CLS Linked Data Policy

This document includes data that is **PUBLIC** and can be disclosed outside UCL IOE CLS and shall not be duplicated, used, or disclosed in whole or in part for any purpose other than to evaluate and implement procedures defined within this document.

Contents

Scope

This document covers the acquisition, management and dissemination of Linked Data within the Centre for Longitudinal Studies (CLS). Linked Data within CLS are any data linked to CLS cohort data that are not collected or derived from a CLS data collection such as a survey question, physical sample, device from a participant or other person who has been contacted as part of the survey process. This includes both individual and aggregate level data.

Rationale for Data Linkage

CLS seeks to obtain Linked Data for the following purposes:

1. To enhance contact information from administrative records for the purposes of tracing or contacting cohort members;
2. To enhance survey data through linkage of administrative or other relevant records;
3. Linked Data provide a rich source of additional data that potentially:

- cannot be obtained directly from the study participants
- validates responses by participants, or the linked data itself
- reduces the burden on participants for research
- provides information for retention of cohort members for tracing purposes.

Linked Data Programme

CLS has a programme of data linkage that is based on the survey data collection from participants where consent to linkage of health, education, economic and crime records from administrative sources, has been secured. Appendix 5 provides an example of education data linkage. In addition, there is a programme of non-consented linkage of updated personal contact details for tracing purposes, and of mortality data, geographic data and other data for research purposes.

Researchers may also submit a data linkage proposal for data linkages that are not part of the above programme. Applications for new data linkages can be made for any of the four CLS cohort studies through the *CLS Application for Record Linkages*, available on the CLS website. All proposals will be discussed by the CLS Data Access Committee (CLS DAC).

Information Governance

CLS Linked Data are categorised and shared in accordance with the *CLS Data Classification Scheme*, here shown in Appendix 6. This scheme sets outs the classification of data and the appropriate levels of data security and data segmentation that accompany this.

UCL holds NHS Data Security and Protection (DSP) toolkit (previously known as NHS Information Governance Toolkit) accreditation, reference EE133902-SLMS. All CLS staff complete mandatory IG training and all staff processing personal data must have basic disclosure checks. In addition, CLS aims to comply with any specific information governance requirements of Data Providers.

Data Management Environment

CLS holds and manages its potentially identifiable personal data within the UCL Data Safe Haven[1] (DSH). The DSH is a computing environment separated from the rest of the UCL network and has been set up with additional technical measures aimed at increasing security by reducing risks. The DSH is certified to the international information security standard ISO27001:20131 and also complies with the NHS Data Security and Protection Toolkit. The data managed by the following two CLS teams:

1. The Cohort Maintenance team who are responsible for and have access to study participants' contact details and other personal identifiers and also those of their relations (mainly, partners, parents), such as names, addresses, NINOs and NHS numbers
2. The Research Data Management team who have access to the research data i.e. de-identified research data collected during surveys and external data linked through administrative systems or geographical databases.

Each team has access to a logically separated DSH area. This is managed by access control, so that the two teams reside in separate access groups.

Legal Basis for Processing and Sharing Linked Data

a. Our legal basis for processing and sharing linked personal data is GDPR Article 6(1)(d) **-** 'processing is necessary for the performance of a task carried out in the public interest or in the exercise of official authority vested in the controller'. The processing of special categories of personal data, such as health data, follows the additional condition from Article 9(2)(j), paraphrased: 'processing is necessary for scientific research purposes, subject to appropriate safeguards'. For operational (tracing) purposes, an ethics application is submitted by the Records Linkage Team to a Research Ethics Committee and a separate application is submitted to the Confidentiality Advisory Group (CAG) for Section 251.
b. A Data Sharing Agreement is agreed and signed with the external data provider. This must include the scope and content of the data being transferred to accurately link the cohort participant to the administrative records. The agreement also includes a list of those data items which the data holder will supply and the period of retention of personal identifiers supplied from CLS for linkage. CLS applies (jointly with a co-applicant if appropriate) to the data provider for acquisition of data, for those participants that have consented (where applicable).

Data Protection Impact Assessments

CLS has completed a UCL Data Protection Impact Assessment of its records linkage programme. This is periodically reviewed, including for any new data linkages, to ensure that any new risks are identified and mitigated.

Linkage for Tracing Purposes

Where CLS has lost contact with participants who took part in any of its four studies (NCDS, BCS70, Next Steps and MCS) an attempt will be made to trace these cohort members so that they can be re-contacted and invited to take part in the next survey. One of the ways CLS attempts to trace participants is by requesting information from other data providers such as NHS-Digital and the Department for Education (DfE). CLS applies to these data providers to request new addresses for these participants; CLS also regularly applies to NHS-Digital to receive notifications of deaths or moves.

[1] https://www.ucl.ac.uk/isd/services/file-storage-sharing/data-safe-haven-dsh

1. **Data Flow for tracing purposes**

Where the data request is for tracing purposes, CLS requests information for those participants whose contact addresses are found to be outdated in the CLS's database.

- CLS prepares a file containing the CLS member IDs, names, addresses, dates of birth, gender, NHS numbers or school details (depending upon the data provider) and sends it to the data provider.

The file sent to data providers for linkage contains either data on the full cohort, or only a subset of the cohort. This will depend on CLS's needs and the data providers' requirements and working procedures.

a) CLS sends a file containing a subset of the cohort.

