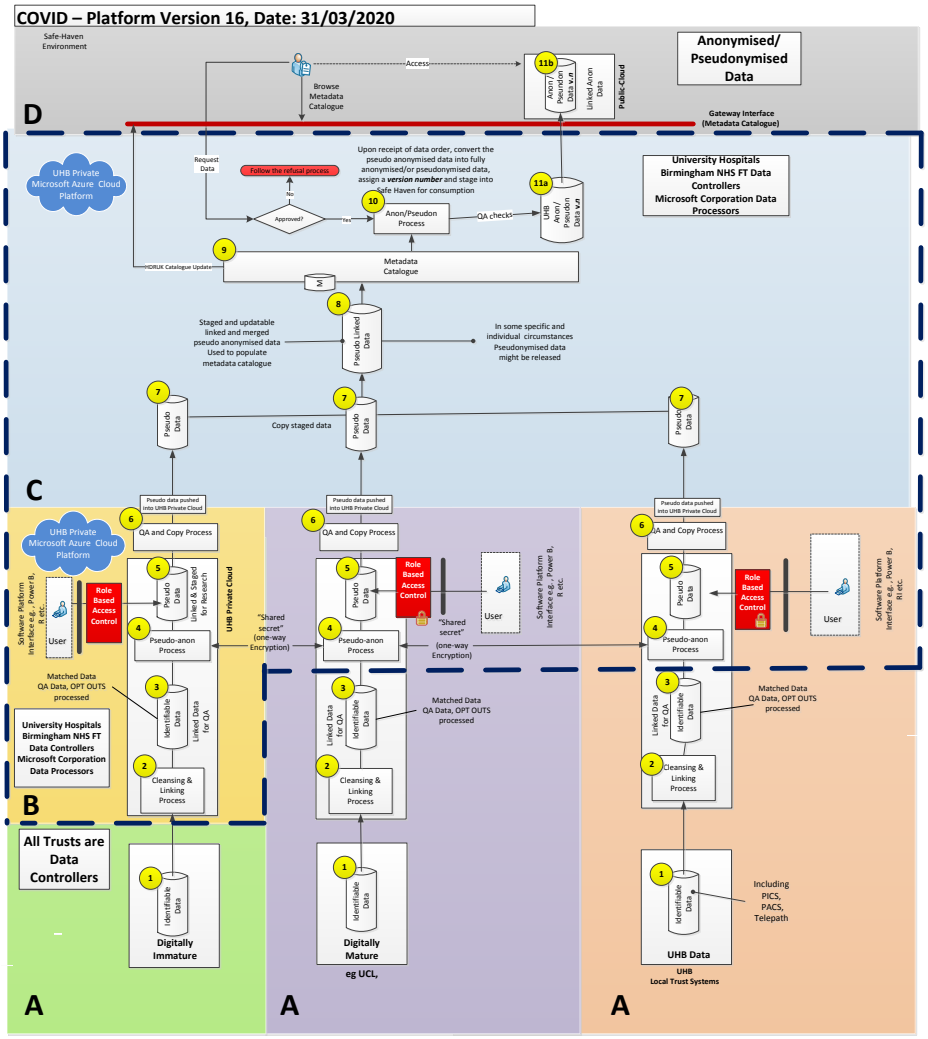
DECOVID collects data specifically related to the impact of the COVID-19 pandemic and thus meets the above requirements as a legal basis for data collection processing.

The same Department of Health dictate also releases data providers from removing "opted out" patients from data sharing until September 2020. Therefore, CAG application is not necessary at the current time but will be sought during this period and opted out patients will be removed after the period has ended as per DHSC directives.

As this is an important consideration, the use of data without explicit consent was specifically discussed with both 168 members of the public and with >300 people to specifically test if the majority of the public would support data use in this way – see section 2.5.

Patient identifiable data is being processed by the care provider as part of usual healthcare processes and within healthcare governance. The diagram clearly shows where data would then be used for the purposes of research and following approval of the DECOVID team and data controller(s) using the procedures outlined in the protocol. The pathway below describes processes.

Figure 1. Data Flow diagram for DECOVID



- For all examples of data flow, Section A is where the individual NHS Trust who collected healthcare data is Data Controller.
- For Trusts which cannot pseudonymise their own data, Section B is where UHB becomes Data Controller
- For all examples of data flow, Section C is where UHB is Data Controller
- Section D is where data is anonymised specifically for individual research projects as approved by processes outlined in section 6.0. This is the default position for data release
- In necessary and individually justified and approved circumstances, researchers may have access to pseudonymised data – the rationale and process for this are listed below in section 2.11.8.1.

2.11.2 UHB NHS Foundation Trust (UHB Data Controllers)

Internal data collected as part of routine clinical care is pooled from across the multiple UHB data systems that hold different types of imaging or other patient centred data as part of usual routine healthcare provision (Step 1). Data is cleansed and linked to ensure quality for clinical purposes (Step 2). The data will be checked in an identifiable form for QA purposes and any patients who have "opted

out" of data sharing will have their record removed (Step 3) after the COVID-19 Notice has expire on 30th September 2020.

The dotted line signifies the start of the research area within the flow sheet and this is where the research protocol begins. Data is pseudonymised using a confidential hash, which refers to a one way cryptographic algorithm that makes it computationally difficult to reverse engineer and identify patients without specific knowledge of the algorithm, which the Data Controllers for the Trusts and the database Data Controller (UHB) will retain to allow for data refreshing. QA checks will ensure data accuracy and validity following pseudonymisation (Step 4-6). Data will remain in the pseudonymised state at all times. See freedom of information request principles (section 2.11.9). Pseudonymised data will remain on UHB servers during the QA checks. Pseudonymised data within this area may be processed for purposes including research, quality improvement projects, audit and service evaluation by UHB staff under role-based access control to improve UHB hospital services and processes (step 6). Data will then be moved to the secure and private Microsoft Azure cloud and will be pooled from all hospital sites (Step 7) and combined (Step 8). These data will be used to develop the metadata catalogue (Step 9). Here the data remains until an approved request is received. The pseudonymised data will be refreshed with data up-dates to gather more longitudinal data and an increased number of patients' data.

2.11.3 Example of Digitally Mature Trust (Non UHB Health Data Partner with ability to pseudonymise data)

Internal data is pooled by the health care partner for routine clinical practice as a data controller (Step 1). Data is cleansed and linked as part of routine clinical care, as described above (Step 2). The data will be checked in an identifiable form for QA purposes and any patients who have "opted out" of data sharing will have their record removed (Step 3) after the COVID-19 Notice has expire on 30th September 2020.

The dotted line signifies the start of the research area and this is where this research protocol starts. Data is pseudonymised using a confidential hash, this will be shared with UHB to permit data refresh as patients utilise care services across healthcare providers (Step 4-6). Following further QA checks to ensure data accuracy and validity following pseudonymisation conducted by the data provider who will still be the data controller at this point (step 5). Then pseudonymised data will be provided to UHB (Step 6) and at this point, UHB will become the Data Controller and the healthcare partner will

act as Data Processor. Pseudonymised data will be moved to the private and limited access Microsoft Azure UHB cloud and data will be pooled from all hospital sites (Step 7) and combined (Step 8). These data will be used to develop the metadata catalogue (Step 9). Here the data remains until an approved request is received. The pseudonymised data will be refreshed with data up-dates to gather more longitudinal data and an increased number of patients' data.

2.11.4 Example of Trust which is not digitally mature (Non UHB Health Data Partner without ability to pseudonymise data)

Some potential healthcare data partners lack the digital maturity or staff capacity to be able to pseudonymise their own health data at pace or scale. To enable their participation, a third way for data inclusion to DECOVID has been developed.

Internal data is pooled by the health care partner who will be the data controller as per usual clinical practice and as part of routine clinical care (Step 1). From step 2 onwards, data use is for research purposes and the research protocol pathway is initiated. Identifiable data will be sent in a selected and staged manner to the private and limited access Microsoft Azure UHB cloud. At this point, UHB becomes the data controller and the health care data provider is the Data processor (Step 2). This is included in the dotted line within the figure as this represents research activity. Data is cleansed and linked (Step 3). The data will be checked in an identifiable form for QA purposes and any patients who have "opted out" of data sharing will have their record removed (Step 3) after the COVID-19 Notice has expired on 30th September 2020. Data is pseudonymised using a secret one-way hash (Steps 4 and 5). QA checks will ensure data accuracy and validity following pseudonymisation (Step 6). Pseudonymised data will be held in the private and limited access Microsoft Azure UHB cloud system and data will be pooled from all hospital sites (Step 7) and combined (Step 8). These data will be used to develop the metadata catalogue (Step 9). Here the data remains until an approved request is received. The pseudonymised data will be refreshed with data up-dates from the NHS Trust to gather more longitudinal data and an increased number of patients' data.

Of note, individual data controllers may choose to restrict data flows after the pandemic, but the aim of DECOVID is to provide a longer-term platform to understand the impact of the COVID-19 pandemic on patients outcomes for at least the following year.

2.11.5 Data sensitivity

Data sensitivity: The data within the DECOVID Research Database is data relevant to an individual's systemic health. A number of highly sensitive types of data relevant to health are excluded, notably data on sexual health and sexual orientation.

