Data Sharing Policy

March 2017
# CCP Data Sharing Policy

<table>
<thead>
<tr>
<th>Version</th>
<th>Release Date</th>
<th>Reason for Change</th>
<th>Updated By</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.0</td>
<td>22/Mar/2017</td>
<td>First Version</td>
<td>Richard Ostler</td>
</tr>
</tbody>
</table>

**Approved by**

<table>
<thead>
<tr>
<th>Name</th>
<th>Position</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jack Cuzick</td>
<td>Director, Wolfson Institute of Preventive Medicine and Head, Centre for Cancer Prevention</td>
</tr>
</tbody>
</table>

**Signature**

**Date** 22 March 2017
# CCP Data Sharing Policy

## Table of Contents

Data Sharing Policy Statement .................................................................................................................. 4

1. Objective .................................................................................................................................................. 4

2. Introduction ............................................................................................................................................ 4

2.1. Need for a Data Sharing Policy ......................................................................................................... 5

2.1.1. Maximises the research benefit of data .......................................................................................... 5

2.1.2. Ensures data are shared appropriately .......................................................................................... 5

2.1.3. Assists data management best practice .......................................................................................... 5

2.1.4. Required by funding bodies .......................................................................................................... 5

2.2. Compliance ......................................................................................................................................... 5

2.3. Who is affected by this policy and what are their responsibilities .................................................... 5

2.4. Where this policy applies .................................................................................................................. 5

2.5. Dataset Asset Registry Toolkit ......................................................................................................... 5

2.6. Classifying Datasets ............................................................................................................................ 5

2.6.1. Active Dataset ............................................................................................................................... 6

2.6.2. External Dataset ............................................................................................................................ 6

2.6.3. Closed Dataset ............................................................................................................................. 6

2.6.4. Limited Value Dataset ................................................................................................................ 6

2.6.5. Effectively Lost Datasets .............................................................................................................. 6

2.7. Documenting Lost and Limited Value Datasets ................................................................................ 6

3. Limitations to Data Sharing .................................................................................................................... 6

3.1. Dataset status ...................................................................................................................................... 6

3.1.1. Datasets for ongoing studies ......................................................................................................... 7

3.1.2. Insufficient institutional knowledge .............................................................................................. 7

3.1.3. Insufficient documentation .......................................................................................................... 7

3.1.4. Spaghetti Data .............................................................................................................................. 7

3.1.5. Recoverability ............................................................................................................................... 7

3.2. Ownership ......................................................................................................................................... 7

3.3. Insufficient Resources ....................................................................................................................... 7

3.4. Value reduction .................................................................................................................................. 7

3.5. Uncontrolled risk of patient re-identification ................................................................................... 8

4. The Data Sharing Plan ............................................................................................................................ 8

4.1. Data Sharing Plans for Funding Bodies .............................................................................................. 8

4.1.1. Digital Curation Centre DMPOnline tool ....................................................................................... 8

4.2. Mention in the Study Protocol .......................................................................................................... 8

4.3. Features of the data sharing plan ...................................................................................................... 8

4.3.1. Dataset Description ....................................................................................................................... 9

4.3.2. Data Security ............................................................................................................................... 9

4.3.3. Data Access Governance ........................................................................................................... 9

4.3.4. Timescales for release or Period of exclusive use .................................................................... 9
4.3.5. Discoverability and the mechanism for sharing ................................ 10
4.3.6. Limitations on data sharing ......................................................... 10
4.3.7. Data Sharing Pack ................................................................. 10
5. Data Access Governance ........................................................................ 10
  5.1. Controlled Access ................................................................. 10
  5.2. Discovering a Dataset ............................................................. 10
  5.3. Dataset request process ................................................................ 11
    5.3.1. Submitting an Access Request ................................................ 11
    5.3.2. Reviewing an Access Request ................................................ 11
  5.4. Maintain a Data Access Request Log ........................................... 12
  5.5. Data Use Agreement ................................................................... 13
6. Preparation of the Data Pack .................................................................. 13
  6.2. Custom Data Packs ...................................................................... 13
    6.2.1. Adequately resourced ............................................................ 13
    6.2.2. Data Pack Documentation ...................................................... 13
    6.2.3. Dataset Validation .................................................................. 14
7. Data Release Process .............................................................................. 14
References ..................................................................................................... 14
Appendix 1. Glossary .................................................................................... 15

Abbreviations
  CRF  Case Report Form
  BCC  Barts Cancer Centre
  CCP  Centre for Cancer Prevention
  CRUK  Cancer Research UK
  DAP  Data Access Panel
  DCC  Digital Curation Centre
  ISMT  BCC Information Security Management Team
  MRC  Medical Research Council
  NIHR  National Institute of Health Research
  RCUK  Research Councils UK
  UKCRC  UK Clinical Research Collaboration
Data Sharing Policy Statement
The purpose of this policy is to provide appropriate guidance for CCP staff in order to share CCP datasets appropriately in line with regulatory requirements to protect patient confidentiality and funding organisations’ requirements to share data.

