British Heart Foundation Data Science Centre
Led by Health Data Research UK

HDRUK
Health Data Research UK

CVD-COVID-UK / COVID-IMPACT Ways of Working

Table of Contents

1 Introduction 3
1.1 How to contact us 4
2 Becoming a consortium member 4
3 Communication 4
3.1 Meetings 4
3.2 Collaboration 5
4 Patient and Public Involvement and Engagement (PPIE) 5
5 Project proposals, approval process and amendments 6
5.1 Proposing a project 6
5.2 Funding the cost of data access and conflicts of interest 6
5.3 Submitting a project proposal and initial triage 7
5.4 Approvals & Oversight Board 8
5.5 Reviewing / approving projects 8
5.6 Contributing to projects 9
5.7 Amendments and proposal versioning 9
6 Project lifecycle 10
6.1 Stage 1: Start-up 10
6.1.1 Establishing your project's presence 10
6.1.2 Forming your project team 11
6.1.3 Data access and safe researcher/information governance training 12
6.1.4 Confirming the number of planned outputs (publications)/subprojects 14
6.1.5 Developing and version controlling protocols / SAPs 14
6.1.6 Creating your phenotypes and analysis code 16
6.2 Stage 2: Analysis 16
6.2.1 TRE responsibilities 17
6.2.2 TRE best practice 17
6.2.3 TRE technical and data wrangler support 17
6.2.4 Requesting new R/Python packages 18
6.2.5 TRE outputs / statistical disclosure control 18
6.3 Stage 3: Manuscript/report writing 19
6.3.1 Authorship 19
6.3.2 Patient/public input and plain English summary of results 20
6.3.3 Manuscript/reports: consortium review 20
6.3.4 Manuscripts: journal peer-review 20
6.3.5 Communications 21

6.4 Stage 4: Sharing 21
6.4.1 Uploading your protocol, code and phenotype code lists to GitHub 22
6.4.2 Submitting phenotype definitions to the HDR UK Phenotype Library 22
6.4.3 Website and Gateway 22
6.4.4 Technical offboarding meeting 23
6.5 Stage 5: Completion 23
7 Project planning and status reporting 23
7.1 Project planning 23
7.2 Status reporting 23
8 Acknowledgements 24
9 Abstracts, presentations and posters 24
Appendix A: CVD-COVID-UK / COVID-IMPACT Consortium: Principles for Participation 25
Appendix B: Responsibilities framework for COVID research in the NHS England Secure Data Environment Service for England 26

1 Introduction

CVD-COVID-UK

CVD-COVID-UK, co-ordinated by the British Heart Foundation (BHF) Data Science Centre (DSC), is one of the NIHR-BHF Cardiovascular Partnership's National Flagship Projects.

CVD-COVID-UK aims to understand the relationship between COVID-19 and cardiovascular diseases such as heart attack, heart failure, stroke, and blood clots in the lungs through analyses of anonymised, linked, nationally collated healthcare data sources in secure trusted research environments (TREs) across the nations of the UK (note that TREs will also be referred to as a Secure Data Environment (SDE) or Safe Haven in this document).

As at the date of this document, we have access to data for:

- **England,** in NHS England's (NHSE) SDE until January 2026. We have submitted a request to extend data access to January 2027 to allow existing projects to complete.
- **Wales,** in the SAIL Databank until February 2026, by which time existing projects are required to have submitted their manuscript(s) to a journal/complete, or transfer to access data via the Big Data for Complex Disease research programme.

As a result of well-publicised changes at NHS England and the need to think beyond COVID-19 related research, NHSE will not be accepting new project proposals, or amendment requests, under the research programme's data sharing agreement. However, discussions are already underway to put future arrangements for data access in place.

Our ethical approval to access data currently runs to January 2027.

Linkable datasets include those from primary and secondary care, COVID-19 lab tests and vaccinations, deaths, critical care, prescribing/dispensing and cardiovascular, stroke and vascular surgical audits, maternity services and mental health. An up-to-date list of datasets provisioned in each TRE can be found on the programme webpage and more detail on each dataset in the Dataset Summary Dashboard.

We have established a consortium of over 400 members across more than 50 academic institutions and NHS trusts including data scientists, clinicians, data custodians and public contributors, all of whom have signed up to an agreed set of principles which embrace our ethos of being inclusive, transparent, and always working in a collaborative way.

Expansion to COVID-IMPACT

Building on the success of CVD-COVID-UK, the BHF Data Science Centre gained ethical and regulatory approval to broaden the scope of the programme to all COVID-related research (using data in NHS England's SDE service for England only). This is known as COVID-IMPACT and it helps to support research projects from the wider community.

Research questions

CVD-COVID-UK / COVID-IMPACT aims to answer the following broad research questions:

1. What are the effects of prior health conditions, their risk factors and medications on susceptibility to and outcomes from COVID-19 disease?
2. What is the direct impact of COVID-19 disease (and interventions used for its prevention and treatment) on acute complications as well as on medium- and longer-term health conditions?
3. What is the indirect impact of the COVID-19 pandemic and the government and NHS response to it on the presentation, diagnosis, management and outcomes of health conditions?

Further detail can be found on the programme webpage, which provides links to the main protocol, the consortium membership (updated periodically) and the consortium's principles.

1.1 How to contact us

Please email us at bhfdsc@hdruk.ac.uk with any consortium related questions, comments, suggestions or requests and we will respond as soon as possible.

2 Becoming a consortium member

Contact us if you would like to become a member of the consortium and contribute to one of our active projects.

You will be asked to sign up to an agreed set of principles (see Appendix A), which will apply at all times during your membership of the consortium, and to opt into receiving email communication from the centre. This will enable the BHF Data Science Centre (BHF DSC) team to invite you to consortium meetings, provide you with access to programme folders in Box, add you to our dedicated channel on the Health Data Research UK (HDR UK) Slack, and invite you to access our private repositories in our GitHub organisation.

We will hold your personal data on our systems for the purpose of carrying out consortium-related activities only. Your data will not be passed to any third parties, except for those members requiring data access, where limited detail will be passed to the national TRE providers for the purpose of setting up user accounts. You can unsubscribe from receiving communications/cease your consortium membership at any time, by emailing the centre.

3 Communication

3.1 Meetings

The main consortium meetings are currently as follows (note that these are subject to change and updates will be communicated to all consortium members, where necessary):

- **Consortium Update:** this takes place monthly and comprises:
    - A **consortium-wide programme overview** chaired by Angela Wood, Professor in Biostatistics and Health Data Science, University of Cambridge, and Associate Director of the BHF DSC's Whole Population Data area of work. The overview includes updates on projects, publications, and dataset provisioning in the

respective TREs. As a consortium member, you will be invited to this 30-minute update, but we appreciate that you might not always be available to attend.

- The second half-hour is devoted to the **Methods Group**, and is chaired by one of our senior health data scientists with a focus on data management and curation assets, to facilitate analysis for each project. You can choose whether to stay for this part of the meeting if this is not your area of interest.
- Minutes of these meetings are circulated to all consortium members.

- **Projects:** each project lead is responsible for setting up meetings as required. Further detail on contributing to the projects can be found below.
- We do not currently run our **Monthly Webinar**, but recordings of previous sessions, including brief updates on the work in the centre across our areas of work and presentations on emerging research outputs from the work of the consortium, can be found on the BHF Data Science Centre's playlist on HDR UK's YouTube channel.