- The file contains only cohort members which CLS intends to trace.
- The data provider matches the participants against the information available in their database and prepares a file containing the latest addresses they have for each participant and sends it back to CLS, including the all other additional variables provided.
- Once information is provided to CLS the data provider deletes the file from their database.

b) CLS sends a file containing the full cohort

- The file contains all cohort members who have ever participated in the study except for those cohort members who have requested to withdraw from the study or who have died.
- The data provider flags the full cohort on their database
- Data provider sends CLS latest updated addresses for the cohort members flagged on their database on a regular basis, as agreed in the contract with the provider.

On receipt of the data, CLS will:

- Update their database with new addresses.
- Provide new information agency field workers where necessary.
- Provide new information to third party mailing house of the new address where necessary.
- Please see Appendix 3 **-** *External data request for tracing purposes data flow diagram* for detailed information.

2. **Data flow for Notifications of Deaths or Embarkations**

Where a data request is to receive notifications of deaths and moves, CLS requests information for the full cohort.

- CLS prepares a file containing the CLS member IDs, names, addresses, dates of birth, gender, NHS numbers for the full cohort and sends it to NHS-Digital.
- NHS-Digital flags the full cohort in their database.
- NHS-Digital notifies CLS on a regular basis, as agreed in the contract with NHS-Digital.
- Please see Appendix 4 - *External data requests for notifications of deaths data flow diagram* for further details.

Linkage for Research Purposes

1. **Consent checking**

CLS has sought informed consent for each type of data linkage from the participants of its studies and participants can withdraw this consent at any time. Any proposed data linkage is only conducted on those participants for whom consent has been given. Full details of the consent

collection procedure and the numbers of individuals who have agreed are the basis upon which identifiable information is passed to the Data Provider to enact participant linkage. If a respondent withdraws their consent for data linkage no further linkage will be conducted but the data will not be withdrawn from the UK Data Service (or other similar service). If a respondent requests that their linked data be deleted data will not be deleted immediately but will not be included in new deposits. More details can be found in the CLS Withdrawal and Data Deletion Policy.

2. Participant matching

a. This involves the CLS Cohort Maintenance Team sending in a secure manner pseudonymised uniquely identifiable information to the Data Provider for those participants who have consented. This information enables the Data Provider to identify the participants in the Data Providers' systems. In addition to names, addresses, date of birth, gender, CLS might send information such as National Insurance Number, NHS Number or institutional identifiers such as a school or GP.
b. The identifiable information is sent with a proxy serial ID for the purposes of the individual matching only. This proxy serial ID is generated by the Research Data Management Team and passed to the Cohort Maintenance Team together with an internal identifier shared by both teams. The identifiable information is sent to the Data Provider with ONLY the proxy serial ID. Please see Appendix 2- *Proxy Serial ID generation data flow diagram* for further details.
c. Resolution of data matching queries between the Data Provider and the CLS Cohort Maintenance Team.
d. Agreement on what constitutes a valid match, e.g. name, NHS number, date of birth, etc.

3. Data linkage and extraction

a. Following the agreed criteria for matching individuals, the Data Provider is responsible for linking and extracting the relevant data related to that individual.
b. The content of what needs to be extracted should be agreed in advance of the linkage process between CLS and the Data Provider and specified in the Data Sharing Agreement between the two parties.
c. The Data Provider utilises the proxy serial ID to identify the linked data and removes all of the personal identifiers used for linkage (names, addresses, NINOs and NHS numbers, etc), ensuring that the minimum amount of identifiable information is included in the matched dataset.
d. The Data Provider sends the linked data to the CLS Research Data Management team in a secure manner.
e. Once CLS has checked the linked data, the Data Provider securely deletes the uniquely identifiable information provided to them by CLS. Please see Appendix 1 *CLS External Data Linkage data flow* for detailed information on the data linkage process.

4. Data storage

The CLS Research Data Management team stores the linked data returned from the Data Provider in the UCL Data Safe Haven for de-identification, curation and data documentation.

5. Data enhancement and validation

Subject to the terms of the agreement with the Data Provider, CLS securely stores and manages

access; carries out validation of the data received and combines the linked data with that collected from the surveys or other linked data (where appropriate). In the case of an application from an external researcher, CLS enables their access to the data.

6. **Evaluation of disclosure risk**

   CLS evaluates the linked data in terms of whether the data might contain information that could re-identify an individual (disclosivity) and how damaging re-identification might be to an individual (sensitivity). Data may be modified to reduce the risk and are classified according to the CLS Data Classification Scheme.

7. **Pseudonymisation**

   Versions of the data prepared for research purposes are 'pseudonymised' with 'Research Identifiers' used for the research survey data available from the UK Data Service. These 'Research Identifiers' are shared across different sweeps of the data to enable longitudinal research analysis. In some cases a dataset is prepared with its own unique set of Research Identifiers so that the data cannot be linked to other data from the same study for reasons of data confidentiality.

8. **Data transfers**
   Data transfers are always encrypted in transit and logged. For transfers between CLS and the Data Provider, this may be done using the UCL DSH secure transfer system or the Data Provider's own secure FTP. For transfers of data to UKDA, these are securely uploaded as an encrypted archive via the University of Essex ZendTo portal.

9. **Data dissemination**
   See *Data dissemination and governance* below.