The scope of the policy is to define how CCP will share research datasets in respect of:

- Conforming to the requirements of the Data Protection Act and protecting patient confidentiality.
- Conforming to the expectations of funders to share data.
- Ensuring datasets are appropriately documented.
- Respecting the privilege of the research team to conduct and publish from first analyses of the data and to protect researchers' intellectual property.
- Mitigating shared datasets being used to misrepresent CCP research.
- Providing procedures for external persons to request datasets and for CCP to review dataset requests.
- Providing procedures for CCP staff to prepare datasets for sharing.

1. Objective
The objective of Data Sharing is to maximise the value of a dataset by promoting further opportunities for research.

2. Introduction
Sharing medical research data is recognised as providing opportunities to advance understanding and benefit patient care. For example, it allows new research questions to be asked using existing data; promotes collaboration between research teams; allows for the independent verification of results; and provides an added incentive to ensure data are cleaned and effectively curated.

Queen Mary, University of London (QMUL) is the host institution for CCP and follows the Research Councils UK policy on data access. This policy commits QMUL to the general principle of open access to research data, subject to any funder, legal or ethical constraints.

In the case of medical research, the data collected relate to individual patients participating in a research study. Therefore consideration must be given to:

- Participants’ having provided informed consent to have their data shared.
- The risk of participant re-identification from shared anonymised datasets.
- The risk of data being mis-used or mis-represented.
- The ownership of data and the data lineage from which a dataset may be created.

QMUL recognises the particular concerns around medical research data and therefore defers policy to the relevant funders’ data sharing policy. If the funder does not have a policy, QMUL advises the MRC policy on data sharing should be followed.

The UKCRC recommends a controlled access framework to data sharing as it:

“facilitates the responsible sharing of ‘richer data’ for research purposes within the confines of a system that aims to protect patient privacy.”

This guidance has been endorsed by key funders including Cancer Research UK, MRC, and Wellcome Trust. The NIHR is supportive of this guidance.

This document is based primarily on MRC and Cancer Research UK data sharing policies and the UK CRC document, Good Practice Principles for Sharing Individual Participant Data from Publicly Funded Clinical Trials.
In addition to reading this document, and in line with QMUL policy, all CCP staff involved with research studies are encouraged to read and understand the MRC data sharing policy.

2.1. Need for a Data Sharing Policy
Planning for data sharing should be an integral part for any new study and requests for previously created datasets should always be assessed for feasibility of sharing when a request for access is made.

This policy is to ensure CCP has in place procedures to enable the appropriate sharing of CCP data in a research culture which increasingly encourages access to publicly funded research data.

2.1.1. Maximises the research benefit of data
CCP has an extensive collection of datasets built up over many years of research.

This policy is to ensure CCP has in place procedures to enable the effective re-use of research data by others, maximising the potential value of important research data to society and for patient care.

2.1.2. Ensures data are shared appropriately
This policy advocates controlled access for data sharing. Controlled access means data are shared only with suitable persons and shared data protect patient privacy.

2.1.3. Assists data management best practice
A data sharing culture can help embed best practices within CCP by ensuring datasets are appropriately documented and effectively curated. This directly benefits CCP by ensuring dataset knowledge is retained at the institution level and data remain accessible throughout their retention period.

2.1.4. Required by funding bodies
Funding Bodies, including CR-UK, MRC and the Wellcome Trust, increasingly require grant holders to submit data sharing plans both for a funding bid and after funding has been secured.

2.2. Compliance
The sharing of medical research data is governed by legislation, notably the Data Protection Act 1998.

2.3. Who is affected by this policy and what are their responsibilities
The Information Security Management Team is responsible for ensuring staff have appropriate operating procedures and tools for adhering to the policy.

All principal investigators are directly responsible for implementing the policy within their study teams.

It is the responsibility of each employee with a responsibility for reviewing access requests and/or preparing a dataset for sharing to adhere to the Data Sharing Policy.

2.4. Where this policy applies
This policy applies to all datasets for which CCP is the data owner or data controller, or has agreement from the data controller to share data.

2.5. Dataset Asset Registry Toolkit
CCP Datasets should be documented using the Dataset Asset Registry Toolkit[5]. This is an online resource developed by the CCP ISMT for documenting CCP datasets and dataset access requests.

2.6. Classifying Datasets
To assist with the discussion of data sharing activities and for determining whether or not a dataset is appropriate for sharing it is useful to follow a dataset classification. The following classification takes
account of a dataset’s lifecycle, external influencers and limitations. For example, is the dataset being added to with fresh data from a clinical trial? Is the research team currently analysing the dataset ahead of publishing results? Is the data stored in an old file format?

2.6.1. Active Dataset
Active datasets are those which are collecting data, data are being added to, and/or the research team has yet to publish results. CCP is the data controller for these datasets.