Risk of identification: The data within the DECOVID Research Database is pseudonymised data and in normal operation only anonymised data is released. However in certain circumstances which are justified as necessary to ensure the integrity of the data, pseudonymised data will be released with additional requirements on data requestors (see section 2.11.8.1).
With regard to class of identification:

- Direct identifiers
  - The DECOVID Research Database does not contain:

    direct and recognisable identifiers such as name, address or image of a face
  - The DECOVID Research Database does not contain:

direct but not recognisable alphanumeric identifiers such as NHS number
  - The DECOVID Research Database does contain:

images that are not recognisable but may be unique

- Indirect identifiers
  - The DECOVID Research Database does contain:

post code, age, ethnicity and gender

diagnoses including rare diseases

DE-COVID randomly generated ID to replace NHS Number

Risk will be managed proportionately with regard to providing access to any data that might alone or through combination lead to identification of an individual. Specific examples include:

- Post code: the DECOVID Research Database holds postcode data to support studies into equity of access and enable greater understanding of the health impacts of social deprivation. To reduce risk, however, access will not be provided to the post code directly, but rather DECOVID will provide the required linked data on demand and provide it as part of the anonymised dataset, for example providing a less specific geographical unit such as the Lower layer Super Output Area (LSOA) or the associated data of interest such as the Index of Multiple Deprivation score. This approach reduces risk whilst ensuring that the research value of this data is not compromised.
- Age: data of birth is not provided to reduce likelihood of identification; age is provided to the nearest year.

- Diagnoses including rare diseases: a rare diagnosis may enable identification if combined with enough additional indirect identifiers; this will be evaluated on a case-by-case basis and appropriate restrictions will be placed on accompanying data (such as the specificity of any age or geographical data provided) that might significantly increase the risk of identification.
- Data in combination: the combination of enough data fields will at some point result in a unique profile for an individual. This provides a theoretical risk to identification, but such identification is still only possible if that same set of data is provided from some other source. Such datasets are not in the public domain, making this risk extremely low.

A general principle of DECOVID is that data made accessible should be necessary and proportionate to the purposes required i.e, there is data minimisation. When requesting access to a dataset, the applicant must justify the inclusion of each data field. DECOVID reserves the right to refuse an application or limit the data fields available based on concerns around possible identification.

2.11.6 Process of pseudonymisation

This is a technical process of replacing person identifiers in a dataset with other values (pseudonyms) from which the identities of individuals cannot be intrinsically inferred. DECOVID maintains an association between the original value and replacement value. Examples of this process are replacing an NHS number with another allocated random number curated within DECOVID. The allocated number has been generated using a specific encrypted 'salt code' added to this before the combined data is then encrypted using a SHA2-256 hashing algorithm.

Despite this process, the very nature of the linked data across primary care and secondary care providers, even when pseudonymised, means it may not require considerable effort to potentially identify a patient. For this reason, only anonymised data will be shared with researchers, as outline in the data flow diagram, unless there are specific reasons why pseudonymised data are needed (see section 2.11.8.1).

Data will not be currently not be shared with commercial entities and CAG approval would be needed for this to occur.

2.11.7 Metadata catalogue

A metadata catalogue is produced, detailing a summary of the data available. A copy of the metadata catalogue will be placed on the HDRUK metadata catalogue. Applicants may browse the HDRUK Metadata catalogue and submit requests for data via the Gateway or apply directly to DECOVID.

The metadata required by HDRUK is pre-determined and utilises the MoSCoW rating, which seeks to categorise users requests into 'Must have', 'Should have', 'Could have', 'Won't have'. We are required to provide the 'Must' and will aim to provide as much of the remaining information as requested, which comprises summary level data.

2.11.8. Process of anonymisation/ data minimisation

On receipt of an approved request for anonymised data, the requested data will be extracted from the pseudonymised data hub (Step 10) and anonymised (Step 10 - 11). When data is used for purposes beyond individual care and treatment it is normally anonymised, which means that information that identifies an individual patient has been removed. The intent of anonymisation is to turn data into a form which does not directly identify individuals and where re-identification is not likely to take place.

This is a technical process of replacing person identifiers in a dataset with other values from which the identities of individuals cannot be obtained. DECOVID does not maintain any association between the original value and replacement value. Examples of this process are replacing an NHS number with another allocated random number.

Anonymised datasets will be created on demand for specific data requests and projects and will be held within the DECOVID data safe haven (Step 11) for the applicant to access but this will not be retained and DECOVID will not retain a copy of any data that was supplied to/ shared with any user.

The anonymised data will be version controlled, so that researchers can request updated data to reflect the increasing numbers of patients seeking acute care over time. The pseudonymised data process will allow linkage of data records (for example, if a patient has multiple admissions over the data collection timeframe) and this new anonymised data set will be provided to the researchers.

2.11.8.1 Access to pseudonymised data by researchers

There are circumstances where pseudonymised data may need to be accessed by researchers for specific activities.

These include:

QA checking and further data harmonisation.

Checking modelling predictions against real patient data outcomes in an ever-refreshing data model

Answering specific research questions to feedback into clinical care which might benefit an individual

Full justification for pseudonymised data must be given, to ensure data minimisation.

Each pseudonymised data request will be considered by both the SCC and DMC, including lay members.

Where pseudonymised data is requested, the anonymisation process will be followed with the exception of data which has been identified by the applicant and agreed by the SSC and DMC.

All granted requests for access to pseudonymised data will be reported to the ethical committee and publicly reported for transparency.

2.11.9 Opt Out Process for implementation after COVID-19 DHSC notice period expires

DECOVID will comply with the NHS Digital "National Data Opt-out" policy after the COVID-19 Notice has expired on 30th September 2020, using the process from NHS Digital outlined below. Data subjects will be informed of the process for 'opting out' via the Trust privacy notice; Patients who wish to be excluded can opt out either via online or phone registration. This is recorded on the Spine by NHS Digital and we will cross check against this prior to data being utilised for research purposes.

The national data opt-out was introduced to give patients a choice on how their confidential patient information is used for purposes beyond their individual care. The information that the opt-out applies to is special category data as it includes information about a patient's health care and/or treatment that has been collected as part of the care we provide for the patient." DECOVID follows the NHS Digital process so that, "patients can set or change their national data opt-out choice using an online process or contact centre service. When a patient sets a national data opt-out it is in held in a repository on the NHS Spine against the patient's NHS number".

The Data controller for each dataset will be asked to check to see if any patients have opted-out of data use, however, it is recognised that some smaller data providers may not have the expertise to perform this and that patients may choose to 'opt out" after their data has entered DECOVID. In accordance with the patient's wishes and national data opt-out policy, as a health and care

organisation located in England, DECOVID is "required to apply national data opt-outs when applicable to a use or disclosure of confidential patient information for purposes other than the patient's care or treatment".

In line with the NHS Digital process, DECOVID will check, by using the NHS numbers of patients, whether a patient has registered an opt-out before the data is used/disclosed. To do this a separate list of the NHS numbers in the data that is going to be used/disclosed needs to be created. The list of NHS numbers is then submitted to the Check for National Data Opt-outs service via the secure Message Exchange for Social Care and Health (MESH) messaging service. The Check for National Data Opt-outs service is an external service provided by NHS Digital. The service checks the list of NHS Numbers against a list of opt-outs created from the repository on the NHS Spine. Where a match is found it removes the NHS number from the list and then returns an updated list of NHS numbers (with opt-outs removed) back to UHB via MESH.

DECOVID will then match the updated list of NHS numbers against the original set of data that was going to be used/disclosed. The records that need to be removed can then be identified and the entire record for those patient records can then be removed. This creates a 'cleaned' set of data with opt-outs applied that DECOVID can then use/disclose. If a patient chooses to opt out after data processing has occurred then their record will be removed provided a link to this record still exists (ie if this is pseudonymised data). The opt out does not apply to fully anonymised data since at that point there is no link back to the patient from which it derived.

2.11.10. Freedom Of Information Act Principles

- DECOVID will process any Freedom of Information (FOI) Act requests to meet all requirements.
- Each FOI request will be considered individually.
- DECOVID cannot eliminate the risk of re-identifying the individual following pseudonymisation without agreed data sharing agreements in place. Further DECOVID would not be aware of additional datasets that a FOI requestor may hold. Therefore, all DECOVID data will be treated as personal data and not subject to FOI requests.
- All releases of data are for a specified purpose and the use of the data is restricted by conditions specified within the Data Sharing Agreement (DSA).

- Any attempt by a receiving organisation to re-identify any patients whose records are provided in pseudonymised form would be considered a breach of the Data Protection Act and the DSA
- Anonymised datasets will be created on demand for specific projects and will be held within the DECOVID data safe haven (number 11b) for the applicant to access but DECOVID will not retain a copy of any de-identified data that was supplied to any user.

3.0 DECOVID Schedule

Due to the quickly emerging health challenges associated with the COVID-19 pandemic, there is an expedited pathway for DECOVID set up and delivery. The target schedule for the setting up of the database is as follows:

3.1 Milestone 1: Database Established

Phase 1: March 2020 – April 2020

The following will be completed within the first month:

- Consortium Agreement in place with all key initial partners
- DECOVID protocol agreed and submitted for expedited ethical approval through approved COVID-19 pathways
- Initial data Platform ready for sharing and access
- Data Sharing Agreement with Alan Turing Institute named personnel
- Scientific Steering Committee established including Constitution, Terms of Reference, Membership, Modus Operandi, Powers, Memoranda of Understanding and Codes of Practice (see section 6).
- Pilot projects serviced for data sharing
- Data from at least 3 sites normalised and made ready for analysis

3.2 Milestone 2: Delivery and expansion

Phase 2: May 2020 – September 2020

Milestone 2 (Service Delivery)

- Further projects serviced for data sharing
- Evidence that the quality of the datasets has been improved (curated) and that the curated data is discoverable through the Gateway.
- New data provider partners added to DECOVID with data sharing agreements in place
- Provide publishable enhanced service case studies that demonstrable impact (and expected impact) and value to the NHS and patients globally.

