

# ODAP Manual

**Rule for all communications: We absolutely must include both sides of the platform in all communications/diagrams. If we don't people get confused about what ODAP is.**

## **What is the Outbreak Data Analysis Platform (ODAP)?**

The ODAP is the overarching term for a range of computers supporting research within a specific scope. These include the ODAP TRE (formerly called: flexible compute space; ultra; ultra2), and various project spaces in the National Safe Haven.

It is divided into two sections:

1. The “UNDER EMBARGO” section – in which data are wholly controlled by contributors
2. The “NO EMBARGO” section – in which data are accessible to appropriately-qualified researchers studying questions within the scope of the platform.

The foundation for the ODAP is the International Severe Acute Respiratory Infection Consortium (ISARIC) Clinical Characterisation Protocol (CCP).

## **Scope of the ODAP**

**The purpose of the ODAP is to facilitate biomedical research to advance understanding of severe infectious disease\* and other exposures of public health interest.\*\* Research within the ODAP is strictly limited to this purpose.**

\* Severe infectious disease - this term describes all severe infectious agents, including new, re-emerging or therapy-resistant forms of existing infectious agents.

\*\* Other exposures of public health interest: this term describes new or unexplained poisoning, or exposure to harmful energy sources such as electromagnetic radiation.

# ISARIC Clinical Characterisation Protocol (CCP)

The scope of the ODAP mirrors the objective of the CCP, an ethically-approved research study in the UK (Joint Chief Investigators: Calum Semple(Liverpool, Oxford) and Kenneth Baillie (Edinburgh, Oxford)). A broad range of scientists with relevant expertise have come together to form a UK-wide group: the ISARIC Comprehensive Clinical Characterisation Collaboration. Membership of this collaboration is by invitation and is extended to researchers performing high-quality biomedical research to advance understanding of severe infectious disease and other exposures of public health interest.

## The ISARIC Spine

Some studies, addressing aims that are within the scope of the CCP, have chosen to contribute data into the ODAP for linkage to CCP and other datasets. Because these studies are addressing questions that are directly within the scope of the CCP, where appropriate, participants in these studies are eligible to be enrolled under the CCP.

## Examples

The ISARIC4C consortium will incorporate participants from the PHOSP, GenOMICC and COG-UK studies into the ISARIC4C study, creating a superset of participants in a single “spine”. Doing so would allow them to match to NHS data under existing ISARIC4C data agreements.

To do this, we need to have an ID and an NHS number, or CHI number, as a minimum standard. Some participants in the other studies may already be in ISARIC4C. These need to be identified and tagged as being the same individual. This will be done by comparison of NHS or CHI numbers. The spine will contain all the joins between projects and participants. Those participants that are not already in the ISARIC4C system will have an ISARIC4C ID allocated to them.

The data from the spine will be used for several purposes:

- it will feed into the NHS Digital data matching system. This system requires ID, NHS number, date of birth and postcode (the latter two are optional). This will allow the ongoing monthly matching of study participants to English & Welsh Health Data for the superset of ISARIC4C participants.
- it will allow study participants to be identified against a cohort of NHS or CHI numbers for individual research studies, and as requested by Trusted Research Environments (TREs).

Currently the data for participants in these studies are held in a series of REDCap databases. These are the source of truth for the data and update on a regular

basis. We will therefore need a process to update the spine from these sources.

## Computer architecture

The ODAP consists of two main compute areas.

### Flexible Compute Space

The Flexible Compute Space (FCS) is a Trusted Research Environment with large scale compute capacity and access to a range of software tools for data analysis and machine learning. Data within the FCS can be either under embargo, or not under embargo.

Access to the FCS is provided to researchers through *private project zones*

### PHS National Safe Haven

Data within the National Safe Haven can be either under embargo, or not under embargo.

Embargo area is data not yet published or available for use, non-embargoed is data that can be accessed by bona fide researchers with no DSA required.

### API

There will be a secure API in place from the Flexible Compute Space to External TREs and UK Public Health Agencies. This is to accelerate research by enabling data linkage across multiple data sets using unique identifiers to facilitate secure transfer of specified data fields e.g., viral sequencing data. The future purpose of the API is more in line with providing a mechanism for quick pandemic updates to national health services.