3.2 Collaboration

We have various Slack channels where our TRE users and project leads come together to discuss ways of working within the environments and how best to perform their respective analyses. Contact us to ask for a list of these channels and to be added.

Our consortium documentation and project-related files are held in shared folders in Box, which can be accessed by all consortium members.

4 Patient and Public Involvement and Engagement (PPIE)

The BHF DSC works with patients and the public to ensure transparency, and to build trust in the use of health data for research. We actively seek to record and describe tangible examples of the impact of this involvement in changing, shaping and improving research throughout a project.

We expect you to have engaged in PPIE activities **before** proposing a project, and you will be asked to provide details of any specific groups you are engaging with, how their involvement has influenced the project/research question(s), examples of how their involvement has brought about other changes in your project, and how you will continue involving public/patients throughout the project and in the dissemination of outcomes. Our team of public contributors will review and provide constructive feedback on these plans as part of the project approval process.

You will also be offered the opportunity to present your planned outcomes back to our public contributors. At this meeting, they can advise on routes for effective dissemination with patient groups and the wider general public, and provide suggestions for engagement activities to raise public awareness. You can also work with the public contributors to produce a plain English summary of the results and any other public facing messages.

If you require help with PPIE before submitting a project proposal or as part of an active project, please contact our PPIE Manager at bhfdsc@hdruk.ac.uk. We also encourage you to approach existing PPIE networks or contacts at your host institution, or locally, for support and guidance.

5 Project proposals, approval process and amendments

New projects or amendment requests are no longer being accepted. This section sets out the approvals process that was used prior to that. You can also read about data access costs, how to contribute to existing projects or how to manage your project if it has already been approved.

5.1 Proposing a project

If you would like to lead your own project[1], a project proposal form can be found in Box. The sections to be completed include a plain English summary, project background, the research questions that will be answered, plans for patient/public contributor involvement, an overview of the methods to be used (a detailed plan is not required at this stage), data access funding/conflicts of interest, names of analysts that will work in the respective TREs and required data sources.

Projects may be proposed that will utilise the datasets available in the respective TREs in one or more nations.

Before drafting or submitting a proposal, you are advised to refer to the informative Dataset Summary Dashboard, or liaise with our Health Data Science team (bhfdsc_hds@hdruk.ac.uk), to determine whether your planned study is feasible with the available data.

You should also check whether a project has already been approved which either covers your planned area of research, or overlaps with it, by:

- reviewing the plain English summaries of approved projects on our webpage
- reading the proposals of interest in their respective project folders

Your proposal may overlap with and/or can be completed as a sub-project of an existing project without the need for additional approvals, or with just a minor amendment (see below). Where this is the case, the BHF DSC team will connect you with the project lead or convene a meeting to discuss how to collaborate and/or create a sub-project.

Given the size of our consortium and breadth of expertise amongst our membership across the nations and multiple institutions, we encourage you to take a **team-science** approach when formulating your proposal and to seek contributions/input from as many relevant members as possible. The BHF DSC team will assist with this.

5.2 Funding the cost of data access and conflicts of interest

To make the provision of data for the research community via the consortium more sustainable, **all projects are expected to contribute towards the recovery of the costs of data access, charged by the TRE providers to the centre** (including software licenses, compute, desktop upgrades).

From January 2026, this will include projects with a focus on cardiovascular disease and/or its risk factors. Unless and until funding is in place to cover the full costs of data access, SDE accounts for researchers working on your project(s) will be deactivated and/or new user accounts will not be

[1] Note: if proposing a project to access data for England in NHSE's SDE, the project lead(s) must either be directly employed by, or have an honorary contract with, one of the institutions party to the data sharing agreement with NHSE

activated. Please contact us in advance to discuss the expected level of contribution per analyst, per month, to avoid any interruption in data access.

Note the following:

- Funding for access to data, whether via the centre, previously via the Data and Connectivity National Core Study or elsewhere, must be appropriately acknowledged in any related outputs.
- Data access costs will be pro-rated where an analyst is working on more than one project.
- Funds should be set aside for post-submission access to data to address reviewer comments.
- For the month in which data access is activated or deactivated, you will be charged for a full month regardless of the date in the month that occurs (for access to data for England only).

When submitting a project proposal, you will also be asked to provide detail about any potential **conflicts of interest** in accessing data, for example, whether you have:

- Been asked by and/or have received funds from an industry partner or other organisation to answer the research questions; or
- Have links with an organisation that may have an interest in the outcomes.

The responses provided will be taken into consideration as part of the project approval process.

5.3 Submitting a project proposal and initial triage

Once you have drafted your project proposal and considered how data access will be funded, completed forms should be sent to bhfdsc@hdruk.ac.uk

On receipt, your proposal will undergo an initial triage by the BHF DSC team, including the Whole Population Data Research Project Manager, the Health Data Science team and PPIE Manager, to check that the:

- Project falls within the scope of the approvals in place for CVD-COVID-UK/COVID-IMPACT and is feasible using the available datasets/environments;
- Aims and objectives are clear;
- Project does not unnecessarily duplicate existing projects;
- Requested data sources are available in the TRE(s) (including whether there is sufficient coverage within specific datasets);
- Data source requests align with the needs of the project;
- Language used within the plain English summary is clear;
- Plans for PPIE are described in the proposal.

The initial triage will result in one of three outcomes:

- **Standalone project**: the proposal will proceed to the Approvals & Oversight Board for consideration. Feedback will be provided, and you may be asked to make amendments to your proposal and submit an updated version prior to submission to the Board.

- **Withdrawal**: the proposal may need to be withdrawn for a variety of reasons, including the lack of data availability and/or coverage. Suggestions may be made on adapting the study design or data used.
- **Overlap**: if overlap with an existing project is identified, the BHF DSC team will connect you with the relevant project lead and/or convene a meeting to discuss collaboration, including whether an amendment request to broaden the scope of the existing project is required.

5.4 Approvals & Oversight Board

Following the initial triage, new, standalone project proposals are submitted to the Approvals & Oversight Board for their consideration.

Board membership comprises the following, or their nominated delegate:

- National TRE provider representatives, where projects are being run in their environment (NHS England and SAIL Databank);
- Consortium members representing academic institutions party to the data sharing agreement with NHS England; and
- Public contributors from our panel.

The Board is coordinated/chaired by a member of the BHF DSC core team, but they do not have any voting rights.

On average, the current period from submission of a proposal to outcome notification is four to six weeks.

5.5 Reviewing / approving projects

Project leads attend a meeting with two public contributor members of the Approvals & Oversight Board and provide a brief (approx. 5 minutes) overview of their project (with slides). The public contributors will then discuss the proposal with you in more detail, as well as raise questions and provide constructive feedback on:

- Whether the project is relevant to patients and the public;
- The clarity of the aims and objectives;
- The clarity of the plain English summary;
- Whether adequate PPIE plans are in place.

In parallel, the proposal is shared with the other Board members for review against the same criteria as per the initial triage.

The Board then decides whether each project is approved, approved with conditions, or not approved. The final approval decision for a project lies with the national TRE/data custodian members of the Board, i.e. currently, NHS England and the SAIL Databank.

The BHF DSC team notifies the project lead of the outcome via email. If approved (with or without conditions), the BHF DSC team lists the project reference and title on our webpage, and the HDR UK Gateway, together with the plain English summary (modified where necessary for clarity).

