**CLS data access application form**

The CLS Data Access Committee (CLS DAC) welcomes applications to access CLS research data not publicly available via the UK Data Service (UKDS) or other data repositories and biological samples.

***Please read the*** [***CLS data access guidelines below***](#_CLS_Data_Access_2) ***prior to completing this form.***

## Main data applicant

For applications by students, the main data applicant must be the project supervisor.

|  |  |
| --- | --- |
| **Name:** |  |
| **Affiliation and work address:** |  |
| **Academic email address:** |  |
| **Website (e.g., ORCID or institutional website)** |  |
| **Telephone:** |  |

## Research team and collaborators

Please list the names of all members of your research team who need access to the data, their affiliation, and their contact email addresses.

|  |  |  |
| --- | --- | --- |
| **Name** | **Affiliation** | **Email address** |
|  |  |  |
|  |  |  |
|  |  |  |

You may add rows for additional users if required.

## Type of data requested

[ ]  Main survey data

[ ]  Linked administrative records

[ ]  Paradata

[ ]  Genetic data

[ ]  Biological samples

[ ]  Other (please specify)

|  |
| --- |
|  |

## CLS cohorts

[ ]  National Child Development Study (NCDS, or the 1958 Birth Cohort Study)

[ ]  1970 British Cohort Study (BCS)

[ ]  Next Steps (Next Steps)

[ ]  Millennium Cohort Study (MCS)

## Title of the research project

|  |
| --- |
|  |

## Plain-language abstract summarising the research project

Please provide a short description of the proposed research you plan to do with the data, including the main research question(s). Maximum 150 words.

The plain-language abstract will be published online and should be written in accessible language that cohort members would readily understand. For further guidance on writing a plain-language abstract, please see Appendix 1 of [the CLS Data Access Framework](https://cls.ucl.ac.uk/wp-content/uploads/2020/12/CLS_Data_Access_Framework.pdf).

|  |
| --- |
|  |

## Project keywords

Please add at least five keywords related to your project, for publication online.

|  |
| --- |
|   |

## Description of the research project

Please provide a full and detailed description of the purpose for which the data are requested, describing the aims of the study/research, methodology, and ethico-legal considerations. Please specify if the research project is part of a larger programme.

|  |
| --- |
| 1. Research project description *(up to 500 words)*
2. Brief methodology description *(up to 500 words)*
3. Ethico-legal issues, in particular regarding sensitive or socially controversial topics. *We strongly advise that you consult section 6.8 of* [*the CLS Data Access Framework*](https://cls.ucl.ac.uk/wp-content/uploads/2020/12/CLS_Data_Access_Framework.pdf) *for guidance.*
 |

## Incidental findings

Please describe the likelihood of generating clinical incidental findings as a result of a genetics project or analysis of biological samples. We strongly advise that you consult see section 6.7 of [the CLS Data Access Framework](https://cls.ucl.ac.uk/wp-content/uploads/2020/12/CLS_Data_Access_Framework.pdf) for guidance.

|  |
| --- |
|  |

If applicable, do you have a clinical expert or geneticist to assess potential incidental findings?

[ ]  Yes. Please provide the name of the expert:

|  |
| --- |
|  |

[ ]  No

[ ]  Not applicable

## Has the project been funded?

Funding is not mandatory but provides an additional reassurance for the project.

[ ]  Yes. Please provide the name of the funder:

|  |
| --- |
|  |

[ ]  No. If the project is not funded or you are waiting to hear from a funder, please provide a statement on whether the project will still go ahead without funding and how you will proceed:

|  |
| --- |
|  |

## Project dates

Please specify when you estimate that the project will start and finish.

|  |  |
| --- | --- |
| Start date (MM/YYYY) |  |
| Finish date (MM/YYYY) |  |

## Non-genetic data requested

Please also complete **section 15** for requests of genetics data, and **section 16** for requests of biological samples.

### Description of data requested

Please describe the data requested and provide as detailed a description as possible.

For phenotypic data requested as part of a genetics application, please specify how the variables will be used in the research analysis (e.g., as exposures, confounders, outcomes, exclusions, etc.).

|  |
| --- |
|  |

### List of phenotypic variables to be combined with genetics data

For phenotypic data requested as part of a genetics application, please use the MS Excel [NCDS data dictionary](https://cls.ucl.ac.uk/wp-content/uploads/2020/12/NCDS_Data_Dictionary.xlsx) or the [MCS data dictionary](https://cls.ucl.ac.uk/wp-content/uploads/2020/12/MCS_Data_Dictionary.xlsx) to provide a full and exact list of phenotypic variable names that need to be linked to the genetics data. The data dictionaries can be found [here](https://cls.ucl.ac.uk/data-access-training/data-access/accessing-data-directly-from-cls/).