### Data tranfers

The preferred way to move data between another TRE and ODAP are the already established methods for that TRE.

## Governance

### Background

ODAP creates a UK-wide capability by curating and linking outbreak relevant data from clinical records, research studies and audit data. It brings together key initiatives and leadership across the UK including ISARIC, COG-UK, MRC CLIMB and GenOMICC.

The platform combines a national Trusted Research Environment (TRE) infrastructure collocated with >£100M of world-class computational and data science capacity including the UK National Supercomputer, with a UK-wide governance framework.

## **The ODAP Partnership Steering Group**

The ODAP Partnership is a UK-wide body which decides on strategic matters to facilitate vital research, supporting additional data linkages, and democratising access to research datasets with full agreement from the contributing studies.

The Partnership aims to strengthen the UK-wide NHS/academic partnership between the CLIMB COVID viral sequence analysis system, the Outbreak Data Analysis Platform, ISARIC-4C, COG-UK, the four UK Public Health Agencies, ICNARC, ONS and HDR UK.

The ODAP Partnership steering group exists to provide advice, support and challenge to ODAP delivery team and will have voting rights on decisions not made by consensus.

## **The ODAP Delivery Team**

ODAP Delivery team to report progress to Data and Connectivity Delivery Group

## **Specific Roles**

University of Edinburgh (as the lead institution for ODAP) is accountable to Health Data Research UK for the delivery of ODAP.

HDRUK is accountable to UKRI for delivery of the Data and Connectivity National Core Study

## **Data management**

1. The Lead Institute will become a Data Controller to facilitate the onboarding of key data sets to the ODAP and to facilitate external access to ODAP only.
2. All Data Contributors (those who hold the key data sets) will sign a Data Contributor Agreement (DCA) with the Lead Institution for the following purposes:
  - To allow the Lead Institute to onboard, curate and link the data using ODAP.
    - The Lead Institution will manage process to enable access to data however Data Contributor must give permission.

- This Data Contributor Agreement will require a due diligence process to ensure any data can be accessed by researchers for the agreed purposes. These agreed purposes will clearly articulated and decided with public/patient input.
3. This access facilitated through the Gateway using the 5 safes data access form.
    - This governance structure sets out appropriate safeguards for privacy and effective use of sensitive data when it is embedded throughout each stage of a project
  4. The Lead Institution + Data Contributor can together approve researchers access to data on ODAP if:
    - Research is for already agreed purpose. There will be specific research purposes detailed in the DCA which will all relate to patient outcomes.
    - Any linkage has been detailed in the methodology and is in an approved format. This may include only exporting certain agreed data points such as a co-morbidity score, vaccination status, covid-19 severity score, and/or outcomes.
    - All aspects of the project satisfies the 5 safes.
  5. If a request is outside of the agreed purposes the Lead Institution will go back to Data Contributor and request permission.
    - This should be done either through a Data Access Committee formed between the Data Contributors or by the relevant Data Contributor only. The DAC should include public/patient representatives.
    - New project approvals will always be subject to the Data Contributors approval and a DAC should be convened of Data Contributors whose data is involves.
  6. The Lead Institution will sign a Data Access Agreement with any Institutions accessing the data).
  7. All aspects of any projects will be subject to clear, transparent and detailed processes.
  8. The Data Contributor will remain a Data Controller of their data sets. Lead Institution becomes independent controller for facilitating access and onboarding to ODAP only.

## Information for data contributors

### How do I add my dataset (within scope) into ODAP?

Check the scope

Within scope, arrange a discussion by emailing: ODAP@ed.ac.uk

Speak with the ODAP team

Agree what data and where it needs to be hosted: clinical consented data SNSH TRE and/or analysis data ODAP TRE

Does the dataset need to be linked with other datasets? Scottish health data; English health data

Can the dataset be added to the ISARIC Spine?