5.6 Contributing to projects

Approved project proposals will be emailed to all consortium members requesting them to contact the BHF DSC team or project lead to register their interest in contributing to the project(s), either from a scientific, clinical and/or analytical perspective.

New consortium members will also be asked to register their interest in contributing to projects during the onboarding process and will be connected with the relevant project leads.

5.7 Amendments and proposal versioning

You will need to submit an amendment request to the BHF Data Science Centre for review by the Approvals & Oversight Board where a:

- **substantial** change is made to the scope of your planned analyses beyond what has already been approved. This includes where the scope of the approvals in place needs to be broadened to allow a team proposing a similar, but not completely overlapping, project to collaborate with you.
- **non-substantial** change is made which does not change the scope of the approved proposal, e.g. requesting access to additional datasets (already available within the relevant TRE), change in start/end dates, minor amendments to the wording of the proposal for clarification purposes only etc.

Notify the BHF DSC of your intention to submit an amendment request and you will be emailed your project proposal document with any sections and/or datasets added, amended or removed (as per the standard form) since your project was initially approved or last amended. You should then:

- Update the sections in the proposal document, as necessary, **using tracked changes**;
- Complete the amendment request form to briefly describe why you need to make an amendment, to state whether the amendment changes the scope of the original proposal and to highlight which sections of the proposal have been updated;
- Return the updated proposal document **and** completed amendment request form.

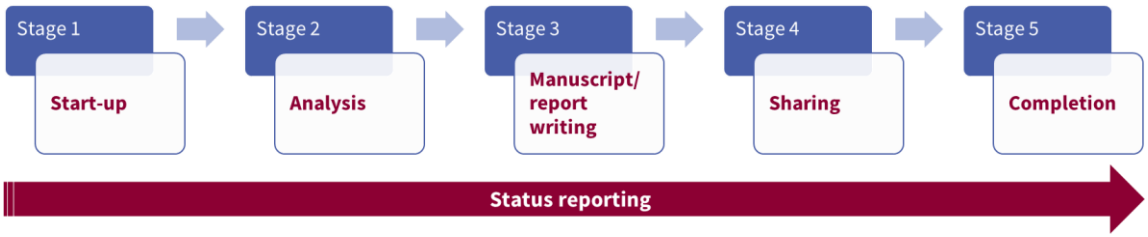
The BHF DSC team will submit a substantial amendment request to the Approvals & Oversight Board for **review and approval** (which may involve a further meeting with the public contributors), whereas non-substantial amendments will be submitted **for information only**. Once approved and/or you have been notified of the outcome, the BHF DSC team will save a clean version of the proposal in your project folder in Box and update the webpage, HDR UK Gateway and GitHub as necessary.

Project proposals will be numbered/versioned as follows:

Version	Description
0.1, 0.2, 0.X	Submitted proposal and iterations before/during review by the Board
1.0	The version approved by the Board
1.1	Approved version, updated in line with conditions/comments from the Board
2.0, 3.0, X.0	Subsequent amendment requests (version number increases by one)

6 Project lifecycle

Once approved, we track the progress of the project through the following stages:



Note that some projects might have multiple outputs/subprojects and/or perform analyses in more than one TRE which run at different rates.

6.1 Stage 1: Start-up

Your project proposal has been approved and it is now time to get your project underway. The Start-up stage is where you:

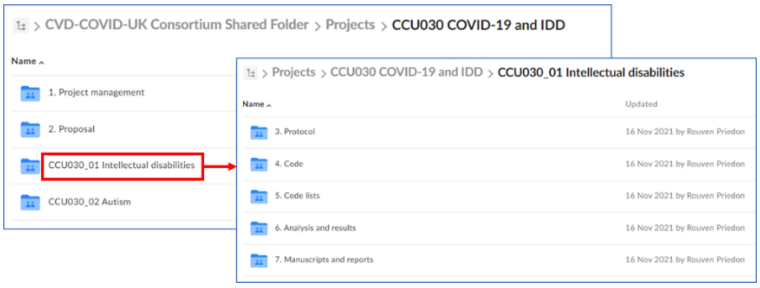
- establish your project's presence in Box, GitHub, webpages and the HDR UK Gateway;
- form your project team;
- request TRE access;
- confirm the number of planned outputs/subprojects and their working titles;
- develop and version control your protocol/statistical analysis plan (SAP);
- create your phenotypes and analysis code, drawing on what already exists.

6.1.1 Establishing your project's presence

The following tasks will be carried out **by the BHF DSC team**:

6.1.1.1 Box

- A project folder will be created here in Box, named with the project reference and an abbreviated name, e.g. CCU030 COVID-19 and IDD.
- The project folder will contain a folder to store your approved project proposal (and any subsequent, approved amendments), with additional folders for each planned output/subproject (see below) to store the related protocol and analysis plans, code, phenotype code lists, files relating to your analysis and results, and manuscripts and reports. These folders will be subdivided by nation, where necessary.



- There will also be a separate folder for project management including meeting minutes, project reporting and project completion activities.

6.1.1.2 GitHub

- A repository will be added to the BHF Data Science Centre's GitHub organisation named with the project reference, e.g. CCU030, containing the approved plain English summary and links to repositories for the related subprojects/outputs.
- Additional repositories will be created for each planned subproject/output (usually once you have drafted your protocol/SAP) containing three standard folders, for each nation in which you are performing analyses, to store the protocol, analysis code and phenotype code lists (note: none of these should contain any identifiable or personal information). Repositories will be named with the subproject reference, e.g. CCU030_01.
- Repository visibility will initially be set to **private**, enabling it to be viewed by invited BHF DSC GitHub organisation members only.
- Project leads can elect to change their project repository visibility to **public** immediately, in the knowledge that the full contents of the repository will be accessible to all (including any interim code etc.).
- Alternately, if project leads wish to make just their protocol/SAP public from the outset, these can be placed in the open-access **Protocols** repository (see Developing and version controlling protocols / SAPs).
- In all cases, repository visibility will be changed to **public** in Stage 4 when you submit your manuscript/letter/report to a preprint server and/or for publication.
- Having a repository per subproject/output allows for protocols/code etc. to be made public for projects with multiple planned outputs, but with differing reporting timelines.
- All code in GitHub remains the intellectual property of the contributing authors and is made available under the Apache version 2.0 license.

6.1.1.3 Website

- The project details and plain English summary will be added to the programme webpage.

6.1.1.4 Gateway and Phenotype Library

- The project will be added to the Gateway (as a data use) and linked to our collection.
- Phenotyping algorithms will be submitted (by you) to the BHF Data Science Centre's consortium collection in the HDR UK Phenotype Library (also see below).

6.1.2 Forming your project team

- Project leads and named analysts working on a consortium project **must be onboarded as a consortium member** by the BHF DSC team and have signed up to the consortium's principles. Other project team members are also welcome to join the consortium.
- As above, we encourage you to **take a collaborative, team-science approach** in forming your team and to work across institutions with any consortium member who has expressed an interest in contributing to your project. This includes using analyst resource with existing

TRE domain/data experience, wherever possible, as this will be the most efficient way of performing your analyses, both from a time and cost perspective.

- The BHF DSC team is also able to assist with providing example job descriptions, advertising open roles or with related funding applications, as appropriate.