3.3 Milestone 3: Expansion and Sustainability

Phase 3: October 2020 – March 2025

- A programme of research serviced by data sharing
- New national and international data provider partners added to DECOVID with data sharing agreements in place
- Changes in clinical care processes based on DECOVID outputs
- Published studies that demonstrable impact (and expected impact) and value to the NHS and patients globally.
- Provide evidence that DECOVID is continuing to engage and involve patients and the public in a meaningful manner.
- Partners and collaborators will be engaged early and often and relevant partners responsible for implementation will be integral to goal setting and planning for each delivery iteration to ensure a common understanding and commitment towards targets.

4. Data Management

4.1 Data Collection

Data will consist of routine, pre-existing acute care data. This will include demographic data, data of care processes (time acute care presentation, first assessment, first investigations and treatment, time to discharge, grade of staff, place of care) and health care delivery including investigations and treatments, diagnosis and onward care plans). Investigations will include imaging (radiographs, computerised tomography, Magnetic resonance images etc) as well as physiological data captured as reports and images (such as electrocardiograms and echocardiograms). The images are already stored on the trust local servers in data warehouses within the trust's own legal entity and can be moved to the DECOVID data set. Within the UHB data warehouse data is stored from sources PAS and Medisoft. See section 2.11.4 for special category data.

Quality assurance of the data will take place within the local trust servers and it is at this point that any records for patients who have opted out will be removed after the COVID-19 Notice has expired. The data will only be pseudonymised when these processes are complete. Identifiable patient data is then pseudonymised using a one-way hash. before transfer to DECOVID.

Pseudonymised data is stored on a Research Data Platform database under the data controller (UHB). Pseudonymised data undergoes quality assurance and copying, it then enters a Private Microsoft Cloud Platform. Refer to Figure 1 DECOVID Dataflow Process for the data collection process.

5.0 Source Data and Documents

5.1 Data Handling and Record Keeping

Data will be submitted directly to a secure cloud-based environment, maintained on the Microsoft Azure cloud platform in accordance with the UK Cyber Cloud Principles which are outlined here:
https://www.ncsc.gov.uk/collection/cloud-security?curPage=/collection/cloud-security/implementing-the-cloud-security-principles
The cloud provision will follow the standards below:

ISO 27001
An international specification for information security management. The corresponding code of practice is ISO/IEC 27002.

ISO 27017
Code of practice for information security controls based on ISO/IEC 27002 for cloud services

ISO 27018
Code of practice for protection of Personally Identifiable Information (PII) in public clouds acting as PII processors

The database platform complies with the Department of Health Information Governance policies and standards for secure processing of patient healthcare data as set out in the Information Governance Toolkit of the Health and Social Care Information Centre.

5.2 Data Validation and quality

Data will be cleansed and matched in each trust's local server as per usual data controllership activities. Data cleansing is the process of detecting and correcting (or removing) corrupt or duplicate or inaccurate records from a record set, table or database. It refers to identifying incomplete, incorrect, inaccurate or irrelevant parts of the data and then replacing, modifying, or deleting the dirty or coarse data.

Secondly, the data will be normalised, this is the systematic process to ensure the data structure is suitable or serves the purpose. Here, the undesirable characteristics of the data are eliminated or updated to improve the consistency and the quality. The goal of this process is to reduce redundancy, inaccuracy and to organise the data. The data will only be pseudonymised when these processes are complete. OMOP has been identified as target data model as we feel that it provides the correct balance between the ease of data normalisation and preparation for advanced analysis.

Quality assurance of the data will take place within the local trust servers prior to transfer to Microsoft Azure Clouds, and during the anonymization process in the Shared Private Microsoft Cloud. Quality checks include but are not limited to checks as to whether the record counts are correct to expectations; whether mandatory fields are populated; whether the primary and foreign keys work; appropriate use of SNOMED codes and whether the pseudonymisation was successful.

The project will require ongoing access to clinical systems, by doing so it will support maintenance of accurate data. Data will be either refreshed (pulling new accurate information) or cross checking of data will occur on a frequent basis. The published date is a mandatory field in the metadata catalogue and will clearly identify this. Once the data has been anonymised for external agencies there will be no ability to update the anonymised data sets. Another full anonymised data set can be produced and this would be version controlled by date against when it was produced.

To ensure the quality of data contained within datasets the quality processes below will be used against the datasets

Firstly, the processes will performed against particular "Standards"

- ISO 11179 Metadata standard

- ISO 8000 Data Quality
- ISO 25012 Quality Assurance

Secondly each dataset will be checked for completeness and consistency; that the data contained is appropriate for that dataset and that the data is accurate and cleansed.
To help achieve the required data quality a 'Plan-Do-Review' process will be used
This will be coupled with the following controls:-

- Dataset Version Management
- Access control for curated datasets under version control
- Risk-management Controls
o e.g. Security controls
- Role-based access
o e.g. Manual quality-check 'gateways'
- e.g. 'Sensitive/Personal Information' removed
- Pseudonymisation correct and traceable
- Anonymisation correct and untraceable
- Categorized Reference Data (aka Master Data)
- Categorized Transaction Data
- ASCII character set or Unicode
- Mandatory fields populated
- Range constraint on data fields (e.g. Age 0 to 150)
- Remove leading and trailing non-visible characters
- Primitive Data-type constraint (e.g. integer, decimal, string, Date)
- Entity Data-type constraint (e.g. DoB, Country Code, Disease, Postcode, SNOMED)
- Uniform spelling
- Duplication alerts
- Missing data alerts
- Semantic compatibility / ontology-checked (e.g. NHS & DECOVID data-dictionary)
- Foreign-keys matched to Primary keys within included tables
- Auditability built-in / considered from the start

The aim of the database is to curate granular, real-time and ever-green data of the acute care impact of COVID-19 in the UK both during the pandemic and longitudinally. To facilitate this, a

shared ‘secret salt’ will enable this process to happen, by facilitating consistent pseudonymisation so the same patient would always have the same pseudo ID regardless of the Trust undertaking the pseudonymisation for future care episodes to be captured.

These processes will help ensure the data quality of the DECOVID data.

Only when these processes are completed will the data be pseudonymised.

5.2.1 Training

As this is an innovative project there will be ongoing development to support application of the data, including but not limited to:

- Data Protection and Information Governance – including institutional GDPR and Cyber security training and Data Security Awareness Programme provided by NHS Digital and Health Education England (see https://www.e-lfh.org.uk/programmes/data-security-awareness/)
- Data dictionary
- Support with analysis of data
- Development and testing of algorithms to improve patient care delivery

5.3 Data Security and Access

DECOVID is committed to promoting the protection of privacy and data security in line with the OECD Recommendation of the Council on Health Data Governance, and to use a proportionate approach to the governance of data access based on the five “safes''(8). DECOVID recognises the model’s key feature that the five dimensions ‘severally and jointly’ contribute to the safety (or risk) around data access.

5.3.1 Safe Projects: Is this use of the data appropriate?

‘Safe projects’ refers to the legal, moral and ethical considerations surrounding use of the data.

One of the essential criteria for all projects requesting access to data will be to demonstrate likelihood of patient benefit. Specifically, the project will be evaluated against:

- Does the research aim to bring patient benefit ('public good')?
- What is the predicted size of that benefit?
- What is the likelihood of the project being successful and this benefit being realised?
- What is the risk of unintended harms including potential discrimination?

It should be noted that there may also be a risk of 'loss to public benefit' through not doing the project. The scientific value of the research question and the likelihood of successful in answering the question were the data access permitted will be assessed by the Scientific Steering Committee. See section 6.0.

5.3.2 Safe People: Can the researchers be trusted to use it in an appropriate manner?

'Safe people' reviews the knowledge, skills and incentives of the users to store and use the data appropriately.

One of the essential criteria by which DECOVID will evaluate all applications will be whether the applicant is deemed to be appropriate. Specifically, the applicant will be evaluated against:

- Can the applicant be trusted to use the data exclusively for the purpose agreed and on the terms agreed?
- Does the applicant understand the reasons for the restrictions of use, including restrictions on onward data transfer, linkage or manipulation?
- Do they have the necessary skills to undertake the work described and deliver trustworthy outputs?
- Do they have the resources to complete the project?

Evidence for answering the above questions will be supported by the DECOVID Due Diligence Process (DDP), which is outlined in Appendix 2.

Part of ensuring 'Safe people' is that a condition of access for successfully approved projects is for the applicants to undertake relevant training provided by DECOVID and to engage constructively throughout the life of the project to ensure understanding and active acceptance of access conditions, which will support appropriate safe behaviour.

5.3.3 Safe Data: Is there a disclosure risk in the data itself?

This is discussed in section 2.11.5.

5.3.4 Safe Settings: Does the access facility limit unauthorised use?

DECOVID provides a safe setting through technical and physical security, education and culture, and contractual safeguards. Enhanced by high-powered computing services, secure access, analytics and data exchange support, leveraging proven delivery expertise through UHB and Microsoft.

Access rights to data are limited by dual factor authentication for cloud access. DECOVID password policy will following NCSC guidance (as laid out in https://www.ncsc.gov.uk/section/advice-guidance/all-topics) with specified role rights. Data will be stored on a central web-based platform that is secured. The platform will be located on a private shared cloud provided by Microsoft Azure. Central data will only be accessible as approved by the Data Controller(s) following a use-based access control for the purpose of audit, QA checks and reports.

The DECOVID system will be installed on the Microsoft Azure platform and will have the backup and recovery tools provided by Microsoft to protect data and installations.