Linkage to Data at Small-Area Level

CLS has linked publicly available data such as geographical information at small-area level to its cohort data on a non-consented basis; for example, air pollution data. There is the potential for further geographical linkage to a wide range of other external data such as Ordnance Survey, housing, environmental, energy, broadband, weather and other publicly and non-publicly available data. Where the data are freely and publicly available they can be linked without the requirement for specific agreements with the data owners, subject to the terms and conditions of use. There is also potential for geographical linkage to data held by private organisations, for example those holding credit rating scores. In this case specific agreements will need to be in place to allow for onward sharing of the linked data at no cost (although there may be a one-off fee to cover the linkage process). There is a programme of work to explore which other data it would be most useful to link to in future and to plan for such linkages. Such linkages are subject to disclosure control reviews before being disseminated.

Data Dissemination and Governance

CLS is committed to ensuring that as much linked data as possible are made available to the research community through the existing mechanisms of data dissemination. This is a key aim of the Economic and Social Research Council (ESRC), which is the main funder of CLS, which has an international reputation for supporting high quality data for the research community, enacted primarily through the UK Data Service (UKDS).

The linked data are pseudonymised, de-identified, curated and documented by the CLS data Management

team for research purposes.The dissemination of these linked data to the research community can only be carried out with the explicit agreement of both CLS and the Data Provider. There are three channels by which linked data can be disseminated:

1. **UK Data Service (UKDS)**
   Our strong preference is that data are disseminated through the UKDS at the University of Essex. This has several advantages including:
   a. The governance framework utilised by the UK Data Service is compliant with best practice in both its security arrangements (ISO27001) and good governance e.g. the Five Safes principle (http://blog.ukdataservice.ac.uk/access-to-sensitive-data-for-research-the-5-safes/):
      i. A single process can be used across different studies by the same Data Provider
      ii. The Data Provider only has to provide the data once and then oversees applications from the research community to utilise that data
      iii. The Data Provider does not have to individually vet each application other than for the purpose of the application (mitigating the need to monitor data security compliance).
      iv. Robust security safeguards are in place. See Appendix 7.

   b. Data are deposited at the UKDS under the relevant UKDS Depositor Licence Agreement, which specifies the terms under which UKDS will make the data available to the data users and, where relevant, that copyright is held jointly with the Data Provider. This also provides a mechanism for the UKDS to allow joint governance of the data which means that both CLS and the Data Provider (where appropriate) will jointly approve applications for use of the data.

2. **Other data service providers**

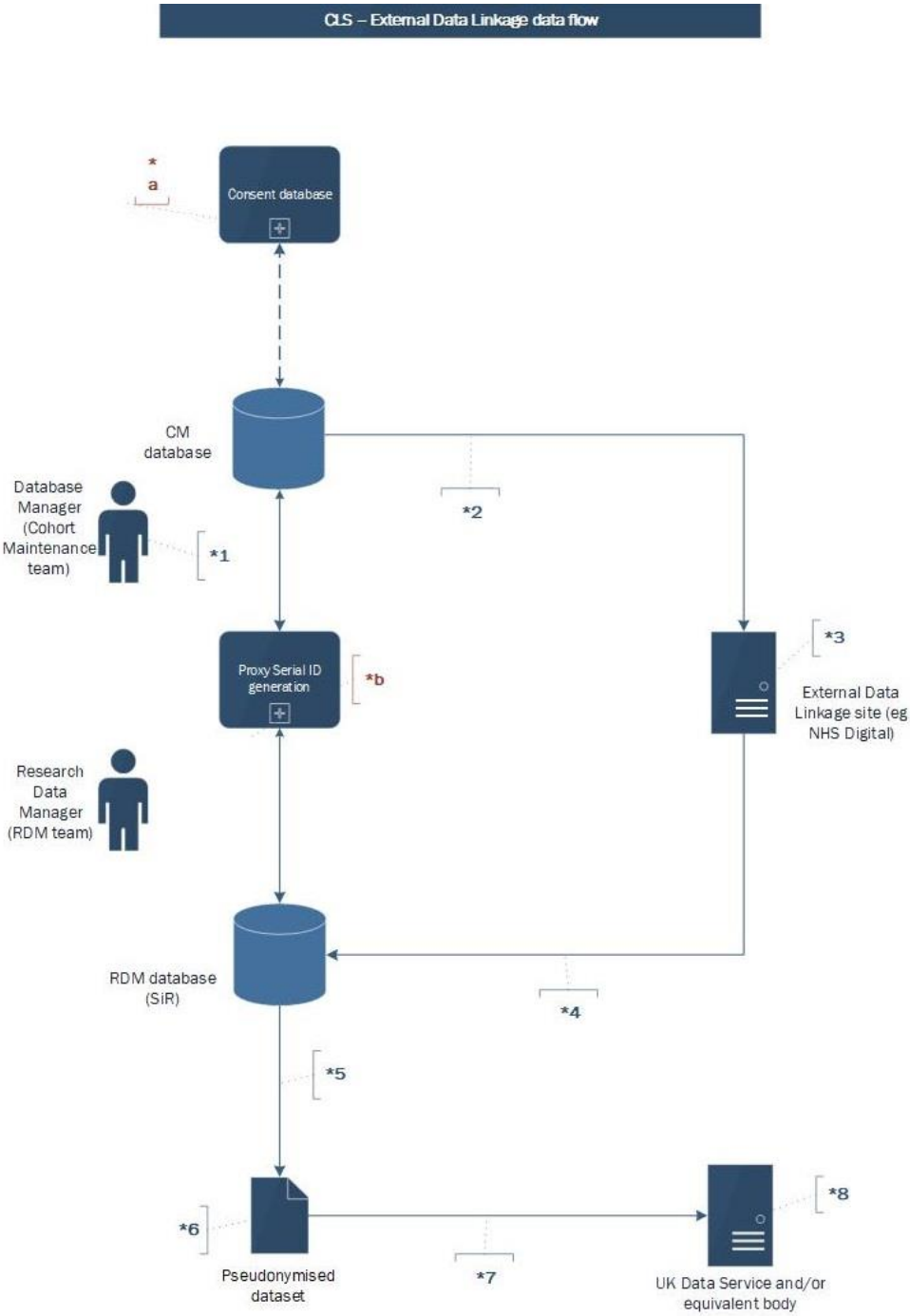
These are other UK organisations that provide data sharing services to the research community, such as the SAIL (Secure Anonymised Information Linkage) Databank, the Office for National Statistics Secure Research Service (ONS SRS), Dementias Platform UK (DPUK), etc. These organisations will also have their own established policies and protocols. CLS and the Data Provider (where appropriate) will jointly approve applications for use of the data.