6.1.3 Data access and safe researcher/information governance training

Data access for new TRE users

- The project proposal form specifies analysts needing TRE access per nation. The BHF DSC team will discuss these requirements with the project lead to ensure that TRE access is granted in a manner which both maximises efficiencies and is cost-effective. This can be achieved by:
    - Naming analysts to work in one TRE only for any given project, ideally based on existing domain or dataset experience (an updated list of analysts per TRE/project is maintained by the BHF DSC team).
    - Performing analyses in the respective TREs in parallel, which may include working with analysts from within the consortium who already have access to and are familiar with a TRE and its data, and might be able to assist either on a part-time or full-time basis.
    - Requesting access only when the analyst has confirmed their readiness to start actively working in the TRE(s) as soon as access is granted.
    - Requesting access during the first 10 days of each month (for access to the NHSE SDE service for England only, due to NHSE's charging structure).
- To be provided with access to data in any of the TREs, named analysts must:
    - Be **accredited**, i.e. they need to provide evidence of completion of an appropriate safe researcher/information governance (IG) training course[2] within the last 3 years (and will be asked to refresh it as necessary).
    - Have **sufficient analytical capability** to directly contribute to the planned research, with appropriate supervisory and oversight mechanisms in place to support this. **TREs are not to be used as training environments.**
    - Provide a copy of their **honorary and/or student contract**, where appropriate.
- Once the list of TRE users and the date from which they require access is agreed for your project, the BHF DSC team will request their access to the relevant TREs, if not already set up:

[2] Accepted IG training courses include the MRC's 'Confidentiality and Data Protection in Health Research' course, the MRC's 'Research, GDPR and confidentiality – what you really need to know course' (if completed before October 2025), or the ONS Safe Researcher Training course (run by the ONS via the Research Accreditation Service, the UK Data Service, and the Scottish Centre for Administrative Data Research).

- **England (NHS England):** the analyst's institution may need to be added as a party to the data sharing agreement (if this is not already the case) or an honorary contract put in place with an institution that is already party to the agreement[3], both of which can take several weeks (sometimes months).

  Analysts will need to provide evidence of completion of IG training, their job title, and a copy of their honorary/student contract, where applicable (as per sections 5a and 5b of the data sharing agreement). The BHF DSC team will then add/activate the account in the SDE's User Management portal. This will trigger automated emails advising the analyst of the steps to take to fully activate their account.

  First-time users will be required to undertake an online SDE induction, covering navigation, basic usage and output guidance, before they can access the data. As part of **technical onboarding**, they will be invited to a mandatory meeting where a health data scientist will provide a) a further introduction to the SDE; b) details on our ways of working; c) insights on requested datasets; and d) signposting to relevant data curation resources that might help accelerate their project.

  Emails about SDE access will be sent to the email address used to set up the account, i.e. for those with an honorary contract, it will be the email address of the institution with whom they have the contract. In all cases, if users have been advised that their SDE account has been activated, but haven not received the email, they should check if it has landed in their junk mail folder.

  Note: Before receiving the account activation email, analysts might also receive an email from databricks.com – **please ignore this email.**

  If access to **Stata** is requested, it will take NHSE between one to three weeks to assign a licence and an additional cost will be charged.

- **Wales (SAIL):** the analyst's institution will need to be added as an organisation with the Information Governance Review Panel (IGRP) to show the list of organisations involved in the project. Analysts will be added to the list of approved users and must then attend an induction session (arranged and led by Ashley Akbari) and provide evidence of completion of IG training before their access to SAIL and related pages in Confluence are set up. The SAIL new user application process will be completed and the user agreement will need to be signed and countersigned by someone within the analyst's institution.

[3] Imperial College London, King's College London, London School of Hygiene & Tropical Medicine, Swansea University, University College London, University of Bristol, University of Cambridge, University of Dundee, University of Glasgow, University of Leicester, University of Liverpool, University of Manchester, University of Nottingham, University of Oxford, University of Sheffield, University of Southampton

Deactivating TRE user accounts

- Following discussion with users, as appropriate, TRE accounts may be deactivated either on a temporary or permanent basis where analysts:
    - No longer need data access;
    - Have been inactive for over a month and/or are no longer contributing to the analyses for any projects;
    - Have not refreshed their safe researcher/information governance training on request (if completed more than three years ago);
    - No longer have a valid honorary/student contract, where applicable, i.e. the end-date has passed.
- Project leads must **immediately** notify the centre if, for whatever reason, one or more of the named analysts working on their project no longer needs to access data in the relevant TRE(s). This also ensures that data access costs are appropriately controlled.
- **Note:** Data access costs will continue to accrue and be charged back to the project lead, until such time as notification to deactivate the account for one or more users working on their project is received.

6.1.4 Confirming the number of planned outputs (publications)/subprojects

Once your project has been approved, you will be asked to confirm the number of planned outputs (publications)/subprojects and their working titles. It is accepted that these details may change over time, but they enable the BHF DSC team to structure the folders in Box and set up repositories in GitHub, as appropriate.

6.1.5 Developing and version controlling protocols / SAPs

6.1.5.1 Developing protocols/SAPs

A protocol/SAP should be developed for each planned output/subproject. This provides clarity on expected outputs and also removes the risk of methods for other planned outputs being made public whilst still undergoing analyses (unless this is agreed – see above).

As a guide, your protocol/SAP should contain the following sections:

- Version control
- Title (this may differ to the project title if there is more than one planned output)
- Plain English summary and/or background
- Research hypothesis, aims and questions
- List of datasets and reasons required
- Study design with detailed definitions of study population, exposures, covariates and outcomes
- Statistical methods
- References

You should liaise with the following people, as necessary, to understand what curated data sources are available in the respective TREs, as well as any known issues with data coverage, quality and missingness (also referring to the Dataset Summary Dashboard), as this will determine which of your desired analyses you are able to perform (plans may need to be adjusted accordingly):

- BHF DSC Health Data Science team:
  - bhfdsc_hds@hdruk.ac.uk
- NHSE/SDE service for England data wrangler/science team:
  - england.sdeservice@nhs.net
- SAIL Databank data wrangler/science team:
  - Ashley Akbari: a.akbari@swansea.ac.uk
- Methods Group:
  - Lead - Prof Angela Wood: bhfdsc_hds@hdruk.ac.uk

If your project proposal states that you will be accessing data from more than one nation, this should be reflected and taken into consideration when developing your protocol(s) and SAP(s).

Where possible, we encourage you to:

- Perform analyses in the relevant TRE environments concurrently, working with others as appropriate.
- Develop your SAP, and data/variables to be used, to enable outputs and results to be comparable across the nations or to articulate what the differences in approach are, the reasons why and what impact this may have on being able to compare/pool the results.
- Publish results for all nations included as part of the same report/letter/paper.

Once you have drafted your protocol/SAP(s) and defined your output tables, the BHF DSC team will assist you with sharing the document(s) with the wider consortium for review and feedback and to determine whether other consortium members may like to actively contribute to the (sub)project. Consortium members will be given **one week** to provide feedback. You can progress your curation and/or analyses in parallel with this review, although this may be subject to change based on feedback received by the consortium.