A comprehensive audit trail is in place for the DECOVID system and the data sets it contains these tracks:

- who has accessed the system and when,
- when data items are created and who by
- when data items are edited and who by
- when data sets have been browsed or information (with correct permissions) has been accessed and downloaded

5.3.5 Technical Security

Enhanced by high-powered computing services, secure access, analytics and data exchange support, leveraging proven delivery expertise through UHB, UCLH, the Alan Turing Institute and Microsoft.

DECOVID structure is guided by FAIR Data Principles (findable, accessible, interoperable and reusable). Access and usage of the secure infrastructure will implement the DSP Toolkit and BS-ISO-27000 Series of Information Security Standards.

5.3.6 Physical Security

The database will sit on a secure UHB tenancy on a Microsoft Azure Cloud instance. This Cloud instance will be in either the UK South or UK West Microsoft data centres. However, should capacity be in question, DECOVID may use Microsoft data centres in Europe and the USA as Microsoft's Privacy Shield registration provides robust security including meeting requirements of GDPR.

The Azure Cloud data centre physical security features a layered security model, including safeguards like custom-designed electronic access cards, alarms, vehicle access barriers, perimeter fencing, metal detectors, biometrics, and the data centre floor features a laser beam intrusion detection system.

Microsoft data centres are monitored 24/7 by high-resolution interior and exterior cameras that can detect and track intruders.

Access logs, activity records, and camera footage are available in case an incident occurs. Microsoft data centres are routinely patrolled by experienced security guards who have undergone rigorous background checks and training. Access to the data centre floor is only possible via a security corridor which implements multi factor access control using both security badges and biometrics. Only approved employees with specific roles may enter.

Data is broken into subfile "chunks," which are stored on local disks and identified by unique chunk IDs. Microsoft encrypts data as it is written to disk with a per-chunk encryption key that is associated with a specific Access Control List (ACL). The ACL helps ensure that data in each chunk is only decrypted by authorised Microsoft employees and services that were given permission at the time of encrypting the data. This means that different chunks are encrypted with different encryption keys, even if they belong to the same applicant.

These chunks are encrypted using 128-bit or stronger Advanced Encryption Standard (AES).

5.3.7 Network Security Management

Within UHB and UCLH the network security will be controlled with the Trust network security protocols. Any data leaving UHB will be encrypted in transit and at rest. Data transfers from organisations contributing datasets will be done via sFTP between servers (secure File Transfer Protocol).

Data stored on Microsoft infrastructure is automatically encrypted at rest and distributed for availability and reliability (as above). This helps guard against unauthorized access and service interruptions.

Penetration tests for external-facing systems:

Data on internal UHB systems will be protected by Sophos. When the system sits on the secure UHB Informatics tenancy on Microsoft Azure Cloud instance where it will be protected by Azure's Security Centre. (Security Centre helps safeguard Windows servers and clients with Windows Defender Advanced Threat Protection and helps protect Linux servers with behavioural analytics. For every attack attempted or carried out, we would receive a detailed report and recommendations for remediation.

The DECOVID system will have been penetration tested by an external ethical hacking company if required but will conduct IT health checks to meet NHS DSP toolkit requirements. Microsoft themselves utilise Red Teaming, a form of live site penetration testing, against Microsoft managed infrastructure, services and applications

5.3.8. Access Control

Technical authorisation/access will include specific access points via two-factor authorisations combined with recorded Media Access Control (MAC) address, and secure, encrypted transport layers.

5.3.9 Contractual Safeguards

Access to data will include contractual obligations which:

- expressly preclude any attempts at re-identification
- limit the use of the data to the purposes described within the contract
- require clients to seek approval from the database before transfer to a third party and to "flow down" all requirements through sub-contracts.
- Require clients to provide evidence of data destruction.
- Provide UHB with the right to audit any activity by the client and its subcontractors.

5.3.10 Safe Outputs: Are the statistical results non-disclosive?

It is important that researchers publish their findings, and with sufficient detail to maximise the value of the study. However, the way that data is presented, particularly in tables, may provide sufficient detail for inadvertent disclosure at individual level. DECOVID requires authors to ensure that this is avoided through an 'output statistical disclosure control' in which they evaluate all statistical output for risk of disclosure. A common example is for tables where any cells may have less than five units. In such cases, we would ask authors to either: (1) consider collapsing categories if possible; or (2) replace the cell count with '<5'.

5.4 Database Software

The software will be compatible with all modern web browsers, Internet Explorer, Firefox and Safari. The software has high level security and encryption. It has multilevel security, data encryption for storing sensitive information, and dual factor authentication password protection for data entry and retrieval. Access to the data is controlled through a Roles Based Access control (RBAC).

5.5 Record Retention

The application for the DECOVID Research Database is initially for five years, but with the expectation of future applications for renewal to continue its benefits longer term. Anonymised datasets created on demand will be timestamped and made available under contractual arrangements for pre-specified time periods in line with the nature of the projects.

5.6 Down-stream Security/Integrity

Access to the data under the agreed approval will be on condition of a 'safe setting' for its analysis and use. DECOVID will require assurance of compliance with relevant standards (notably ISO 27001 and the DSP Toolkit) and the Data Controller for the database may request evidence of systems/policies/procedures to ensure such. This will be reflected in data sharing agreements.

6.0 Data Sharing

Pathways to enable appropriate data sharing have been developed with reference to the principles of the Open Research Concordat(9) and in partnership with patient and public partners. This concordat sets out ten principles with which all those engaged with research should be able to work. These principles are:

1. **Open access to research data is an enabler of high-quality research, a facilitator of innovation and safeguards good research practice**
2. **There are sound reasons why the openness of research data may need to be restricted but any restrictions must be justified and justifiable.**
3. **Open access to research data carries a significant cost, which should be respected by all parties.**
4. **The right of the creators of research data to reasonable first use is recognised.**
5. **Use of others' data should always conform to legal, ethical and regulatory frameworks including appropriate acknowledgement.**
6. **Good data management is fundamental to all stages of the research process and should be established at the outset.**
7. **Data curation is vital to make data useful for others and for long-term preservation of data**
8. **Data supporting publications should be accessible by the publication date and should be in a citeable form.**
9. **Support for the development of appropriate data skills is recognised as a responsibility for all stakeholders.**
10. **Regular reviews of progress towards open research data should be undertaken.**

Pathways to enable appropriate data sharing have been developed in partnership with patient and public partners.

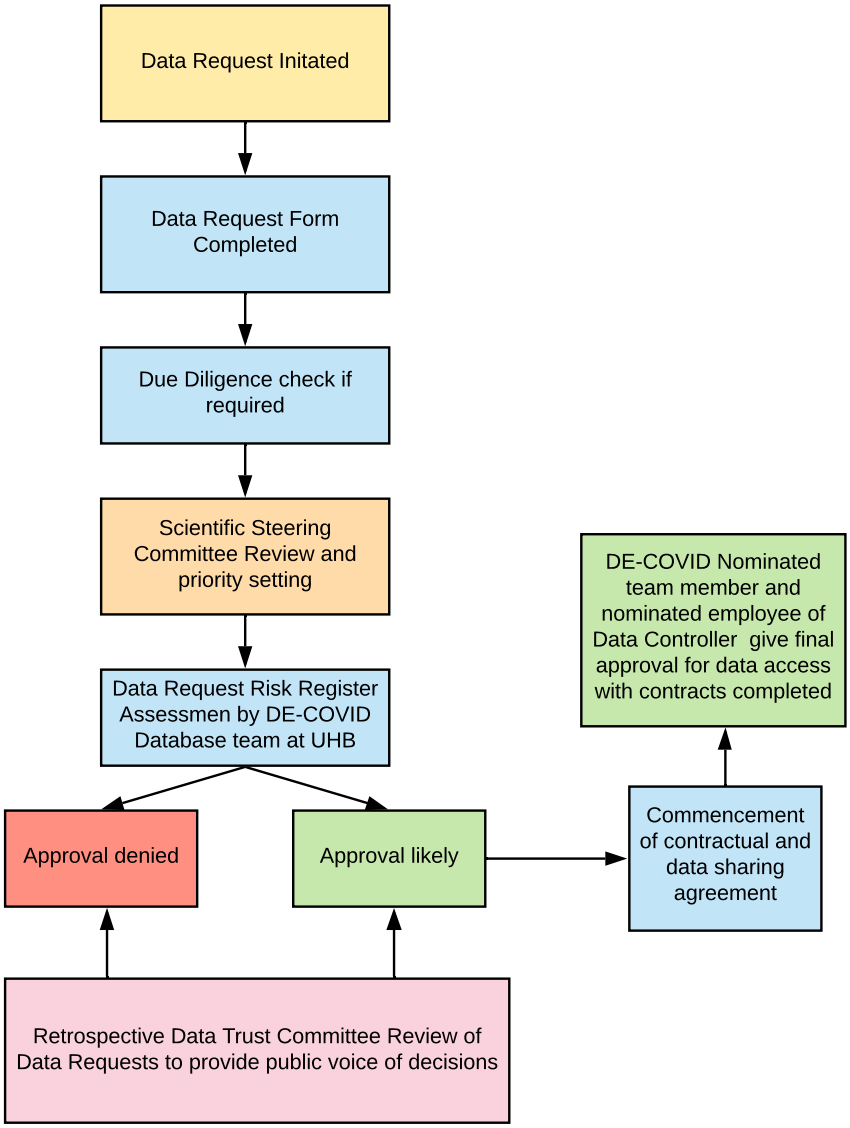
DECOVID is committed to the following principles:

1. Maintaining the highest standards of rigour and integrity in all aspects of research and data access;
2. Ensuring that research which includes DECOVID data is conducted according to appropriate ethical, legal and professional frameworks, obligations and standards;
3. Supporting a research environment that is underpinned by a culture of integrity and based on good governance, best practice and support for the development of researchers;
4. Working together to strengthen the integrity of research and to review process for data requests regularly and openly.