3. **CLS**

The CLS Data Access Committee (DAC) is responsible for approval of applications for use of the data not currently available through the UKDS. Applicants need to follow the recommended application process which is set out in the CLS data access framework. The CLS DAC determines the most suitable onward sharing arrangement, which will usually be through the UCL Data Safe Haven. These applications may include methodological research projects to access linked data prior to their deposit through the UKDS in which case researchers are required to register their research with the UCL Data Protection Office and may need to complete a DPIA.

The dissemination and access arrangements for data are described in detail in the CLS Data Access Framework

Appendix 1 - CLS External Data Linkage Flow diagram



Process Notes

***1** - The Database Manager in the Cohort Maintenance (CM) team sends the CLS IDs of the cohort members to be included in the data linkage to the RDM team. These are limited to those cohort members for whom CLS has consent. The RDM team generates Proxy Serial IDs to send to the External Data Linkage site in place of the CLS IDs.
***2** – The CM team compiles the personal data of the cohort members (eg. name, address, DOB, etc.) plus the RDM team-generated Proxy Serial IDs.
***3** – Data transfers are encrypted in transit and logged. This is done using the UCL Data Safe Haven secure transfer system or the Data Provider's own secure FTP.
***4** – Matched data file returned securely with relevant requested data (eg. NHS Digital variables) plus Proxy Serial IDs.
***5** – Assigned staff solely within the RDM team ingest the returned data into the RDM database and prepare the data for deposit with the UK Data Service (or other similar service).
***6** – The pseudo-anonymised dataset for deposit contains Research IDs which are different from the CLS IDs and the Proxy Serial IDs.
***7** - Transfer of data from CLS to the UKDS has to comply with their security policies.
***8** – The UKDS classifies the data by its level of sensitivity : either End User Licence for low risk, Special Licence or Secure Lab, the highest level of security.

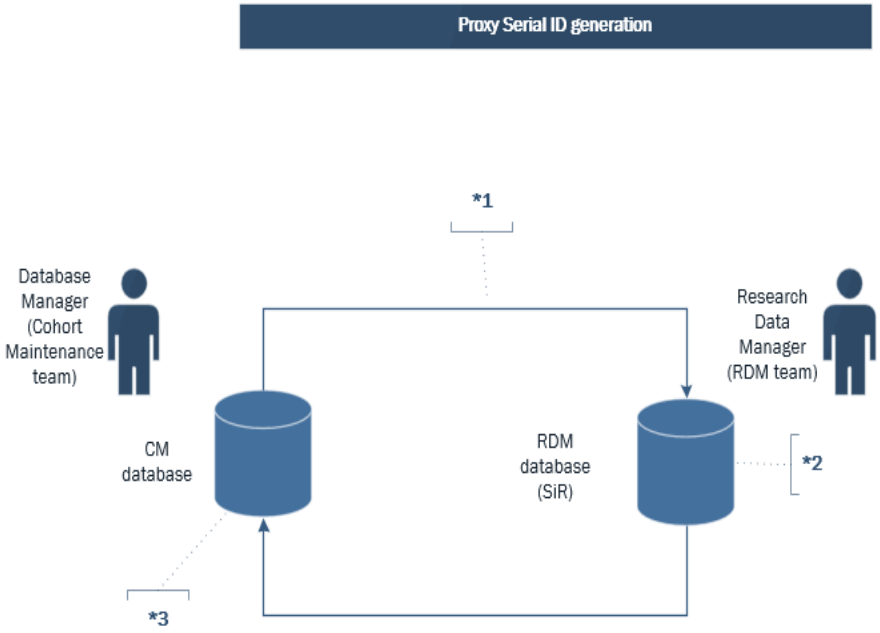
Misc info

***a** – The consent database records which cohort members have consented to administrative data linkages and is kept up-to-date to reflect consent withdrawals.
***b** See Proxy Serial ID generation diagram for details.