6.1.5.2 Version controlling protocols/SAPs

Version controlling protocols/SAPs must take place **before commencing** any analyses (regardless of whether or not you elect to make your protocol publicly accessible before results are published). You should:

- Add a 'version control' section at the start or end of your document where you can track the version number, date, who made the changes and what the changes are (see below).
- Save the protocol and tables, incorporating any feedback from consortium members, in Box with the filename format **CCUXXX_XX_protocol_vX.X_YYMMDD**
- Upload a pdf version of the file to your project's GitHub repository and/or to the open-access protocols GitHub repository (where the document will be time and date stamped).

It is accepted that protocols/SAPs are likely to evolve whilst performing your analyses. Where this proves to be the case, you should carry out the following steps **before continuing** your analyses:

- Create a new, updated version of the relevant document in Box with the filename format **CCUXXX_XX_protocol_vX.X_YYMMDD**
- Update the version control table within the document to summarise the changes made.
- Upload a pdf version of the file to your project's GitHub repository and/or to the open-access protocols GitHub repository (where the document will be time and date stamped).

Example version control table

Version	Date	Author(s)	Comment
0.1	02/07/2020	Will Whiteley	Initial draft.
0.2	06/08/2020	Will Whiteley	Incorporated feedback from consortium members in sections XXX.
1.0	17/08/2020	Will Whiteley	Approved version (uploaded to GitHub).
2.0	14/11/2020	Will Whiteley	Revised version reflecting changes in variables used in section XXX (uploaded to GitHub).

6.1.6 Creating your phenotypes and analysis code

6.1.6.1 Phenotype code lists

Many phenotypes have already been created and curated for the programme, available via the:

- NHSE SDE service for England in the **bhf_cvd_covid_uk_byod** database
- SAIL Databank
- HDR UK Phenotype Library (and indexed in the HDR UK Gateway)

The Health Data Science team can provide support to identify or create additional algorithms, if required (email bhfdsc_hds@hdruk.ac.uk).

6.1.6.2 Analysis code

See the section below for analysis (and curation) code best practice.

6.2 Stage 2: Analysis

You have formed your project team, got access to the relevant TREs, specified your SAP(s) and defined your code lists and phenotypes. You can now start the analysis stage of your project.

You should always:

- Speak with the Health Data Science team to identify, re-use or modify research-ready code, if available. As above, for first-time analysts, this will be covered in a mandatory onboarding meeting that the team will invite you to (Note: remember to acknowledge this in any outputs and to take a copy of any existing code **before** modifying it – you should never modify or delete existing code in other projects' folders);

- Make your code available (share in the collaborative space in the TRE);
- Ensure that your code is fully annotated;
- Advise other analysts (via Slack) if you identify a bug in shared code;
- Have your code reviewed, ideally by an analyst not directly involved in your project;
- Arrange for code lists to be added to/updated in the phenotype library.

If, for whatever reason, you need to alter the methods used in your analyses, you must reflect any changes in an updated version of your protocol/SAP and upload the new version to GitHub **before continuing** your analyses (see above). Please also refer to the amendment process for substantial and non-substantial changes to the scope of your project.

6.2.1 TRE responsibilities

- In line with our principles, analysts are responsible for only conducting analyses that comply with the approvals in place for both the research programme and the project(s) that they have approval to work on.
- Partner institutions vouch for the analysts they employ to enable access to data within the TRE(s). For access to the NHS England SDE, this duty is specified within the Data Sharing Agreement that participating institutions sign (also refer to the responsibilities framework in Appendix B). For SAIL, this is specified in the SAIL Data Access Agreement signed by both the analyst and a representative from their institution.
- Approved projects will have an assigned lead responsible for overseeing the analyses for their project in the TRE(s). The Methods workstream is responsible for coordinating the data management, processing, checking and methodological research carried out to benefit all projects.

6.2.2 TRE best practice

Best practice guidance for working in the respective TREs, including folder/table naming and data management, can be found in the locations below:

- England (NHS England): in Box here
- Wales (SAIL): see the SAIL Wiki (via desktop link in SAIL environment)

This guidance is work in progress and will continue to be improved and updated over time.

6.2.3 TRE technical and data wrangler support

If you experience any technical issues with the respective TREs when performing your analyses, you should collate as much detail about the issue as possible, including screenshots of error messages, and then:

- **England (NHS England):** raise a ticket with the National Service Desk by emailing ssd.nationalservicedesk@nhs.net. Please also post a message on the **bhf-dsc-cvd-covid-uk-analysis** Slack channel to see if any other analysts have the same issue or know of any interim workarounds.

  Suggested detail to include in the body of the email to assist the National Service Desk in directing your ticket to the right group, as well as screenshots, is as follows:

- **System**: SDE
- **NIC number**: 391419_j3w9t
- **Tool**: Databricks / RStudio / Stata (state as appropriate)

Queries on datasets and curation should be emailed to england.sdeservice@nhs.net for a response by one of the members of the NHS England data wrangler team.

The NHSE data wrangler team also runs **drop-in clinics** on Teams every Monday, Wednesday, and Friday from 10-11am to answer any queries: link to the data wrangler clinic.

- **Wales (SAIL):** the preferred option for technical or data issues is to raise a ticket for the Helpdesk (https://jira.hiru.swan.ac.uk/servicedesk/customer/portal/6) to respond to (note that a helpdesk group is set up for our consortium so that tickets can be viewed by other users to avoid multiple tickets for the same issue being raised). Alternatively, you can email helpdesk@chi.swan.ac.uk. You can also post a message on the **bhf-dsc-tre_discussions-wales** Slack channel for the SAIL team to respond to.

In all cases, posting a message on the relevant Slack channel will alert other analysts of the issue (who should follow the process above if it is also impacting them).

Our Health Data Science team are also available to offer support and guidance across all TREs by emailing bhfdsc_hds@hdruk.ac.uk.

6.2.4 Requesting new R/Python packages

- **England (NHS England):** email the request to ssd.nationalservicedesk@nhs.net putting "SDE TOOLS REQUEST: " at the start of the Subject of your email.
- **Wales (SAIL):** users can install any R package if it is on CRAN. Other packages not on CRAN may be imported into SAIL by users themselves, and installed by users themselves, but users must ensure that any files coming into SAIL are from a trusted source and do not breach terms laid out in the SAIL Data Access Agreement.

6.2.5 TRE outputs / statistical disclosure control

As a general rule, outputs requested from the TREs should:

- Not contain any personal or identifiable information or be requested in such a way to enable re-identification (e.g. differencing from previously released outputs);
- Be made at an aggregate/summary level, i.e. not person level;
- Avoid low counts, i.e. <5 and in some cases <10 (e.g. when the output contains sensitive information);
- Be clearly presented (e.g. in the case of code, annotated with comments), so that the output checkers can accurately review;
- Be submitted to allow the output checkers sufficient time to review and assess outputs

Please refer to the TRE-specific guidance below for more detailed information on outputting tables, charts, and code.

Note: interim/draft outputs can be shared with members of the project team or wider consortium only. They must not be onward shared or used in publications.

Outputs can be requested in the respective TREs as follows:

- **England (NHS England):** Please see the "Exporting and downloading results" sections of the Databricks and RStudio User Guidance documents. Please also refer to the Safe Output Service guide which contains a section on disclosure control rules and details the service level agreement (including only requesting 10 outputs per user, per day).
- **Wales (SAIL):** Outputs are requested via the "Data Transfer In-Out" link on the desktop of the SAIL environment. If approved, the user will receive an email with a link to the SAIL portal where the approved output can be downloaded.