The figure which provides an overview of this process is shown in Figure 2.

See the sections below for details of these processes.

Figure 2: DECOVID Data Access Process



6.1 Access to Data Application Process

DECOVID will produce a metadata catalogue and data dictionary which will be freely available for researchers to browse to enable an understanding of the data held within DECOVID.

Data requests will be considered from NHS organisations and research institutes, members of the public or any non-commercial agency or body and for the purpose of the protocol, are referred to as Data Requestors. Requests from commercial entities will only be considered at a later phase of the research database and would be sought using a CAG application and protocol amendment. All requests for access to data will be considered as part of a three-stage review and release mechanism.

These are:

- Stage one – Technical and scientific assessment
- Stage two – Data Request Risk evaluation
- Stage three – Record and release.

These processes are described in detail below.

All requests for licensed access to data will be considered against core principles for data access and against the "5 safes" described in section 5.3:

1. Data requests that support a project which is likely to be of benefit to patients, to the NHS or with clear societal benefits
2. Data requests are from Organisations, Researchers or individuals which pass the "5 safes"
3. Data requests which are ethical, appropriate and include sufficient data to answer the proposed question but are not excessive in the data requested.

Requests for access to data may be initiated through the HDR-UK Health Data Research Innovation Gateway or through direct contact with DECOVID team members. The process this initiates is the same for either means of contact.

The Gateway is an application which supports researchers and innovators to discover and access data from the UK Health Data Research Alliance in a safe and responsible manner, and contains a metadata catalogue of all data available through HDR-UK.

All engagement will start with the Data Requestor completing a Data Request Form (DRF). The DRF also includes contact details and an initial description of what the request involves. For some projects, the DRF may be sufficient to permit data release. Where more information is needed, this

will be curated in a bespoke data request application, adding to the information within the DRF as needed.

Research questions can also be submitted from non-academic clinical teams or patients and here the DECOVID operations team will help complete the DRF for review by the Scientific Steering Committee as described below.

Box 1. DECOVID Data Request Form (indicative content)

Data Request Form

SECTION A: THE PROJECT

A1: **Project title.** (200 characters)

A2: **Research question(s) and aim(s)** (up to 200 words)

A3: Background and scientific rationale of the proposed research project (up to 300 words)

A4: A brief description of the method(s) to be used (up to 300 words)

A5: The type and size of dataset required (up to 100 words)

A6: Whether any pseudonymised data is requested? Yes/ No

A7: The expected value of the research (taking into account the public interest requirement) (up to 100 words).

A8: Up to 6 keywords which best summarise your proposed research project. (added here)

A9: Lay Summary.
A lay summary of your research project in plain English, stating the aims, scientific rationale, project duration and public health impact suitable for publication on the DECOVID website (up to 200 words).

A10: The estimated duration of the project, in months. (add here)

A11: How will results be shared/ disseminated? (up to 300 words)

SECTION B: THE DATA, SETTING AND ANALYSES

B1: Level of data access requirement
a) Do you wish to commission DECOVID to conduct the analysis for you minimising your direct exposure to the data? Yes/No

b) Can you undertake the planned project using aggregate data only? Yes/No

c) Do you wish to request access to anonymised individual patient-level data? Yes/No

B2: Selection of data-fields

a) Standard data-fields requested (listed within the DECOVID Metadata catalogue). Yes/No
(List data fields required).

b) Additional data-fields requested (subject to availability). Yes/No

c) Pseudonymised data requested? If so explain what fields and justify why the analysis could not be completed using anonymised data (up to 200 words)

d) Will the data be linked to any existing data set the data requestor holds? (please describe in up to 100 words).

B3: Data environment:

a) Will you access the data solely within the DECOVID Data Safe Haven? Yes/No

b) Will you require transfer of data to an alternative secure environment in order to achieve the project aims? Yes/No

c) If yes, then:
   i) What are the reasons that this transfer is required?
   ii) Are the standards of transfer ISO27000 series compliant?
   iii) Does the alternative Data Environment satisfy all requirements of:
      ISO27001 Yes/No
      NHS Data Security and Protection Toolkit Yes/No

B4: Statistical analysis

i) What forms of statistical analysis are planned? (up to 100 words)

ii) How is it intended that this will be presented in the final output? (up to 100 words)

iii) What is the smallest cell value that is likely to be generated by this analysis, and how will this be managed to avoid disclosure? (up to 100 word)

B5: Machine learning

a) Will the data will be subject to any machine learning (ML) techniques? Yes/No

If Yes, please specify:

b) Type of ML technique(s)

c) Is the DECOVID data for:
   i) Algorithm generation and training Yes/No
   ii) Internal validation Yes/No

iii) External validation Yes/No
iv) Other - please specify

B6: Ethical approvals

a. Do you seek for your project to be approved under the generic favourable ethical opinion of the DECOVID Research Database (REF 20/HRA/1689)?
Yes/No
b. Do you seek for your data access request to be considered under pre-existing ethical approval? (Please attach all relevant documents). Yes/No

SECTION C: THE APPLICANT AND RESEARCH TEAM

C1: Lead Applicant

i) Name
ii) Email address
iii) Current position
iv) Institution
v) Specific role(s) in the project

C2: Evidence of Lead Applicant's expertise and experience relevant to delivering the project including:

i) relevant publications (up to 5 most relevant)
ii) other relevant outputs

C3: Sponsoring organisation

i) Name
ii) Legal name (if different; to appear on any legal documents)
iii) Sector
iv) Size of institution

C4: Co-applicants

i) Name
ii) Current position
iii) their Institutions
iv) Specific role(s) in the project

C5: Other significant project team members

i) Name
ii) Current position
iii) their Institutions
iv) Specific role(s) in the project

C6: Contact person

i) Name
ii) Email address
iii) Preferred telephone contact number

C6: Contracts person

i) Name
ii) Email address

iii) Preferred telephone contact number	
Internal Use only: **Log number:** **Date request received:**	**Time:**

6.1.1 Stage One - Technical and Scientific assessment

Each DRF request is logged by the DECOVID project officer with a unique number, the date and time of the request, on a Data Request Database which captures how the data requests flows through the DECOVID system and the rationale and end results for end decision making.

The DRF will be assessed against the following criteria (as described in Box 2)

Box 2. Initial Screening of DRF by DECOVID Operations Team

Box 2: Initial Screening of DRF by DECOVID Operations Team

PART A:
SUFFICIENT INFORMATION SCREENING

A1) Is the form complete?
A2) Is the potential for patient benefit/public interest present and clearly stated?
A3) Is the data request clear (including number, types of data fields)?
A4) Is enough detail provided for reviewers to evaluate the extent to which the 'five safes' are met?
A5) Is pseudonymised data requested and justification given?

If the answer is NO to any of the above questions then register as an enquiry and return to applicant for further information.

If the answer is YES to all questions then register as a full application and proceed to part B.

PART B:
PURPOSE SCREENING
B1) Is there potential for patient benefit/public interest?

APPLICANT DUE DILIGENCE SCREENING
B2) Does the applicant require and has passed a Due Diligence Process (Appendix 2)?

TECHNICAL SCREENING
B3) Is the data for which access is requested currently or potentially available within the scope of the DECOVID Research Database?

If the answer is NO to any of the above, the application should be declined and the reasons given to the applicant.
If the answer is YES then proceed to Scientific Steering Committee review (see section 6.1.1.2)

6.1.1.1. Due Diligence

DECOVID will undertake due diligence checking for all Data Requestors, in recognition of the need for public trust in DECOVID operations. The due diligence check will be updated at each data request from that requestor.

The Due Diligence process consists of:

1) Checking any previous due diligence checks.
2) Completing the necessary sections of the due diligence paperwork.
3) Researching predefined online media sources by keyword search
4) Check due diligence outcomes of HDR-UK gateway or other data providers
5) Maintaining a log of all of the above.

DECOVID will follow the University Hospitals Birmingham NHS FT due diligence process. The due diligence form (DDF) notes the Requestors organisation sector, legal status, start date, as well as notable media exposure and previous working partnerships with any of the DECOVID partners. A failed due diligence will prevent data access. For the due diligence process, see appendix 2.

6.1.1.2. The Scientific Steering Committee

The scientific value of the research question and the likelihood that the question will be answered by the researchers will be evaluated by the Scientific Steering Committee (SSC). The SSC will be chaired and hosted by UCLH (Professor Williams) and include representation from the DECOVID founding members and the Research Directors of the founding members associated HDR UK sites/HDR UK national programmes (Prof. Hemmingway (London): Prof. Ball (Birmingham): Prof. Landry (Oxford). The aim of this process is to ensure data provision supports projects which meet the requirements of the 5 safes including the scientific rigour of the question and analytic approach. Particular scrutiny will be given to data requests which request pseudonymised data. The SSC may invite additional members to provide expertise identified by the SSC. Invited members will not have voting rights.