ODAP team will advise on most appropriate agreement: joint data controller; data processor

Agreements executed/signed off

Data to be transferred into ODAP and added to spine and/or

Refer to the the ICODA due diligence: <https://icoda-research.org/hdr-uk-as-data-controller-for-icoda>, for incoming: 1. Datasets 2. Legal basis for use 3. Ethical review 4. Public and patient review

## **Access for data analysts**

**What are the datasets are available for research analyses?  
What process(es) must I follow? What approvals must I obtain?**

A live ODAP datasets mapping.xlsx is available for reference. [Further details about processes and approvals to be added]

A parallel document exists for eDRIS

## **Process for Application for data access for non-embargoed data**

1. Details of the dataset will be published on the ODAP website; HDR UK Innovation Gateway and/or the data controller's website. HDR-UK will ensure that appropriately trained researchers (and not just COG-UK or ISARIC4C) can easily access data.
2. Researchers will submit initial data request via the HDR UK Innovation Gateway.
3. Applications will be discussed informally with (i) the data processor (ODAP Data Management and Monitoring Team), and (ii) with the data controller (i.e. UKHSA).

4. Applications will complete the full 'Five Safes' form and submits it via the HDR UK Innovation Gateway.

**TRIAGE/APPROVAL:**

5. The ODAP Internal Triage Secretariat (i.e. the ODAP Data Management and Monitoring Team) will review and triage applications based on the 'Five Safes' form. If required, data requests will be discussed with requestors.
6. Data access approval is required from two sources:
  - a. The ODAP Internal Triage Secretariat or the ODAP Data Access Committee
  - b. The data controller (i.e. UKHSA)
7. The ODAP Internal Triage Secretariat or the full OPAP Data Access Committee, when deemed necessary for complex requests, shall approve or decline requests on behalf of ODAP. The OPAP Data Access Committee should always have sight of data access requests even if only for information.
  - The data controller (i.e. UKHSA) will be sent details of all ODAP approved requests. Where UKHSA is joint controller, requests will be sent to a central UKHSA coordinator generic in-box; Covid19genomics@phe.gov.uk.
  - Where UKHSA is a joint controller, UKHSA coordinator arranges the UKHSA response on whether approval is granted for specific projects to access genomic data.
    - The response will be made within 7-days
    - The UKHSA decision will be based on joint opinion of:
      - \* Ian Harrison (Genomics data product owner)
      - \* Sam Organ (Joint Chief Data Officer)
      - \* Saheer Gharbia (Genomics Board)

**GOVERNANCE & LOGISTICS:**

8. The ODAP Data Management and Monitoring Team initiate process for data access with eDRIS National Safe Haven) and/or Edinburgh (FCE)
9. The ODAP Data Management and Monitoring Team
  - a. Prepares data sharing agreement between the requestor and data processor/controller and relevant contracts team
  - b. Ensures appropriate sign-off.

**DATA RELEASE FROM RELEVANT TRE:**

10. The ODAP Data Management and Monitoring Team will review figures and outputs of data analysis performed in the TRE before they are removed from the TRE .
11. For figures and outputs created using genomic data, the data controller (i.e. UKHSA) will also need to review and authorise removal of data.
12. DATA CREATION PROCESS

#### **LINKAGE IDENTIFIER USED IN TRE (NATIONAL SAFE HAVEN AND FCE)**

- ISARIC4C-ID will be used as the unique identifier. ALL COG participants will be enrolled into ISARIC4C. This means that the ISARIC4C-ID can be used and we can link to NHSD health records
- Data compartmentalised within EPCC prevents data being linked to other datasets held by EPCC without approval. EPCC environment prevents copying or removal of data from the relevant TRE.

### **How can I transfer my data files...? (not outputs but for further analyses)**

Refer to: ISARIC4C Data Transfer Processv1.docx

### **How can I get outputs from ODAP?**

Refer to:

NA National Safe Haven Statistical Disclosure v5.docx

FCS Output Request Form - 20220608.docx

### **How is ODAP communicated? Where can I refer people to about ODAP?**

Discoverability:

HDR UK Innovation Gateway

- CO-CONNECT:
  - CO-CONNECT is HDR UK programme: limited number of cohorts from ISARIC4C