6.3 Stage 3: Manuscript/report writing

You are now ready to write your manuscript and publish your results. In line with the Centre's Publication and Dissemination Policy:

- Manuscripts must be published with **open access**.
- Publications must include '**on behalf of the CVD-COVID-UK/COVID-IMPACT Consortium**' at the end of the author list, with all relevant individual contributions (coordination, writing, analysis, interpretation etc.) listed and the coordinating role of the BHF DSC acknowledged. Please see the Acknowledgements section below.
- All research articles must be made publicly and freely available via Europe PMC within six months of the final publication date.
- Research articles must be published with a Creative Commons attribution license (CC-BY).

Please inform us when you are drafting your manuscript and we will send you the '**Offboarding Declaration Form**' to complete, before you submit your manuscript to a journal. In completing the form, you're confirming that the manuscript aligns with the scope of the research programme and of your approved project's stated aims and objectives, and that you have complied with the required aspects of the principles for participation, ways of working, and publication and dissemination policy.

6.3.1 Authorship

We expect those who have contributed to the work to be fairly represented in authorship or acknowledgements. This includes patient and public contributors, the Health Data Science team (where curation resources and support have been provided) as well as consortium members who have provided relevant/substantive comments during the protocol and/or manuscript review process. Note that:

- We encourage the use of CASRAI1's Contributor Roles Taxonomy (CRediT) system or following the ICMJE recommendations, where possible.
- **The final decision on co-authorship lies with the project lead** (the centre will not adjudicate).

6.3.2 Patient/public input and plain English summary of results

- The results of your research should be presented to and discussed with the public contributors who have been involved in your project from existing PPIE networks or contacts at your host institution, where applicable. You will also be offered the opportunity to meet with the public contributor representatives from our Approvals & Oversight Board. This will help to shape your paper, by highlighting the outcomes of interest to the patients/public, as well as ensuring that the outcomes are communicated clearly and effectively, and that plans to do so have been properly considered.
- If you choose to meet with our public contributor representatives, the BHF DSC PPIE Manager, together with the core team, will convene a 30-minute meeting on your behalf, where you will be asked to provide a brief (up to 5-minutes) overview of your outcomes (with a short slide deck), followed by a discussion.
- Public contributors can also assist with preparing/reviewing a plain English summary of results for each output, which will inform communications for your paper at the point of journal publication, such as web stories, social media posts or press releases.
- In the first instance, project leads should complete the Research Outputs Reporting form, which contains sections for a plain English summary of results and communications contacts at your host institution, and email it to both the BHF DSC Communications Manager and PPIE Manager at bhfdsc@hdruk.ac.uk.

6.3.3 Manuscript/reports: consortium review

- Prior to review by the consortium, you should allow **at least 2 weeks** for all authors and the BHF DSC team to perform their review of the draft manuscript. To mitigate issues with meeting any desired timelines, you should notify the relevant stakeholders of when you plan to issue your manuscript to them for review, to confirm their availability and to enable them to free up time, as necessary.
- The centre will check that the appropriate funding, acknowledgement, ethical approval and data availability statements have been included, and alignment with the research programme's aims and objectives. However, **it is the project lead's responsibility to ensure that the manuscript aligns with the research questions stated in their project proposal**. You will be requested to make changes, as necessary, to address any issues prior to review by the consortium.
- Prior to submission to a journal, manuscripts will be circulated to all consortium members for comment. Please allow a minimum of **2 weeks** for this review cycle to take place (also see above regarding authorship).

6.3.4 Manuscripts: journal peer-review

- After your manuscript has been reviewed by the consortium and changes made as necessary, you should:
    - Save a copy of the submitted manuscript document in the project folder in Box.

- Upload the manuscript to a preprint server and share the DOI with the centre (note: uploading the manuscript to a preprint server is optional).
- Submit the manuscript to your target journal and advise the centre when you have done this, so that they can make the GitHub repo publicly accessible (if this hasn't already been done).
- Request that TRE access for named researchers is deactivated, to avoid incurring the costs of data access, but only once all steps of the Sharing stage have been completed, including researchers attending a technical offboarding meeting. Note: TRE access can be reactivated on request, if rework is required as a result of reviewer feedback. Funds should be set aside to cover the costs of data access.
- **Advise the centre if/when the manuscript is accepted by a journal** and share the type-set proofs, so that the centre can check that all acknowledgements have been included, compliance with the other aspects of the centre's Publication and Dissemination Policy (see above), and that the **Offboarding Declaration Form** has been fully completed – **this must happen prior to publication**.

6.3.5 Communications

- The BHF DSC Communications Manager and PPIE Manager should be given **a minimum of 5 working days' notice** of any expected preprint/publication of results. This allows them time to plan for and provide support with the promotion of your paper, to link with partners, including your home institution and funder, and to prepare any communications, if appropriate. You should also supply the contact details of any additional communications stakeholders who may need to be informed, e.g. your institution's press office, if you haven't already done so via the Research Outputs Reporting form. As the major funder of the BHF DSC, the BHF Research Media team will also be notified of any upcoming papers and sent a link to the preprint.
- The BHF DSC will not publicise research as a web story or press release that has not been peer-reviewed, unless in extraordinary circumstances where the findings have an immediate implication for public health. Researchers are particularly encouraged to get in touch with our Communications Manager and press offices at their own institute as soon as possible where findings require sensitive messaging or are of potentially high interest to the public. This allows for appropriate plans to be made across communications teams.

6.4 Stage 4: Sharing

In accordance with our principles, analysis plans, protocols, code (analysis and curation), phenotype code lists and manuscripts arising from your project must be made **publicly available** via a combination of the BHF DSC website, HDR UK Gateway, BHF DSC GitHub organisation, HDR UK Phenotype Library and through open-access publications (note: none of these should contain any identifiable or personal information).

In addition, prior to TRE access being deactivated, researchers will be invited to attend a technical offboarding meeting with a member of the Health Data Science team.

These tasks can be performed in parallel with writing your manuscript/report, however, they should be completed **before** a manuscript is uploaded to a preprint server and/or submitted to a journal.

6.4.1 Uploading your protocol, code and phenotype code lists to GitHub

The following material should be uploaded to the GitHub repository for your subproject:

- The subproject protocol document (if not already uploaded – see above).
- All code (curation and analysis) used to generate analytical outputs, visualisations and results included in the manuscript, both inside and outside of the TRE. Code should be sequenced and annotated to provide a clear description of what each script is doing.
- All phenotype definitions, including a file citing the accession IDs (including version) or DOI of all phenotype definitions and code lists uploaded as csv files (see section 6.1.6.1) which, as a guide, should contain the following fields:
    - Phenotype name (e.g. Hypertension)
    - Code (e.g. I10)
    - Code term (e.g. Essential (primary) hypertension)
    - Code type (where 1 is incident codes, and 0 indicates codes for historic/prevalent events)
    - Code position (i.e. primary, secondary, or any diagnostic code position in hospital admissions)
    - Code terminology (e.g. ICD-10, Read V2, SNOMED-CT,…)
- A ReadMe including a description of the project.

Please see CCU002_01 as an exemplar GitHub repository and contact the Health Data Science team at bhfdsc_hds@hdruk.ac.uk with any questions on constructing your code-lists for GitHub.