To select the variables:

1. NCDS: highlight the requested variables.
2. MCS: mark which variables are required in the column titled “Requested”.
3. The selected variables must match the research project.
4. Do not delete unwanted variables from the variable list.
5. Do not change the formatting of the data dictionaries.

Please confirm that the spreadsheet of requested variables is attached to this application:

[ ]  I have attached the NCDS data dictionary

[ ]  I have attached the MCS data dictionary

Queries concerning variables can be directed to *clsfeedback@ucl.ac.uk*.

## Data outputs

Please describe what derived variables or other individual-level data outputs you plan to generate or compute as part of your proposal. Relevant data outputs and documentation must be sent to CLS if requested.

|  |
| --- |
|  |

## Data access

Please specify the server where the data will be accessed and stored and indicate if it is different to the organisation/institution servers specified in section 1.

|  |
| --- |
|  |

## Genetic data requested

Information about CLS genetic data resources is described [on the CLS website](https://cls.ucl.ac.uk/data-access-training/genetic-data-and-biological-samples/).

### NCDS genetic data

Please indicate what NCDS genetic data you would need to link to survey/biomedical data:

[ ]  [Genome Wide (GW) genotyping data (GWA)](https://research.ncl.ac.uk/d2k/ourresearch/58forwards/geneticdatatypes/gwadata/). Please specify the subsets:

|  |
| --- |
|  |

[ ]  [Imputed data sets](https://research.ncl.ac.uk/d2k/ourresearch/58forwards/geneticdatatypes/imputeddata/) based on the GW data. Please specify the datasets:

|  |
| --- |
|  |

[ ]  [Epigenetic datasets](https://research.ncl.ac.uk/d2k/ourresearch/58forwards/geneticdatatypes/epigeneticdata/) – one methylation dataset.

[ ]  [Sequencing data sets](https://research.ncl.ac.uk/d2k/ourresearch/58forwards/geneticdatatypes/sequencedata/) – one exome sequence dataset.

[ ]  Other – please specify the data and the data source, e.g., data generated by a previous applicant of NCDS samples or data datasets:

|  |
| --- |
|  |

### MCS genetic data

The MCS genetic data resource is described in the [CLS working paper](https://cls.ucl.ac.uk/wp-content/uploads/2020/08/CLS-working-paper-2020-7-Collection-of-DNA-samples-and-genetic-data-at-scale-in-the-UK-Millennium-Cohort-Study.pdf). Please indicate what MCS genetic data you require:

|  |  |  |  |
| --- | --- | --- | --- |
|  | Non-imputed | Imputed | Illumina GenomeStudio Final Report |
| Cohort member data | [ ]  | [ ]  | [ ]  |
| Parental data | [ ]  | [ ]  | [ ]  |

## Biological samples requested

In order to obtain the biological samples, you are required to agree to return to CLS genotypes and/or lab analytes to enhance the CLS data resource.

If you are seeking access to a finite resource, the CLS DAC will also use the information you have provided to obtain a short report from the CLS laboratory in Bristol to assess whether sample depletion will be an issue.

The CLS DAC may request that you provide a short external peer review document with regards to the sample analysis you propose.

Please indicate what cohort you require biological samples from:

[ ]  NCDS

[ ]  MCS

Please indicate what biological samples you require:

[ ]  DNA samples. Please complete **section 16.1**

[ ]  Other biological samples. Please complete **section 16.2**

### DNA samples

In addition to the information below, you must provide a copy of the protocol(s) to be used for laboratory processing and analysis, including Quality Assurance (QA) / Quality Control (QC) documentation. These may be attached as a separate sheet to the application.