The SSC will review all data applications made from non-partner research agencies (termed "external research agencies") and prioritize them, based on the robustness and rigour of approach, giving each a "high", "medium", "low" or "unsuitable" priority rating. The rating for this will be based on the following criteria:

Box 3. SSC request rating system

Priority Rating	Definition
High	Meets all of the following: 1. Clear research question with demonstrable patient benefit 2. Urgent and important research priority against demonstrable clinical need 3. Sound research methodology and analytical plan 4. Secure information governance framework for data access 5. Strong research team with proven track record in project delivery and quality outputs or emerging reputation for excellence in field
Medium	Meets all of the following: 1. Clear research question with demonstrable patient benefit 2. Important research priority against demonstrable clinical need (perhaps not urgent) 3. Sound research methodology and analytical plan 4. Secure information governance framework for data access 5. Strong research team with proven track record in project delivery and quality outputs or emerging reputation for excellence in field
Low	Meets some of the following: 1. Clear research question with some evidence for patient benefit 2. A potential research priority with some evidence of clinical need 3. Sound research methodology and analytical plan 4. Secure information governance framework for data access 5. Previous outputs in the field
Unsuitable	Meets any of the following:

	1. No clear research question with little demonstrable patient benefit 2. Unclear or unsound research methodology and analytical plan 3. Concerns regarding information governance framework for data access 4. No previous outputs in the field or poor quality previous outputs.
Pseudonymised data request	The panel will be asked to form a consensus view (justified with notes) as to whether pseudonymised data requests are valid and needed, or whether anonymised data (which is the default for the database) will suffice.

An SSC review can also be requested by the Data Controller if concerns were raised against any of the 5 safes for data use by any requester. A project rated as having a high or medium priority will be supported with data access by DECOVID unless the Data Controller rates data release as high risk, as shown in Box 5. A project of low priority may be supported depending on risk rating and capacity. A project deemed unsuitable will not be supported for data release.

SSC members can request access to DECOVID data by completing a data request form, but conflicts of interest will be declared and the panel member will leave the panel for the discussion of their project (if discussed). If the panel member requesting data is the Chair, the deputy Chair will Chair the meeting. For the Terms of reference of the SSC see Box 4.

Box 4. Terms of Reference for SSC

The SSC will:

- Have a named Chair
- Have a named deputy Chair
- Have up to 11 panel members which include representatives of all DECOVID partners
- Include panel members which represent data provision sites (or a proportion of these in a rotational fashion once data providers increase in number)
- Founding NHS Trust members (UHB and UCLH) will be permanent members od the SSC.
- Include 2 lay members from the Data Trust Committee and the UCLH BRC Lay Panels and patient network
- Review Data Requests from external requesters at least monthly
- Include members with a background in academic clinical medicine and science to be able to evaluate projects

- Include members with a background in data science to be able to evaluate projects from a data perspective
- Include appropriate statistical support for project evaluation
- Include a data governance officer to provide IG support
- Non-partner panel members will service for a term of 3 years in the first instance, but may serve longer if requested
- Have a secretariat
- Vote on each data request and provide a consensus opinion as to the priority rating for each external data request
- Form reports on priority rating of reviewed data requests to be fed back to management committee for actioning.
- These reports will be shared with the Data Trust Committee
- Instruct on the type of agreement to be implemented with the data requestor(s)

6.1.2. Stage Two –Data Request Risk Evaluation

Each DRF and SCC review will be collated by the DECOVID Operations Team (shared working between UHB and UCLH) and given a Data Request Risk Rating (green for low risk, amber for moderate and red for high risk) based on the data requested, timelines, potential for reputational risk and potential for patient gain, as outlined in Box 5.

The following Risk Rating will be applied to all data requests and will include past dealings with DECOVID, DECOVID partners or HDR-UK. Potential breaches of contract from past data use will be reviewed on a case by case basis and assessed for seriousness by the DECOVID Management team, and Data Controller's information governance legal and contracts department, as required.

Box 5. Data Request Risk Register

Descriptor	Green/ Low	Amber/ Moderate	Red/ High
Due Diligence (if needed)	Passed	Passed	Did not pass
Previous dealings?	Select one of: **Yes.** Met all contractual obligations for data use, attribution, data security and outputs and acted in	Select one of: **Yes;** previous dealings. Met contractual arrangements but minor deviations from DECOVID	Select on of: **Yes**: One or more serious breach of contract or repeated breaches of contractual obligations

	accordance with DECOVID guiding principles Or: **No** previous dealings but considered low risk of contractual breach (add detail as needed)	guiding principles (for example, open access) Or: **No serious breach of contrac**t and no repeated breaches of contractual obligations Or: **No previous dealings** but considered minor risk of contractual breach (add detail as needed)	Or: **No previous dealings** and considered high risk of contractual breach Or: **Previous serious contractual breach** with other HDR-UK data provider (add detail as needed)
SCC Rating	High or medium	Low	Unsuitable
Data security	Provides evidence of data security measures which meet all requirements (add comment)	Provides evidence of data security measures which meet most requirements with additional support (add comment)	No evidence of data security or evidence to suggest risk of data breach (add comment)
Potential for reputational risk to DECOVID	Low (add rationale)	Moderate (add rationale)	High (add rationale)
Decision by DECOVID Management Committee	Suggestion to support data release (as requested)	Suggestion to support data release (as requested or with modification)	Suggestion not to support data release

Where the suggestion is to support data release, contractual arrangements including data sharing agreements can be initiated.

The final decision for Data Release will be made by the Management Team of DECOVID (see section 6.1.4) and the Data Controller for DECOVID; but all decisions can be reviewed by public members through the Data Trust Committee to ensure DECOVID is operating with the support and expected remit of the public and patients.

6.1.3 Stage 2. Public and patient involvement in DECOVID data release. The Data Trust Committee

6.1.3.1 Data Trust Committee Review

Across the project a Data Trust C (DTC) will be established and terms of reference agreed to, as described below. In essence, the Data Trust Committee will act as the public conscience of

DECOVID and consider all Data requests. The Data Trust Committee is an advisory function for DECOVID and cannot approve data release but instead provides a clear public voice within DECOVID.

The DTC will be made up of at least 7 individuals who will have a term on the DTC of up to four years with the potential to reapply for a further term and will include at least 2 members of the BRC's research patient network. There will be an open application by letter to become members of the Data Trust Committee following open advertisement and these members will also act as the Public Advisory Board for DECOVID. All members of DTC must declare all relevant conflicts of interest, including any relationship to Data Requestors. DTC will be assisted by experts in data research, information governance and UK data law, but these experts will have an advisory capacity only and will not be voting members of the DTC. There will be a nominated professional secretariat. There will be a DTC Chair, selected by the DTC.
DTC members must sign up to the terms of reference of membership. These are given in Box 6.

Box 6. Terms of reference for Data Trust Committee

Terms of reference include:

- Listing of members of DTB if requested
- Meeting at least every 2 – 3 months to discuss data requests, data access and operations of DECOVID
- Being able to review and discuss all Data Requests and data provision decision by DECOVID
- Feedback thoughts about data provision decisions and voice any areas of concern or uncertainly
- All decisions are to be made in accordance with the protocol and principles of DECOVID as laid out in the protocol.
- The DTB will form a six-monthly report to the DECOVID Management Committee (DMC) and contribute to the annual REC review
- Recommend where involvement or consultation with a wider public is needed
- The SCC and DMC will feedback to the DTC on all recommendations for discussion and review
- The DTC will generate lay summaries of their activity for public review
- The DECOVID Operations team and professional PPIE lead will assist in writing all reports for the DTC.

- An exceptional DTC meeting can be called to consider at risk or potentially contentious data applications (as determined by the DECOVID Management Committee).

The Data Trust Committee will review a summary of all data request processes, decisions and outcomes; will review some applications in full at their or the DECOVID Management's Team's request and may request full disclosure of all recorded information pertaining to any data request made through DECOVID. The DTC's role is to provide a public view of data release/ access decisions to inform DECOVID's processes moving forward and not to make data release or access decisions.

DECOVID's operating procedures may be amended over time to reflect the learning gained from the working closely with the Data Trust Committee. The Data Trust Committee's reports will be publicly accessible upon request and a lay summary of the report will be placed on the HDR-UK website.

6.1.4. Stage 3. Record and Release

Data release will require a Data Sharing Agreement (DSA), associated and agreed costing model (if needed) and scheduling for follow up events (such as publication of data requests and actions, request for data destruction, audit).

6.1.5 Specific Ethics Committee Approval of Research Projects

Where Sponsors approach the Research Database with pre-existing ethical approvals the Sponsor will provide any/all necessary documents to enable Technical and due diligence assessment. If the proposed project covers the data requested then the Data Controller and Information Asset Owner (or approved delegate) will consider releasing data in accordance with Stage 3 procedures.

6.1.6 Conditions of Data Release to Other Researchers

6.1.6.1 Open Access

Open access means that anyone with an internet connection can access the output of research, be it a journal article, algorithm or methodology, without the need to pay for access via a subscription or other mechanism.

DECOVID operates with the following guiding beliefs about open access.

- Transparency is a DECOVID core value.

- DECOVID acts for the public good.
- By being open we can share more and learn quicker from each other's successes and failures. Open access makes research more transparent, rigorous and efficient; stimulates innovation; and promotes public engagement.
- The public voice is at the heart of all we do – non-researchers must be able to access the outputs of DECOVID research.

DECOVID will operate within the following open access principles

Noting the above, DECOVID:

- Expects authors to maximise the opportunities to make their results available for free and to encourage data outputs to be publicly accessible with lay summaries freely available.
- Expects outputs of work supported by DECOVID to select publishing routes that ensure the work is available immediately on publication in its final published form, where possible.
- Encourages authors and publishers to licence research papers using the Creative Commons Attribution licence (CC-BY) so they may be freely copied and re-used (for example, for text- and data-mining purposes or creating a translation), provided that such uses are fully attributed.
- Encourages outputs published in a peer-reviewed journal, and supported in whole or in part by DECOVID, to be made available through PubMed Central (PMC) and Europe PMC as soon as possible and in any event within six months of the journal publisher's official date of final publication.

6.1.6.2 "Public Good" Condition for data release

All requests for data must have demonstrable potential for public benefit. This includes but is not limited to;

- The development of new health care processes, pathways, biomarkers, devices, therapeutics and software as medical devices.
- The development of new NHS services or new models of health care and development of new or augmented social care.
- Benefit to the NHS, through products, services, regulatory reports, audits or direct and indirect financial benefits
- Benefit to the public through the creation of new knowledge, products or services.