6.4.2 Submitting phenotype definitions to the HDR UK Phenotype Library

You are required to submit **all** phenotype definitions (including code lists) to the HDR UK Phenotype Library, where they have not already been submitted.

Easy to follow instructions have been prepared, which can be found here. We suggest that researchers familiarise themselves with the format and information required for submission in advance, and keep appropriate records.

If you have any questions or require assistance translating your phenotype definitions and code lists into the required format, please contact the Health Data Science team at bhfdsc_hds@hdruk.ac.uk. Any problems or bugs experienced when uploading to the Phenotype Library should be reported to the Phenotype Library team.

6.4.3 Website and Gateway

The BHF DSC team will add:

- The output title, and links to the preprint and GitHub repository, below the project's plain English summary on the programme webpage (a link to the published paper or letter to editor will also be added, once available)

- Details of the output on the downloadable research outputs summary document on the programme webpage and as an entry on the publications page.
- An entry reflecting the preprint/paper on the Gateway, and linking it to the centre's Gateway collection, and the related TRE dataset and project/data use (which have additional links to the project's repo in GitHub).

6.4.4 Technical offboarding meeting

Before you submit your manuscript to a journal and/or any TRE accounts are deactivated, named researchers are required to accept an invitation from the Health Data Science team to attend a **technical offboarding meeting**, to ensure that:

- **Table storage in the TREs has been tidied up**, with tables following the agreed naming convention and being clearly marked as those which should be kept to aide reproducibility (e.g., the final analysis-ready table(s)) and those that can be deleted once the paper has been accepted/published by a journal (e.g., intermediate or exploratory tables).
- **Code/files are shared in the TRE**, including code and codelists in personal workspaces or that which has not yet been pushed to GitLab.
- Any **lessons learned, developments, and feedback on technical work has been shared** with the Health Data Science team to help accelerate future research projects.

If TRE accounts are reactivated to address feedback following journal peer review, this might require further tidy-up of table storage and sharing of code/files etc.

6.5 Stage 5: Completion

Within one month of all planned outputs for your approved project being published, you should submit an end of project report.

The BHF DSC team will convene a lessons-learned meeting with the project team and relevant stakeholders.

7 Project planning and status reporting

7.1 Project planning

Once your project has been approved, for each subproject you will be asked to provide estimates on when you expect to complete each stage of the project lifecycle. This will assist the BHF DSC team with managing the overall programme.

7.2 Status reporting

The BHF DSC team will meet with or email project/subproject leads, or another designated member of the project team, every quarter to:

- Obtain an update on the status of each subproject;
- Understand if any timelines have changed;
- Highlight any issues/blockers that you need assistance from the BHF DSC team to mitigate.

You may also be asked to provide verbal updates on the status of your projects/subprojects at the monthly consortium meeting, as appropriate.

For projects accessing data in SAIL, the status of outputs will need to be updated on your project's page in SAIL Confluence at least once a month (as this goes back to SAIL, TAG and SAGE via agreed routes in Wales). Please complete these updates in a timely manner, on request.

8 Acknowledgements

To appropriately acknowledge our funders and recognise the use of patient data in the respective TREs in our research outputs and publications, please refer to the acknowledgement guidance document in Box and include the relevant statements.

The BHF DSC team will check that the appropriate statements have been included during their review of the draft manuscript/report.

9 Abstracts, presentations and posters

You are welcome to submit abstracts, give presentations and create posters. **Prior to submission/use**, please send draft documents to the BHF DSC team for review. Where possible, you should allow **a minimum of 5 working days** for this review to take place.

Please ensure that:

- The content aligns with both the aims of the research programme and of your approved project.
- Where the results have not yet been published in a journal, you clearly state that you are presenting preliminary results which have yet to undergo/complete peer review.
- For abstracts, the following is added at the end of the author list:

  on behalf of the CVD-COVID-UK/COVID-IMPACT Consortium
- For presentations, whether to the consortium or externally, the BHF DSC logo is displayed in the slide deck (various logo options are available in Box here).
- For posters, you can use one of the poster templates located in Box here.
- The following acknowledgement statement is included (word-count/space permitting, this may be abbreviated to the first sentence only):

  This work was carried out with the support of the BHF Data Science Centre led by Health Data Research UK (BHF Grant no. SP/19/3/34678). This study made use of anonymised data held in NHS England's Secure Data Environment service for England, the SAIL Databank and the Scottish National Data Safe Haven, and made available via the BHF Data Science Centre's CVD-COVID-UK/COVID-IMPACT Consortium. This work used data provided by patients and collected by the NHS as part of their care and support. We would also like to acknowledge all data providers who make health relevant data available for research. The BHF Data Science Centre's Health Data Science Team provided data curation resources and support.*

  ** Delete as appropriate*

Appendix A: CVD-COVID-UK / COVID-IMPACT Consortium: Principles for Participation

- CVD-COVID-UK / COVID-IMPACT commits to the 'Five Safes' (http://www.fivesafes.org/) and to a transparent and inclusive approach, enabling additional researchers, research groups and, where appropriate, institutions, to join and contribute to the consortium as the work progresses and evolves.
- All project analysis plans and protocols, listing data requirements, will go through the CVD-COVID-UK / COVID-IMPACT approvals mechanism process before work commences.
- All analysis plans, protocols, code, phenotype code lists and reports arising from this consortium's work will be made publicly available via the HDR UK website and/or Health Data Research Gateway (linking to additional institutional documentation if appropriate), HDR UK Phenotype Library, repositories in the BHF Data Science Centre's GitHub organisation and through open-access publications.
- All reports for government advisory groups and policy makers, the lay public and academic publications will be written in the name of the collaborative group ("on behalf of the CVD-COVID-UK/COVID-IMPACT Consortium") with all relevant individual contributions (coordination, writing, analysis, interpretation etc.) listed, the coordinating role of the BHF Data Science Centre and any relevant funding (including BHF Data Science Centre funding) acknowledged. In addition, some data custodians require specific acknowledgement of use of their dataset in reports and publications arising from the work of the consortium. Prior to submission, all reports and manuscripts for potential publication must be checked for compliance with these requirements by the BHF Data Science Centre coordination team who will ensure that these checks are conducted in a timely manner.
- A list of the members of the consortium and their institutional affiliations will be maintained by the BHF Data Science Centre coordination team and placed in the public domain on the BHF Data Science Centre webpages / website.
- Access to data within trusted research environments (TREs) for relevant researchers within the consortium will require institutions to sign various agreements. Relevant members of the consortium should do everything they can to ensure that this is done as quickly as possible so as not to impede the progress of our research. In particular, when institutions are added as joint data controllers to the data sharing agreement required for researchers to access the NHS England SDE, or when new data sources are incorporated into that SDE, existing signatories will be required to re-sign the agreement. Sign off is a simple and rapid electronic process. The BHF Data Science Centre coordination team will work with NHS England to limit the number of times this agreement needs to be re-signed.
- Accredited researchers working within the TREs agree to only perform analyses for the purpose of producing outputs for projects approved by the CVD-COVID-UK/COVID-IMPACT Approvals & Oversight Board, and within the scope of the project approvals.

Appendix B: Responsibilities framework for COVID research in the NHS England Secure Data Environment Service for England

(V2.0 – 17/11/2025)

1. Background

NHS England has developed a secure data environment (**SDE**) service for England, within which approved researchers who are members of the CVD-COVID-UK/COVID-IMPACT consortium are conducting COVID-related research on health data from the population of England.