Intended use of DNA samples

[ ]  SNP analysis. Approximate number of SNPs?

|  |
| --- |
|  |

[ ]  Micro-satellite analysis. Approximate number of micro-satellites?

|  |
| --- |
|  |

[ ]  Sequencing. Targeted or whole genome sequencing?

|  |
| --- |
|  |

☐ Structural DNA work (including copy number variation). Details:

|  |
| --- |
|  |

[ ]  Methylation analysis. How will you use existing methylation data in your analysis?

|  |
| --- |
|  |

[ ]  Other. Please describe:

|  |
| --- |
|  |

DNA sample details and management:

|  |  |
| --- | --- |
| Samples required | [ ]  All available DNA samples[ ]  Subset of DNA samples |
| Description of sample subset required, if applicable |  |
| Quantity of DNA required |  (μg) per sample/subject[Please note standard aliquots of cell-line DNA are 1µg at a concentration of 50ng/µl but larger quantities and concentrations are available on request.] |
| Minimum concentration required |  (ng/μl) [Please note standard aliquots of cell-line DNA are 1µg at a concentration of 50ng/µl but larger quantities and concentrations are available on request.] |
| Number of subjects  |  |
| Sample type | [ ]  Non-cell-line DNA[ ]  Larger sample of cell-line DNA |
| Justify the size of the sample requested |  |
| Are you happy to receive cell-line DNA? | ☐ Yes ☐ No |
| Person responsible for analysis  |  |
| Laboratory address |  |
| Requirements for sample receipt |  |
| Have you attached a copy of the protocol(s) to be used for laboratory processing and analysis?  | ☐ Yes ☐ No |
| Do you require an embargo period before other users can access the data?*Note: Embargo period is up to 1 year from the date of issue of data/samples* | ☐ Yes ☐ No |

### Other biological samples

Intended use of samples

Please specify what will be analysed:

|  |
| --- |
|  |

Sample details and management:

|  |  |
| --- | --- |
| Samples required  | ☐ All available samples☐ Subset of samples |
| Description of sample subset required, if applicable |  |
| Sample types required | ☐ Plasma☐ Serum☐ Saliva |
| For plasma, preferred anticoagulant |  |
| Minimum sample quantity per subject |  |
| Justification for the volume/quantity requested |  |
| Do you require lymphoblastoid cell lines? | [ ]  Yes [ ]  No |
| Person responsible for analysis  |  |
| Laboratory address |  |
| Requirements for sample receipt |  |
| Have you attached a copy of the protocol(s) to be used for laboratory processing and analysis?  | ☐ Yes ☐ No |
| Do you require an embargo period before other users can access the data?*Note: Embargo period is up to 1 year from the date of issue of data/samples* | ☐ Yes ☐ No |

## CLS conditions of data and sample use

This section concerns the conditions of use of Data or Biological Materials that form part of the research Project hereafter referred to as 'the Materials'. The term ‘Data’ refers to the Data requested, as well as any new derived variables or data arising from this CLS DAC application. CLS data sharing is governed by the [CLS Data Access Framework](https://cls.ucl.ac.uk/wp-content/uploads/2017/02/CLS_Data_Access_Framework.pdf).

In signing this Agreement, the main data applicant agrees to use the Materials according to the CLS conditions listed below:

1. Use the Materials only for the research Project approved by the CLS DAC.
2. Understand that the Materials requested, and any Data or documentation returned to CLS, may be shared by CLS for wider use via the UK Data Archive or other repositories.
3. Preserve the confidentiality of, and not attempt to identify individuals, households, or organisations. Materials cannot be linked or combined to other information in a way that could re-identify the research participants, even if access to those Materials has been granted.
4. Store the Materials securely in an organisational setting; if Data are placed in a shared directory or on a Local Area Network (LAN), access must only be available via personal authentication
5. Be aware of, and follow, any information security guidelines provided by the Recipient’s Institution.
6. Ensure that the means of access to the Data (such as passwords) are kept secure and not disclosed to anyone.
7. Notify CLS of any non-compliance they are aware of. Any incidents of unauthorised access to, Processing of, or disclosing of the research Materials must be reported immediately.
8. Not use the Materials for commercial purposes.
9. Notify CLS of any errors discovered in the Materials by emailing *clsfeedback@ucl.ac.uk*.
10. Understand that CLS does not transfer any interest in Intellectual Property to the Recipient’s Institution.
11. Understand that the Materials are provided without warranty or liability of any kind.
12. The Materials must not be passed to a third Party (i.e., to anybody that is not included in the list of applicants on this Project, nor is a direct employee of one of these applicants).  This includes sharing of publicly available individual level Data. If collaborators are based in different organisations, they should sign a separate Data Sharing Agreement (DSA).
13. At the request of CLS, return derived variables and related documentation arising from this request to CLS within 12 months of receipt of the requested Data.
14. Ensure that all reporting of Project results and press releases use careful and balanced language in order to avoid misinterpretation or exaggeration of the findings. Live interactions with the media (including social media) should use careful and balanced language.
15. At the end of the Project, Data must be deleted from the system on which it has been stored using a secure erasure programme. The recycle/trash bin must be emptied and immediately followed by the running of a secure erasure programme.
16. Inform CLS of any publications that arise from this Project, which CLS will cite and publish on the CLS website.
17. Be aware that CLS will publish anonymised basic information about this CLS DAC approved Project on the CLS website.
18. Understand that this Agreement is subject to review and without limitation whenever a change in the law, contracts for services with third Parties, other procedures, or other relevant circumstances take place.
19. Agree to comply with the General UCL Terms and Conditions described in Section 6 and with the Data Protection requirements described further in Schedule 1. In case of a conflict with Schedule 1, the provision of sections 5 and 6 shall prevail.
20. Understand that any non-compliance with these terms and conditions will lead to immediate termination of access to the Data and could result in legal action.