6.1.6.3 Attribution Policy

DECOVID research outputs include typical academic measures of success, such as publications. As publications are increasingly announced on social media platforms, such as Twitter and LinkedIn, attribution of tweets is also set out in this protocol. Core to HDR-UK's mission and DECOVID's policy is the generation of algorithms, code, software and methodologies that facilitate the analysis of large-scale data, so these are also covered by this protocol.

For publications, DECOVID must be included in acknowledgements or funding section. Suggested text includes:

- *This work was supported by DECOVID, which is funded by Health Data Research UK (or any funders to be recognised as part of DECOVID)*

For code and related digital artefacts.

- DECOVID would encourage code (e.g. algorithms, analytical script, source code) and related digital artefacts (e.g. documents) to be made available within the HDR UK GitHub repositories.
- Otherwise, similarly liberal and open source licenses (such as Apache 2.0, BSD, MIT and Eclipse Public License) should be used, permitting anyone to benefit from, improve upon and redistribute the code.

6.1.6.4 Downstream security

Data from the DECOVID research database will be released on condition that data will be held securely, to the standards described in section 5.0 of this protocol and its integrity will be maintained. The Data controller and Information Asset Owner may request evidence of systems/policies/procedures to ensure such and this will be reflected in data sharing agreements.

7.0 Management and Governance

7.1 DECOVID Management Committee (DMC)

DECOVID will convene a Management Committee (DMC). The DMC will meet at least monthly (which can be virtual meetings) and be made up of Database Projects Leads, a representative of the Data Providers, Data Trust Board, Programme Manager, and a secretariat. There will be a representative from each institution on the DMC (Chris Holmes or delegated member (Turing/HDR

UK); Dr Elizabeth Sapey (UHB) or delegated member; Dr Wai Keong Wong or Christina Pagel (UCLH) or delegated member, with equal voting rights.

Box 7. Terms of reference for DECOVID Management Committee

The DMC will:

- Keep an overview of the data within the database control
- Provide an overview and be able to demonstrate the monitoring of data security
- Develop an initial standard operating procedure to ensure compliance with the protocol
- Review potential new data providers
- Review data requests and data requests outcomes and timelines
- Review outcomes from DTC
- Review contractual procedures and timelines
- Monitor and review cost recovery strategy and overall finances
- Monitor and review reports and ensure timely submission
- Commission, approve and submit reports to the Data Controller and their delegated operator.
- Commission, approve and submit reports to funders
- Maintain a risk register
- Review reported data breaches, identify corrective and preventative action and ensure that it is completed to the satisfaction of the Information asset owner.
- Ensure template agreements are available commensurate with requestor risk management

7.2. Research Governance

DECOVID will ensure that researchers are responsible for ensuring that research will be conducted according to this protocol and related written instructions, and current applicable legislation. Agreements with the Trust at each participating centre will be in place covering data collection.

7.3 Reporting Breach of DECOVID Policy

Serious protocol non-compliance will be reported without delay to the Data Trust Board and Data Provider Partners. The DECOVID MC will designate an individual who will ensure that the issue is investigated and appropriate actions taken. The responsible REC will be notified as soon as possible of any serious breach of the REC approval conditions or any serious breach of security or confidentiality, or any other incident that could undermine public confidence in the ethical

management of the data. Information Governance data breach policy will be followed in accordance with UK law.

7.4. Progress Reports and Accountability

DECOVID and collaborating researchers share responsibility for providing accurate periodic progress reports as required by the main REC, host NHS Trust and other authorised agencies (such as funding bodies).

DECOVID will maintain a record of all research projects for which data has been released. The record should contain at least the full title of the project, a brief lay summary of its purpose, the name of the lead researcher. The main REC and host NHS Trust may request access to this record at any time.

An annual report will be provided to the main REC and NHS Trusts at minimum: details of data collection activity and details of all approved projects for which data has been released in the previous year and any related publications. For the purpose of annual reports, DECOVID will standardise on a single anniversary date i.e. the date of responsible REC approval.

8 On-going PPI/E strategy for DECOVID

Public and patient engagement and involvement are central to all DECOVID operations. The PPI/E strategy has been developed with the DECOVID PPI/E group and continuing outputs from the PPI/E group will be co-created, publicly available on the HDR-UK website, through peer review publications and reports.

8.1 PPI/E Overarching Aims

1. Patients and the public are partners in DECOVID
2. The needs, values and interests of patients and the public are understood and embedded in DECOVID decision making
3. People have trust and confidence in the use of health data within DECOVID for research and innovation
4. People have tangible gains from their data being used in research and innovation as part of DECOVID

Patients and public members will be invited to form a Data Trust Board for DECOVID, who will review applications for data access and feedback to the DECOVID Management Committee. Members of the DTB will also sit on the SSC and DMC, as equal voting members so that patient and public voices are included at all levels within the DECOVID structure.

9.0 Protocol Amendments

Any change in the DECOVID protocol which substantively change data processing within DECOVID will require an amendment which will require ethical review and approvals. The addition of extra data sites to DECOVID will not require an amendment. The information asset owner or will sign any amended versions of the protocol.

10 Annual Reports and Dissemination of Findings

DECOVID and collaborating researchers share responsibility for providing accurate periodic progress reports as required by the main REC, host NHS Trust and other authorised agencies (such as funding bodies). DECOVID will maintain a record of all research projects which data release has supported. The record should contain at least the full title of the project, a brief lay summary of its purpose. The main REC may request access to this record at any time. Any publications arising directly from the DECOVID database will be reviewed, approved and written with the acknowledgment of DECOVID and data partners with authorship following recognised international guidelines as described in the International Committee of Medical Journal Editors(10). Publications resulting from access to data will be requested to acknowledge DECOVID as the source of such data and where appropriate and by mutual agreement, to involve members of the DECOVID consortium as contributors to the design, analysis, or other inputs to the resulting work.

11 References

1. Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, Zhao X, Huang B, Shi W, Lu R, Niu P, Zhan F, Ma X, Wang D, Xu W, Wu G, Gao GF, Tan W. A Novel Coronavirus from Patients with Pneumonia in China, 2019. *New England Journal of Medicine* 2020; 382: 727-733.
2. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, Zhang L, Fan G, Xu J, Gu X, Cheng Z, Yu T, Xia J, Wei Y, Wu W, Xie X, Yin W, Li H, Liu M, Xiao Y, Gao H, Guo L, Xie J, Wang G, Jiang R, Gao Z, Jin Q, Wang J,

Cao B. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *The Lancet* 2020; 395: 497-506.
3. Wang D, Hu B, Hu C, Zhu F, Liu X, Zhang J, Wang B, Xiang H, Cheng Z, Xiong Y, Zhao Y, Li Y, Wang X, Peng Z. Clinical Characteristics of 138 Hospitalized Patients With 2019 Novel Coronavirus–Infected Pneumonia in Wuhan, China. *JAMA* 2020; 323: 1061-1069.
4. Yang J, Zheng Y, Gou X, Pu K, Chen Z, Guo Q, Ji R, Wang H, Wang Y, Zhou Y. Prevalence of comorbidities in the novel Wuhan coronavirus (COVID-19) infection: a systematic review and meta-analysis. *International Journal of Infectious Diseases* 2020.
5. Committee; WE. Statement on the second meeting of the International Health Regulations (2005) Emergency Committee regarding the outbreak of novel coronavirus (COVID-19). . *https://wwwwhoint/news-room/detail/30-01-2020-statement-on-the-second-meeting-of-the-international-health-regulations-(2005)-emergency-committee-regarding-the-outbreak-of-novel-coronavirus-(COVID-19)* 2020; DOA 27.3.2020.
6. J.; O, C.; H. Global COVID-19 Case Fatality Rates. 2. *https://wwwcebmnet/global-covid-19-case-fatality-rates/* 2020; DOA 27.3.2020.
7. Denaxas S, Hemingway H, Shallcross L, Noursadeghi M, Williams B, Pillay D, Pasea L, González-Izquierdo A, pagel c, Harris S, Torralbo A, Langenberg C, Wong W, Banerjee A. Estimating excess 1- year mortality from COVID-19 according to underlying conditions and age in England: a rapid analysis using NHS health records in 3.8 million adults. 2020.
8. Desai T, Ritchie F, Welpton R. Five Safes: designing data access for research. *https://www2uweacuk/faculties/BBS/Documents/1601pdf* 2016; DOA 2.3.2020.
9. UKRI.org. Concordat on Open Research Data. *https://wwwukriorg/files/legacy/documents/concordatonopenresearchdata-pdf/* 2016; DOA 7th Match 2020.
10. ICMJE. Defining the role of authors and contributors. *http://wwwicmjeorg/recommendations/browse/roles-and-responsibilities/defining-the-role-of-authors-and-contributorshtml* 2018; DOA 20.1.2020.

Appendix 1: Department of Health and Social Care COPI COVID-19 Directive

Department of Health & Social Care

39 Victoria Street
London
SW1H 0EU

To:
Organisations providing health services
General Practices
Local Authorities
Arm's Length Bodies of the Department of Health and Social Care

20 March 2020

Dear All,

Covid-19 – Notice under Regulation 3(4) of the Health Service Control of Patient Information Regulations 2002

The health and social care system is taking action to manage and mitigate the spread and impact of the current outbreak of Covid-19. Action to be taken will require the processing and sharing of confidential patient information amongst health organisations and other bodies engaged in disease surveillance for the purposes of research, protecting public health, providing healthcare services to the public and monitoring and managing the Covid-19 outbreak and incidents of exposure.