Legal basis

CLS's legal basis for processing and sharing linked personal data is for a public task under GDPR article 6(1)(e). CLS also processes special categories of personal data for research under GDPR article 9(2)(j).

Appendix 2- Proxy Serial ID generation data flow diagram



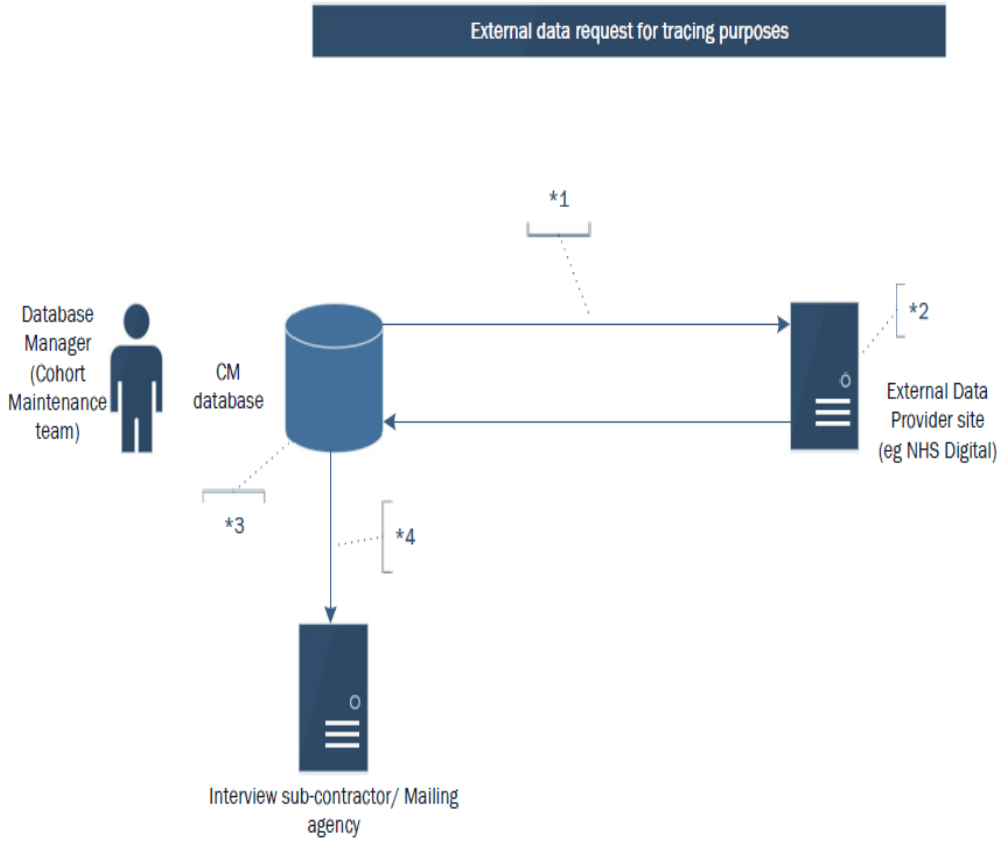
Business Rule

There is an information barrier in place between the two areas of CLS.
The Cohort Maintenance team will only have access to the cohort members personal identifiers (e.g name, address, DoB, gender, etc).
The Research Data Management team will only have access to the research data.

Process Rules

***1** – The Database Manager in the Cohort Maintenance (CM) team will send the '*CLS ID*' to the Database Manager in the Research Data Management (RDM) team.
***2** – The Database Manager in the RDM team will generate a '*Proxy Serial ID*' for each of the '*CLS ID*' and send both IDs back to the CM team.
***3** –The Database Manager in the CM Team compiles the personal data of the cohort (forename, middle name, surname, DoB, gender, latest oldest NHS number, latest new NHS number, address,) plus the '*Proxy Serial ID*' and sends the data to the data provider for linkage.

Appendix 3- External Data Request for tracing purposes data flow diagram



Business Rule

The External Data Provider site contains a secure FTP portal area to drop off data, requiring a login/ user ID and password to gain access.