**Signed:** Click or tap here to enter text.

**Print name:** Click or tap here to enter text.

**Date:** Click or tap here to enter text.

# CLS data access guidelines

The guidelines for the CLS research data access guidelines are derived from the [CLS Data Access Framework](https://cls.ucl.ac.uk/wp-content/uploads/2020/12/CLS_Data_Access_Framework.pdf), which sets out the principles and procedures by which all access to and use of CLS data is governed.

## Introduction

The Centre for Longitudinal Studies (CLS) is responsible for four national cohort studies: the National Child Development Study (NCDS, or the 1958 Birth Cohort Study), the 1970 British Cohort Study (BCS70), the Millennium Cohort Study (MCS) and Next Steps, which follow the same group of people from childhood and throughout their lives.

CLS research data are publicly available to researchers through a number of data repositories. The vast majority of CLS research data are available from the UK Data Service (UKDS):

* [NCDS](http://discover.ukdataservice.ac.uk/series/?sn=2000032)
* [BCS70](http://discover.ukdataservice.ac.uk/series/?sn=200001)
* [MCS](http://discover.ukdataservice.ac.uk/series/?sn=2000031)
* [Next Steps](http://discover.ukdataservice.ac.uk/series/?sn=2000030)

The access mechanisms for data available via UKDS depend on the level of risk of disclosure and impact of that disclosure as follows:

* Most CLS data can be accessed via a standard licence known as an 'End User Licence' (EUL). EUL applications are authorised directly by the UKDS.
* Research data, which are sensitive and/or potentially disclosive of the identities of individuals, households or organisations can be accessed via the UKDS Special Licence.
* Significantly sensitive or disclosive data can be accessed via the UKDS Secure Access user agreement.

Other CLS research data are publicly available via:

* the SAIL Databank, which holds MCS linked to Welsh administrative data
* the UK LLC, which holds CLS survey data linked to NHS Digital linked health administrative data
* the European Genome-Phenome Archive (EGA), which holds NCDS genetic data

## Data access via the CLS DAC

The CLS Data Access Committee (CLS DAC) welcomes proposals for data enhancements to its four cohort studies. The functioning of the CLS DAC is described in detail in the [CLS DAC Terms of Reference](https://cls.ucl.ac.uk/wp-content/uploads/2020/12/CLS_DAC_Terms_of_Reference.pdf).

The CLS DAC welcomes applications to access the following types of data:

* **Main survey data:** some data collected by the CLS cohort studies have not been deposited at the UKDS due to their sensitive or disclosive nature, or because they have not been prepared sufficiently for deposit.
* **Linked administrative records:** CLS has a programme of administrative record linkages underway. Some existing linked data have not been deposited at the UKDS due to their sensitive or disclosive nature but are available for research purposes under secure access.
* **Paradata:** CLS holds data about the data collection process from some data collection sweeps. These are collected primarily for administrative purposes and are not routinely released for research use.
* **Genetic data:** CLS has a programme of genetic data collection. Given the sensitivity and potential disclosivity of the genetic data once they are combined with survey data, these requests are subject to a separate release arrangement that requires the creation of a bespoke survey dataset identified by a specific ID.
* **Biological samples:** CLS has a resource of biological samples stored at the University of Bristol. Access can be requested for genotyping or generation of new analytes.

Requests to create **new linked data**, including new geographical linkages, are governed by separate CLS guidelines and require the submission of the [CLS record linkage application form](https://cls.ucl.ac.uk/data-access-training/data-enhancements/).

Requests to **collect new CLS data** or to **enhance the legacy data resource**, e.g.by digitisation, coding or transcription, are governed by separate CLS guidelines and require the submission of the [CLS data enhancement application form](https://cls.ucl.ac.uk/data-access-training/data-enhancements/).