I am therefore writing to inform you that I am serving notice under Regulation 3(4) of the Health Service (Control of Patient Information) Regulations 2002 (**COPI**) to require organisations to process confidential patient information in the manner set out below for purposes set out in Regulation 3(1) of COPI.

This Notice **does not** apply to NHS Digital or NHS England & Improvement, which are subject to separate notices under COPI.

1. Purpose of this Notice

The purpose of this Notice is to require organisations to process confidential patient information for the purposes set out in Regulation 3(1) of COPI to support the Secretary of State's response to Covid-19 (Covid-19 Purpose). "Processing" for these purposes is defined in Regulation 3(2) and includes dissemination of confidential patient information to persons and organisations permitted to process confidential patient information under Regulation 3(3) of COPI.

I consider this Notice is necessary to require organisations to lawfully and efficiently process confidential patient information as set out in Regulation 3(2) of COPI for purposes defined in regulation 3(1).

2. Requirement to Process Confidential Patient Information

2.1. I hereby provide recipients with notice under Regulation 3(4) that I require you to process confidential patient information, including disseminating to a person or organisation permitted to process confidential patient information, including disseminating to a person or organisation permitted to process confidential patient information under Regulation 3(3) of COPI.

2.2. Organisations are only required to process such confidential patient information:

2.2.1. where the confidential patient information to be processed is required for a Covid-19 Purpose and will be processed solely for that Covid-19 Purpose in accordance with Regulation 7 of COPI; and

2.2.2. from the date of this Notice until 30th September 2020.

3. Covid-19 Purpose.

3.1 A Covid-19 Purpose includes but is not limited to the following:

- understanding Covid-19 and risks to public health, trends in Covid-19 and such risks, and controlling and preventing the spread of Covid-19 and such risks;
- identifying and understanding information about patients or potential patients with or at risk of Covid-19, information about incidents of patient exposure to Covid-19 and the management of patients with or at risk of Covid-19 including: locating, contacting, screening, flagging and monitoring such patients and collecting information about and providing services in relation to testing, diagnosis, self-isolation, fitness to work, treatment, medical and social interventions and recovery from Covid-19;
- understanding information about patient access to health services and adult social care services and the need for wider care of patients and vulnerable groups as a direct or indirect result of Covid-19 and the availability and capacity of those services or that care;
- monitoring and managing the response to Covid-19 by health and social care bodies and the Government including providing information to the public about Covid-19 and its effectiveness and information about capacity, medicines, equipment, supplies, services and the workforce within the health services and adult social care services;

- delivering services to patients, clinicians, the health services and adult social care services workforce and the public about and in connection with Covid-19, including the provision of information, fit notes and the provision of health care and adult social care services; and
- research and planning in relation to Covid-19.

4. Recording of processing

A record should be kept of all data processed under this Notice.

5. Review and Expiry of this Notice

This Notice will be reviewed on or before 30 September 2020 and may be extended by me by further notice in writing for the period specified in that notice. If no further notice is sent to you by me, this Notice will expire on 30 September 2020.

I am grateful for your continued support at this critical time for the nation.

Yours sincerely

On behalf of
Secretary of State for Health and Social Care

Appendix 2. Due Diligence Form and Process

Due Diligence Log Number (xxxx)		
Date of Due Diligence Review:		
PROPOSED FUNDER/PARTNER INFORMATION – Initial assessment		
1a	Principal address of organisation:	
1b	Number of years (months if less than 1 year) the entity has been in existence?	
1c	Any relevant parent/subsidiary companies and other affiliations?	
2	Is the entity involved in any aspect of the tobacco industry (including investment in/by the business)?	Yes/No If YES please elucidate:
3a	Is the entity involved in any aspect of arms manufacturing or trade?	Yes/No If YES please elucidate:
3b	Are you aware of any links between the entity and state governments, companies or individuals with current or past history of serious human rights violations?	Yes/No If YES please elucidate:
4	Are you aware of any reputational or relational difficulties for DECOVID in entering in to the proposed relationship? i.e. damaging media interest?	Yes/No If YES please elucidate:
5	Within the last five years, is there any published evidence that the entity, any predecessor of the entity, or any member of the entity has been associated with any of the keywords listed in the 'check for controversies'?	Yes/No If YES please elucidate:
6	**Any previous relationships with HDR-UK or DECOVID partners?** If Yes –Were concerns highlighted?	**Yes/ No** **If YES please elucidate**
APPROVED DUE DILIGENCE CODE		

SIGNATURE - DECOVID Staff performing due diligence**:**	
Print Name:	
Date:	

Please note: The content of this form, and any attached due diligence documentation is subject to the Freedom of Information Act and the Data Protection Act. Please do not include any content that is unsuitable for dissemination

Special Form 1: Laws and agreements to consider in regards to arms manufacturing and trade

Type of Arms	3ai. Does the entity manufacture or trade in this type of arms?	Relevant Treaty	3aii. If yes to manufacture or trade, does the entity comply with this treaty?
Explosive projectiles weighing less than 400 grams		Declaration of Saint Petersburg (1868)	
Bullets that expand or flatten in the human body		Hague Declaration (1899)	
Poison and poisoned weapons		Hague Regulations (1907)	
Chemical weapons		Geneva Protocol (1925)	
		Convention on the prohibition of chemical weapons (1993)	
Biological weapons		Geneva Protocol (1925)	
		Convention on the prohibition of biological weapons (1972)	
Weapons that injure by fragments which, in the human body, escape detection by X-rays		Protocol I (1980) to the Convention on Certain Conventional Weapons	
Incendiary weapons		Protocol III (1980) to the Convention on Certain Conventional Weapons	
Blinding laser weapons		Protocol IV (1995) to the Convention on Certain Conventional Weapons	
Mines, booby traps and "other devices"		Protocol II, as amended (1996), to the Convention on Certain Conventional Weapons	
Anti-personnel mines		Convention on the Prohibition of Anti-Personnel Mines (Ottawa Treaty) (1997)	
Explosive Remnants of War		Protocol V (2003) to the Convention on Certain Conventional Weapons	

Cluster Munitions		Convention on Cluster Munitions (2008)	

1. Researching online media sources to identify controversies

The most significant aspect of the due diligence process will be to undertake research online to screen for controversies. .

- ***How do I carry out an online search to check for controversies?***

The below keywords/phrases should be used to carry out Google searches on prospective corporate funders against a pre-defined list of sources. Where these identify potential controversies, ad-hoc searches may also be used to research these further.

Keyword/phrase
Ethical
Abuse
Bribery
Controversy
Corporate Manslaughter
Corruption
Discrimination
Extremism
Financial Irregularity
Fraud

Keyword/phrase
Human Rights
Illegal
Litigation
Slavery
Tobacco
Arms Trade
Defence
Trade Embargoes
UN sanctions
Health and Safety Breach

Proscribed list of credible sources (October 2018):

- www.reuters.com
- www.bbc.co.uk
- www.wsj.com
- www.economist.com
- www.nytimes.com
- www.theguardian.com

For example, for a funder called *Paradigm Shifting Research Funding Ltd*, the Google search terms used would be:

- "Paradigm Shifting Research Funding Ltd" bribery
- "Paradigm Shifting Research Funding Ltd" controversy
- "Paradigm Shifting Research Funding Ltd" "corporate manslaughter"
- …etc.

***Tip**: Use the 'Revised DD Google tool' spreadsheet saved HERE to generate the full list of search terms. Once generated copy & paste the list into the 'Multiple Tabs Search' Chrome browser extension. (Add this extension to your browser using the following LINK)*

- ***What counts as a controversy?***

The keywords act as a helpful guide as to what constitutes a controversial issue. Any media coverage relating to any of these issues has the potential to negatively impact on DECOVIDif funding is accepted, and as such should be recorded as part of the due diligence review.

- ***Why is tobacco included as a keyword?***

The University's Code of Ethics states "The University's investment policy excludes direct investment by/in the tobacco industry." This is due to the terms of our agreement with Cancer Research UK (CRUK). Any investment in/from a company involved in the production of tobacco or tobacco-related products (i.e. cigarettes, etc.) could jeopardise our contract with CRUK. Thus it is imperative that any link between a company offering funding and tobacco is included in the due diligence paperwork, however small. This can be as apparent as a cigarette manufacturer, or as subtle as a company supplying machinery to that manufacturer.

- ***How should findings be recorded?***

The role of the researcher is to present an objective, rounded summary of any news stories which point to controversies. This may require including some contextual background to findings so that, when it comes to sign-off, findings are conveyed fairly and accurately.

As an example, imagine a large pharmaceutical organisation is offering funding to DECOVIDfor data access. In carrying out research, you discover that the company is in ongoing litigation. This should be included in the findings, but the specific nature of the lawsuit will also have a bearing on the decision to accept or reject funding. If the lawsuit is in relation to claims over the side effects of Powercetemol in earlier trials, this will have a material bearing on the due diligence. On the other hand, if the lawsuit is in relation to a different drug altogether, or a shareholder dispute, or anything unrelated to the proposed DECOVID research project, the impact is less acute. Context, therefore, is clearly very important when recording findings, and will prove helpful when it comes to making a final risk assessment at the point of sign-off.

- ***What timeframe should be considered when researching findings?***

Generally speaking, any reports in the past five years should be considered when researching controversies. However, if a matter of particular concern is identified outside of this timeframe, it should be included.

- ***Do I need to include references?***

Yes, footnotes linking to the news articles discovered should be included. Remember, even reliable sources need to be treated with care, so it is best practice to 'dual source' wherever possible. This is the act of locating a second article from a separate reputable publication which covers the same issue, and adds to the rigour of findings by presenting multiple touch points.