Process Rules

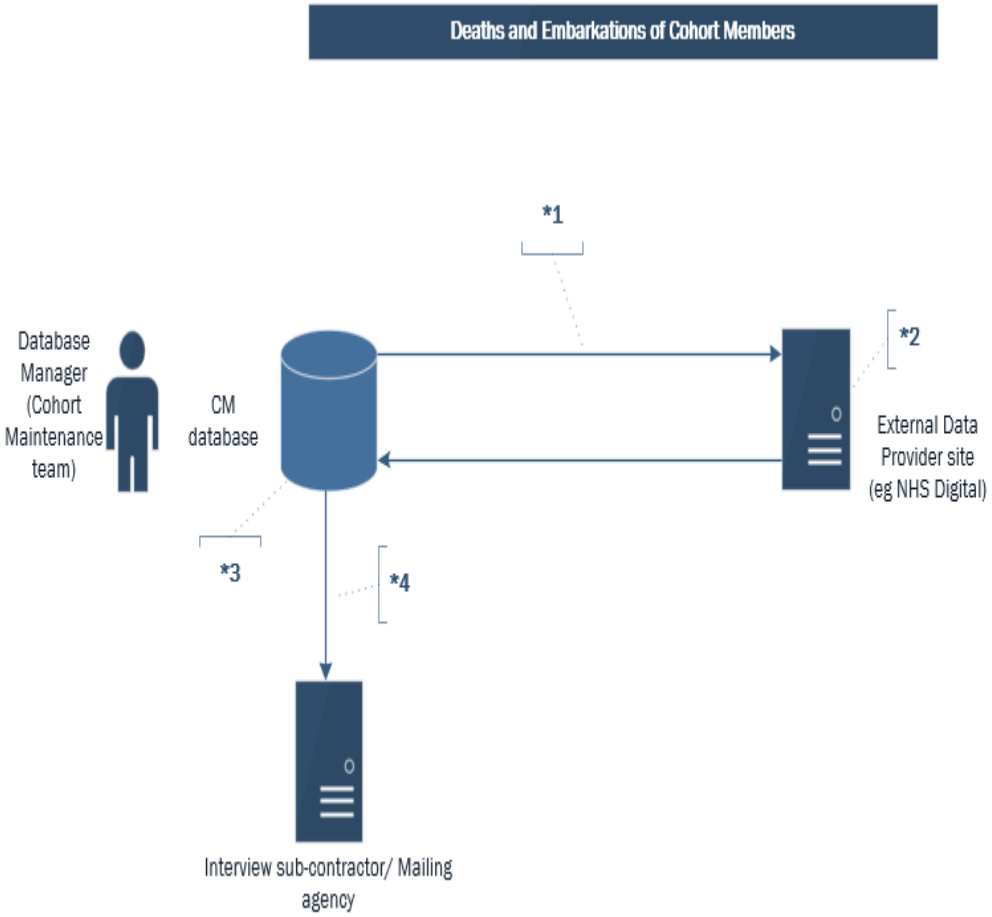
*1 – The Database Manager in the Cohort Maintenance (CM) Team compiles the personal data of the cohort (forename, middle, surname, gender, DoB, NHS number, address, postcode, country and '*CLS ID*') for those cohort members who need to be traced.

*2 – The external data provider finds the cohort in their database and returns to CLS the available data for the cohort member (forename, middle name, surname, DoB, gender, latest oldest NHS number, latest new NHS number, address, postcode and '*CLS ID*'). Data file is returned via secure FTP portal.

*3 – The Database Manager in the CM team updates the CM database where the External Data Provider provided a new address.

*4 - The Database Manager in the CM team updates new contact details with the sub-contractor agency and mailing agency where needed.

Appendix 4- External Data Request for Notifications of Deaths and Embarkations data flow diagram



Process Notes

***1** - The Database Manager in the Cohort Maintenance Team compiles the personal data (Forename, Middle name, Surname, Gender, DoB, NHS number, Address, Postcode, Country and 'CLS ID') for the full cohort and sends it to the external data provider for notification of deaths and embarkations.

***2** - The external data provider flags the cohort in their database and periodically notifies CLS of any deaths or moves.

***3** - The Database Manager in the CM team updates the relevant record in the CM database.

***4** - Database Manager informs relevant external agencies using the CM data (eg. Interview sub-contractors, mailing agencies, etc.)

Appendix 5 - Data Linkage Use Case – National Pupil Database

In the 2008 survey sweep of the Millennium Cohort Study (MCS), CLS asked the parents / carers of the cohort members to consent to linkage of their educational records (the National Pupil Database or NPD). This yielded a 94% consent rate.

CLS approached the Department of Education (DfE i.e. the Data Provider) to link those who consented and were interviewed in England or attended English schools to the NPD to obtain information on schooling history and attainment.

CLS passed the DfE information on name, address, date of birth, sex and schools attended. This yielded an initial 88% linkage rate, which was later increased to 97% by a further linkage exercise.

DfE has a well-established mechanism for sharing the NPD with the research community. This relied upon an application for a specific research project, the provision of the data under conditions which involved individual vetting of both the scientific content of the application and the data security arrangements of the researcher(s). Whilst the NPD is anonymised, there is concern by the DfE that there still a risk of disclosure once the school is known, especially in smaller schools. Any administrative dataset of this nature carries an inherent risk that needs to be assessed as part of the application process. Part of the data sharing agreement between DfE and any researcher with whom the NPD is shared is responsibility for disclosure control especially for research outputs.

CLS proposed an arrangement for the sharing of NPD data linked to the MCS, that the linked data be made available under the UK Data Service's Secure Access arrangements, now known as the Secure Lab, which had just been made available. The Secure Lab is a secure research environment, modelled on the ONS Virtual Microdata Laboratory, now known as the Secure Research Service, whereby researchers are able to conduct research, do analysis and write publications in a virtual environment. Access to the virtual environment also included mandatory training. Users would also be required to fill out a Secure Access Agreement. Outputs from the research are vetted by the UK Data Service's staff to check for disclosure. The licence to use the data lasts for 2 years, and may be extended for a further 6 months.

Selected variables from the NPD were agreed between CLS and the DfE as being suitable for deposit, and it was also agreed, that as the children were still in school both the Local Education Authority and the School should be anonymised. This allowed more data variables to be released, whilst allowing researchers to control for school and local authority effects from their analysis.

Data from the NPD would be linked to the standard identifier used on the other MCS datasets available under the UK Data Service and applications would be invited for research projects.

These applications would state both the rationale for the request and the MCS data to be linked to the NPD.