NHS England's SDE service for England is a secure analysis environment that makes anonymised person-level data (i.e. data that has had all directly personally identifying information, such as name, address, exact date of birth, and NHS numbers removed) available for research. Access to the data within the SDE is restricted to approved, named researchers with certified training in the safe use of health-related data for research. These approved researchers access the data held in the environment remotely through a highly secure authentication system. The researchers conduct analyses of data for the purposes of approved research in the SDE. Before researchers can share results of their analyses outside the environment (e.g. within published reports or presentations), trained output checkers at NHS England first check the results of analyses conducted within the SDE to ensure that no person-level data or potentially identifiable data are included.

2. Access to the NHS England SDE service for England

Before researchers based in CVD-COVID-UK/COVID-IMPACT consortium partner organisations can obtain access to the SDE, the following conditions must be met:

- individual researchers must become consortium members and agree to the consortium's principles (see CVD-COVID-UK/COVID-IMPACT Ways of Working section 2);
- the partner organisation must have a Data Sharing Framework Contract (DSFC) in place with NHS England;
- the partner organisation must meet the required standards of the NHS England Data Security and Protection Toolkit (DSPT), or equivalent; and
- the partner organisation must sign the Data Sharing Agreement (**DSA**) between NHS England and all partner organisations jointly accessing data within the SDE supporting the CVD-COVID-UK/COVID-IMPACT consortium's work. In signing the DSA, each partner organisation agrees to ensure that researchers from that organisation who will access the SDE have appropriate information governance and data protection training in relation to the use and storage of health data and have a confident understanding of their responsibilities (DSA ref: DARS-NIC-381078-Y9C5K).

3. SDE Responsibilities

- NHS England will:
  - ensure that an appropriate system security policy and appropriate technical and organisational controls are in place for the SDE in order to ensure the security of all data processing and personal data in the SDE – these measures include the

provision and operation of a secure authentication system for user access to the SDE and ensuring that only anonymised data is uploaded to the SDE for analysis (i.e. with the direct identifiers removed), as defined within the approved DARS application and associated DSA;
    - design and maintain the SDE (including related access processes) in a manner that meets the data minimisation and other principles under data protection law principles and ensures that the data in the SDE is accurate and maintained;
    - make available information to the data subjects to whom the data in the SDE relates (the **Data Subjects**) regarding the processing activities and joint control arrangement in an appropriate and legal manner;
    - be responsible for responding to any requests from the Data Subjects seeking to exercise their rights under data protection law in a timely manner;
    - carry out an assessment of potential high risks to the rights and freedoms of natural persons that may arise from the data processing activities relevant to the SDE and make this data protection impact assessment available to partner organisations upon request;
    - make itself available as the first point of contact for supervisory authorities or individuals that have questions or legal requests relating to the data in the SDE; and
    - be responsible for notifying data protection supervisory authorities and, where applicable, notifying affected Data Subjects in the event of a personal data breach affecting the data in the SDE.
- Each partner organisation that is a party to the DSA will ensure that researchers from that organisation who access the SDE:
    - are aware that they must not share their personal SDE authentication and password with any other person;
    - only access the SDE from a secure device approved by the partner organisation;
    - do not use insecure software when accessing the SDE;
    - do not access the SDE from a public place or other insecure environment (e.g. coffee shop or public transport);
    - follow any guidelines or policies issued in relation to the SDE and wider project;
    - do not attempt to remove any data from the SDE except in accordance with the approved processes for extracting outputs from the SDE (i.e. subject to review of the output by trained NHS England personnel);
    - do not attempt to re-identify any of the Data Subjects to whom the data in the SDE relates;
    - only conduct analyses that comply with the relevant approvals in place for both the research programme and the project(s) that they have approval to work on; and
    - promptly notify NHS England in the event that they receive a request from a Data Subject seeking to exercise their rights under data protection law.
- Each of NHS England and the partner organisations will also be responsible for maintaining a record of its processing activities relevant to the SDE as required by UK GDPR and complying with their respective obligations under data protection law.
- Each of NHS England and the partner organisations will notify Health Data Research (HDR) UK before communicating with a supervisory authority in respect of the processing and take into account any reasonable comments made by HDR UK.

- Only aggregated summary data can be exported from and shared outside of the SDE (for example in reports to health policy makers or manuscripts for publication in medical and scientific journals). NHS England will ensure that analysis outputs and code are subject to appropriate disclosure control by trained output checkers to ensure that no potentially identifiable data (on patients or practitioners) are released.
- A concise, transparent privacy statement will set out the purposes for processing personal data and explain how data is used for these purposes. This privacy statement is maintained by the BHF Data Science Centre on behalf NHS England and the consortium partner organisations and accessible on the BHF Data Science Centre webpage.

Version Control

Version	Date	Author(s)	Comment
1.0	28/05/2021	RP, JMA	• Final version incorporating feedback from WW, AA, SD, CS, LM, DR.
1.1	09/06/2021	RP	• Section 6.3: notice to comms manager changed to a minimum of 5 working days. Included requirement to circulate the manuscript to all consortium members for information/comment. • Section 8: Expanded HDR to Health Data Research in Acknowledgements
2.0	25/05/2022	RP, LM, TB, JN, DR, CS, SD	• Complete refresh of the document
2.1	10/11/2022	RP, LM, KMA, SK, JN, TB	• Section 4 and 6.3.1: offer of a meeting with the public contributors to discuss outcomes/results dissemination • Section 5.2: funding via D&C to March 2023 • Section 6.1.3: addition of detail for NHSD TRE induction and notification when users no longer need TRE access • Section 6.1.6.1: phenotype code list process changes • Section 6.3.3: greater clarity on communications for preprints and publications • Section 6.4.1: further guidance on uploading phenotype code lists to GitHub
2.2	18/05/2023	RP, LM, SK	• Section 3.1: updated link for YouTube • Section 4: requirement to engage in PPIE prior to project submission • Section 5.2: clarify the position on funding data access and conflicts of interest • Section 6.1.3: charging for data access to continue until advised to revoke access • Other minor amendments, including changing NHS Digital to NHS England
3.0	21/08/2023	RP, JM, JF, SD, TB, JN	• Updates to links to direct to the centre's new website • Updates following migration from NHSE's TRE to SDE • Sections 5.1 and 6.1.5.1: directing researchers to the Dataset Summary Dashboard and HDS team to determine what studies are feasible with the available data • Section 6.1.3: addition of detail to highlight that: ○ Data access should only be requested when analysts are ready to actively work in the TRE(s)

			○ Access to NHSE's SDE will only be requested in the first 10 days of each month (due to NHSE's charging structure) ○ Analysts waiting for access to NHSE's SDE should ignore the email received from Databricks before they receive the SDE account activation email. • Section 6.1.6.1, 6.4.2 and Appendix A: clarify the requirement for researchers to submit phenotyping algorithms to the Phenotype Library • Section 6.2.3: amendments to the process for raising issues with the National Service Desk re the NHSE SDE, and new link to join the NHSE data wrangler drop-in
4.0	17/11/2025	RP, LM, AW, JM	• Complete update of the document, to also reflect no new projects being accepted under the NHSE data sharing agreement, with changes made to sections throughout
4.1	04/12/2025	RP	• Minor amendments to various sections in response to feedback from AW and JM.