## Request process

To apply for CLS research data please follow these steps:

1. First check whether the data are publicly available via the [UKDS](https://ukdataservice.ac.uk/get-data/key-data/cohort-and-longitudinal-studies), [SAIL Databank](https://saildatabank.com/saildata/sail-datasets/) or the [EGA](https://ega-archive.org/dacs/EGAC00001000205). In case of doubt, please contact CLS at *clsfeedback@ucl.ac.uk*. If the data are not publicly available, then you can proceed with the CLS DAC application.
2. Complete the [CLS DAC data access application form](#_CLS_Data_Access_1).
3. Send the application form to *clsfeedback@ucl.ac.uk**.*
4. For biological samples, provide sample management protocols when requested by CLS.

All proposals are discussed at a meeting of the CLS Data Access Committee (CLS DAC), which typically meets every month.

Researchers must allow plenty of time between submitting their application and when they plan to undertake research on the data requested. In cases where significant resources of the CLS data management team are required to fulfil the request, we suggest an application is submitted at least 3 months before the planned research will take place.

The data applicant may be asked to respond to any queries raised by the CLS DAC meeting and to modify their proposal to render it suitable for further consideration.

Following approval, further extensions or modifications of the project can be requested by contacting *clsfeedback@ucl.ac.uk*. Substantive extensions or modifications will need to be described in a new version of the DAC application.

Any third party seeking to use data, samples, or derived variables or genotypes arising from this application must approach the CLS DAC to obtain access permission of their own.

CLS data can be used according to the consents and ethical approvals of the studies and cannot be varied under any circumstance. The applicant’s funders’ and publishers’ requirements must be compatible with these and cannot require anything in conflict with the CLS conditions of data and sample use ([section 17 of the CLS data access application form)](#_CLS_conditions_of).

## Genetic data requests

Genetic data linked to survey/biomedical data (phenotypic data) have an increased disclosure risk and are therefore subject to more stringent access arrangements in order to enable secure analysis at an individual level. This level of security is achieved via the creation of a bespoke phenotypic dataset identified with project specific IDs.

Therefore, researchers wishing to access genetic data combined with phenotypic data will need to:

1. provide additional information on the CLS DAC application form, and
2. submit a list of survey variables names they would need to link to the genetic data

Researchers must request only the necessary data required to conduct the research project described in the application.

## The CLS DAC approval process

The CLS DAC’s decisions will be based on the information provided in the application form and accompanying documentation. The approval criteria are listed in section 8.2 of the [CLS Data Access Framework](https://cls.ucl.ac.uk/wp-content/uploads/2017/02/CLS_Data_Access_Framework.pdf).

As part of the application management, the CLS DAC will evaluate the level of disclosivity and sensitivity of the data requested and will assign the relevant “Tier” according to the [CLS Data Classification Policy](https://cls.ucl.ac.uk/wp-content/uploads/2017/02/CLS_Data_Classification_Policy.pdf). The four data tiers are:

* Tier 1: low level of disclosure and sensitivity. This corresponds to the UKDS End User Licence.
* Tier 2a: medium level of disclosure and/or sensitivity. This corresponds to the UKDS Special Licence
* Tier 2b: high level of disclosure and/or sensitivity. This corresponds to the UKDS Secure Access.
* Tier 3: very level of disclosure and/or sensitivity. These data must be accessed via the UCL Data Safe Haven

The CLS DAC considers appeals from researchers regarding the tier level allocated to any CLS data. For further information on how to proceed, please contact clsfeedback@ucl.ac.uk.

The principles applied in depleting biological samples must include the scientific strength of the proposal and the appropriateness of the methodology proposed. CLS holds responsibility for applying these principles to specific applications for the depletion of finite biological resources.

## Data release

Once access has been granted, research data will be released via a suitable mechanism, free of charge. For Tiers 1, 2a and 2b, this will generally be via the UKDS. However, depending on CLS capacity, interim arrangements will be made occasionally to provide the applicants with the requested data prior to these becoming available at the UKDS.

CLS data releases will be governed by a CLS data sharing agreement (DSA) and subject to data security arrangements equivalent to those from the UKDS.

In circumstances where access is granted to Tier 3 data, this would only be permitted via the highly secure UCL Data Safe Haven.

For genetic applications, both the phenotypic and the genetic datasets will be identified by a newly created ID for each applicant.

All researchers working on CLS data will observe the CLS conditions of data and sample use that can be found at the end of the application form ([section 17](#_CLS_conditions_of) of the CLS data access application form), and will not attempt to identify individual cases, or share data with other unauthorised person(s).