Once an application is received, the UK Data Service ensure that the researchers have complied with the access conditions and that they have undertaken the mandatory training. The application is then forwarded to both CLS and the DfE for comment / approval. If both parties are in agreement, then the data is released to the applicant in the secure virtual environment.

There is provision in the application process for the Data Provider or CLS to impose conditions such as variables or datasets that cannot be linked e.g. low level geography, and to assert the right to check publications prior to release.

Since 2012, there have been over 100 successful applications for NPD data linked to MCS. This compares

with approximately 700 NPD datasets released by DfE in the same time period.

Appendix 6 - Data Classification Scheme

Centre for Longitudinal Studies data is made available to researchers through a number of access mechanisms. The majority of CLS data is available from the **UK Data Service** at:

NCDS: https://beta.ukdataservice.ac.uk/datacatalogue/series/series?id=2000032

BCS70: https://beta.ukdataservice.ac.uk/datacatalogue/series/series?id=200001

MCS: https://beta.ukdataservice.ac.uk/datacatalogue/series/series?id=2000031

Next Steps: https://beta.ukdataservice.ac.uk/datacatalogue/series/series?id=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). Applications via EUL are authorised directly by the UK Data Service.
- Access to more detailed data, which are potentially disclosive of the identities of individuals, households or organisations is provided via a Special Licence. Applications for data under Special License is administered by the UK Data Service.
- Access to more sensitive or disclosive data is provided via Secure Access which is also administered by the UK Data Service.

Access to biological samples, and some sensitive data derived from biological samples are subject to separate access arrangements. Access to the majority of genotypes generated from NCDS participants is governed by the WTCCC DAC. Access to genotypes linked to other variables (phenotypes), applications for access to DNA, and for new uses of biological samples is via the METADAC (Managing Ethico-social, Technical issues and Administration Data Access Committee).

CLS Data Classification Scheme

This table illustrates the data classification based on their level of disclosure risk and how the data might be publishable under a particular licence agreement with the UK Data Service at the University of Essex (http://ukdataservice.ac.uk/)

Classification	Description	UK Data Service Data Licence
PUBLIC	Publicly available datasets e.g. Edubase, list of schools http://www.education.gov.uk/edubase/search.xhtml	n/a
RESTRICTED - TIER 1	De-identified survey data e.g. http://dx.doi.org/10.5255/UKDA-SN-5565-2	UK Data Service End User Licence
RESTRICTED - TIER 2a	Data with a medium level of potential disclosure risk or sensitivity, for example counties 1986-2012, e.g. http://dx.doi.org/10.5255/UKDA-SN-5537-1	UK Data Service Special Licence
RESTRICTED - TIER 2	Data with a high level of potential disclosure risk or sensitivity including: Geography: Output Areas OA, LSOA, MSOA, Local authority, Local Education Authority e.g. https://discover.ukdataservice.ac.uk/catalogue/?sn=7763	UK Data Service Secure Access
RESTRICTED - TIER 3	Data with a very high level of potential disclosure. Any information which would allow identification of less than 5% of a population of the data item e.g. postcodes, date of birth, GP Identifier used for linkage and lookup to other data	n/a – as never deposited via UK Data Service
CONFIDENTIAL	Individually identifying information e.g. names, address, email, NHS Number, National Insurance Number (NINO), Unique Pupil Number (UPN), Unique Learner Number (ULN), etc.	n/a – as never deposited via UK Data Service
PRIVATE	Not used. This is used primarily for what can be termed as CLS internal focussed documentation for which there is no benefit or requirement to make it publicly available	n/a

Appendix 7 – UK Data Service Safeguards

Robust safeguards must be in place before limited data can be made available to others. The measures that the UK Data Service ('the Service') put in place for Secure Access data in its collection are outlined below against the ICO's recommended safeguards. Secure Access data are de-identified but include a level of detail and / or a degree of sensitivity that warrants limited access safeguards.

Limited access safeguard	**Service staff**	**External researchers**
Purpose limitation, i.e. the data can only be used by the recipient for an agreed purpose or set of purposes	Individual named staff have access to the data for the purposes of processing them to make them available to researchers, as covered by legal agreements with the data owners. All Service staff sign a non-disclosure agreement.	Researchers agree in a User Agreement that access to the Secure Access data is provided for the statistical and research purpose outlined in their Approved/Accredited Researcher application form and the data cannot be used for any other purpose (unless a new application is made and approved for that purpose).
Training of recipients' staff with access to data, especially on security and data minimisation principles	All Service staff are required to attend information security training sessions and to complete a non-disclosure agreement. Staff handling disclosive, confidential or sensitive data must also attend a Secure Access training session and are trained to follow strict data handling procedures to ensure that data are kept safe.	All researchers requiring access to Secure Access data are required to attend a face-to-face Secure Access training session that covers: an introduction to the UK Data Service and other similar services; data security and personal responsibility, including legal background, security model, breaches and penalties; Statistical Disclosure Control – how to make statistical outputs safe and what principles are used, including hands-on exercise; using the Secure Lab – how to use it, how to obtain outputs and how to use the interface safely.
Personnel background checks for those getting access to data	Staff handling Secure Access data undergo criminal record checks for unspent convictions, in addition to the checks made by the University of Essex during the staff recruitment process.	Checks are made to ensure that researchers applying for access are registered with the Service (if they are at a University they will have registered using their institutional login and undergone background checks as part of their recruitment) and we have their contact information. We use the same model/collect the same information as for an ONS Approved Researcher i.e. researchers need to provide evidence in their application that they are a 'fit and proper' person.
Controls over the ability to bring other data into the	Secure Access data are stored and processed on a dedicated access-controlled Secure Access Processing	Data are accessed remotely by registered, approved and trained users through a secure virtual private network

environment, allowing the risk of re-identification by linkage or association to be managed	server accessible only to those staff who need access.	where researchers have access to just the data they have been approved to access. Any data entering the environment are controlled by the Service and any researcher wishing to bring in external data must apply to do so – these data are checked by Service staff and permission sought from the data owners.
Limitation of the use of the data to a particular project or projects	Dedicated staff have access to Secure Access datasets only in order to be able to process them for the purposes of making them available to researchers and to be able to support users.	To access Secure Access data researchers must specify the specific research purpose the data are required for in their application form which is sent to the data owner for approval. If data are required for a new purpose, a new application must be sent. Access to the data is time-limited but can be extended upon request / permission from the data owner.
Restriction on the disclosure of the data	All staff sign a non-disclosure agreement.	Researchers agree in the User Agreement not to disclose nor compromise any of the individual records obtained or produced from the data nor to reproduce outside of the Secure Access system any original dataset or copies or subsets of the data. Any outputs that researchers require for the purposes of publication are only released to the researcher after being checked by Service staff for Statistical Disclosure Control.
Prohibition on any attempt at re-identification and measures for the destruction of any accidentally re-identified personal data	During data processing of Secure Access data staff check for direct identifiers accidentally included in the data. In the unlikely event of any being found, the data owner is informed and they are removed from the data.	All users that register with the Service agree not to use the data to attempt to identify any individual, household or organisation in the data. Secure Access users also agree to ensure that no attempts are made to link the data to any other files in order to relate the particulars to any identifiable individual person, business or organisation unless such data linkage exercise has been explicitly approved at the time of the application for an Approved/Accredited Researcher status as part of their proposed research project, or approved subsequently as part of a special request to the data owners or their delegated decision making body. It is not possible for data to be accidentally re-identified as only de-identified are sent to the Service

		to make available under Secure Access arrangements.
Arrangements for technical and organisational security, eg staff confidentiality agreements	All staff sign a non-disclosure agreement and follow strict procedures surrounding data security, including marking or designating data into specific categories for which different procedures apply. The UK Data Archive, where the data are stored, is certified against ISO 27001.	Data are accessed remotely by registered, approved (by the data owner or his/her nominee) and trained users through a secure virtual private network (i) via the researcher's own institutional desktop PC or (ii) via the Safe Centre at the UK Data Archive (or a safe room with equivalent security as determined by the Service). Researchers also sign a User Agreement which is countersigned by their institution's legal office.
Encryption and key management to restrict access to data	All Secure Access data remain at all times on an encrypted server at the UK Data Archive. Secure Access workstations must have appropriate security safeguards, including: • authentication by Active Directory (AD) username and password issued by the University of Essex • access to the Secure Access Processing server requires additional authentication via an AD username and secure password • encrypted volumes for processing and storing working requiring a further additional secure passphrase for access • a five minutes session lockout which requires password credentials to be re-entered	Secure Access data remain at all times on an encrypted server at the UK Data Archive. For those data available for remote access, once authorised to do so, researchers are authenticated using windows Active Directory using a username and password that we issue to them and for which they must change the password on first logon (using a 'strong' password) and then every three months. Access to the system is also restricted to static, routable institutional IP addresses (which we confirm with RIPE). Users access their Secure Access account using Citrix technology which fully encrypts the data transmitted between the researcher's computer and the host network, passes all traffic through a Secure Access Gateway and imposes severe restrictions on what can be done once logged in i.e. no internet access, downloading or email.
Limiting the copying of, or the number of copies of the data	Staff work on a single copy of the data for the purposes of processing them. The original data, if sent on hard media, are securely destroyed or returned.	No copies of data are held locally as all access is remote.
Arrangements for the destruction or return of the data on completion of the project	Does not usually apply as the data are held for the purposes of long term preservation and access. If a data owner terminates the agreement that allows the Service to hold the data the data are destroyed should the data owner request this.	Users may create copies of the data in their Secure Access account but these are securely deleted by the Service after a specific amount of time after the project has expired.
Penalties, such as contractual ones that can be imposed on	Under the terms of the Non-Disclosure Agreement that all staff sign, staff are made aware that any	The User Agreement that researchers sign includes information about non-compliances and penalties and links to

the recipients if they breach the conditions placed on them	breach of the Agreement may lead to disciplinary action being taken and may also lead to removal of all access to the Archive and/or its systems. Also, any breach of the Agreement, illegal or unlawful, which results in loss or damage to the resources of the Archive (either their own, or held on behalf of others) may be referred for legal action under relevant UK legislation.	the Service Licence Compliance Policy. There are a number of penalties which can be applied depending upon the severity of the non-compliance ranging from a warning, a requirement to re-train, a temporary or permanent ban on access to the Service or any ESRC data service (individual or institutional), withdrawal of ESRC funding (individual or institutional), legal action. The Service has in place a set of procedures to follow in the event of a potential or actual non-compliance.