2.6.2. External Dataset
External datasets are those where CCP is a Data Processor acting on behalf of a third party and is not the Data Controller.

2.6.3. Closed Dataset
Closed datasets are those where no data is being collected and the original research team have published results. The metadata and documentation for the dataset are good and specific, prepared dataset for sharing may also have been produced. CCP is the data controller for these datasets.

2.6.4. Limited Value Dataset
Limited value datasets are those where the dataset is known but have some known limitation which reduces either their usefulness for data sharing or the completeness of the data which is available to be shared, for example, if the dataset has been insufficiently documented.

Attempts should always be made to maximise the usefulness of the data if possible and to improve dataset documentation.

2.6.5. Effectively Lost Datasets
These are datasets which meet one of the following criteria:
- Cannot be found.
- Cannot be recovered.
- Cannot be reliably identified, for example from spaghetti data.
- Are too poorly documented or understood to meet the limited value dataset classification.

Attempts should always be made to recover lost datasets.

2.7. Documenting Lost and Limited Value Datasets
When a Dataset has been classified as Lost or Limited Value the attempts made to recover or otherwise improve the dataset and its documentation should be documented using the DART application.

3. Limitations to Data Sharing
Data sharing is a desirable end point for the datasets generated by CCP; however, achieving that end point may not always be appropriate or achievable.

This policy recognises that best practices for data management and curation have not always been applied, especially in the case of legacy datasets when a less rigorous data management culture may have prevailed. This may limit the opportunity for such datasets to be shared.

This policy recognises there are limitations to being able share all datasets.

3.1. Dataset status
The current status of a dataset may constrain its data sharing activities.
3.1.1. Datasets for ongoing studies
Data from ongoing studies will not be released by CCP unless otherwise stated in that study’s data sharing plan.

3.1.2. Insufficient institutional knowledge
CCP may have lost the institutional knowledge necessary to share a dataset, for example, if the original team who generated the data have left and inadequate handover documentation was been provided.

A reasonable attempt should always be made to recover the dataset knowledge.

3.1.3. Insufficient documentation
If the documentation and metadata necessary to describe a dataset is inadequate then it may not be possible to share the data.

A reasonable attempt should always be made to improve the dataset documentation.

3.1.4. Spaghetti Data
Spaghetti data occurs when staff create multiple copies of data files but do not document which are the source datasets, which transformation process were used and which are the generated secondary or derived datasets or their purpose. In this case, it may be very difficult to identify which dataset files are appropriate for data sharing.

3.1.5. Recoverability
Legacy datasets may have been stored onto storage media or using file formats which either cannot be read or for which CCP does not have the software to read. Alternatively datasets may have been saved in a password protected format and the original password has been lost.

A reasonable attempt should be made to recover the data, including use of 3rd party data recovery specialists or tools.

3.2. Ownership
For datasets where CCP is not the data controller, but is acting solely as a data processor, CCPs ability to share data or not will be governed by the data sharing agreement in place with the organisation which provided the data.

Unless otherwise stated, line level data must always be considered as being owned by the organisation providing the data, even if transformed to remove certain fields or merged with other datasets.

Aggregated datasets, derived from externally provided datasets, can normally be considered as owned by CCP.

3.3. Insufficient Resources
Preparing datasets for data sharing requires staff and time resources. If the resources to generate a dataset are unavailable this may limit CCPs ability to share a dataset.

Limited resources must not be invoked simply in order to refuse access to a dataset. A reasonable attempt should be made to cost the resources required to prepare a dataset. It should be discussed with the data requester whether or not they are prepared to meet the financial cost of preparing a dataset and the outcome documented using DART.

3.4. Value reduction
Anonymisation and de-identification of individual patient data may limit the data available and reduce the scope for analysis. CCP should endeavour to use appropriate statistical techniques, for example
aggregation, to ensure the data retains as much utility as possible, however, in some cases these processes may degrade the dataset to such an extent that it no longer has any research value.

3.5. Uncontrolled risk of patient re-identification

Data sharing should not be considered if CCP is unable to effectively anonymise a dataset or there remains a risk that individual participants could be re-identified. This may occur for example when:

1. The study population is small.
2. The study of a rare disease or use of rare prescription.
3. The study population is from a known and geographically discrete area.
4. The study population is from a small identifiable demographic.
5. The data is used in combination with another dataset.

4. The Data Sharing Plan

The Data sharing plan provides information on the proposed management of research data for data sharing.

4.1. Data Sharing Plans for Funding Bodies

Some funding bodies, including MRC, CR-UK and Wellcome Trust, require data sharing included as part of the data management plan submitted for grant applications and is maintained for successful bids.

4.1.1. Digital Curation Centre DMPOnline tool

Funding bodies will provide template forms outlining their data management and sharing expectations; however, staff writing applications are advised to make use of the DCC Data Management Plan Online (DMPOnline) tool.

The DMPOnline tool provides template data management plans for MRC, CR-UK and Wellcome Trust. The DMPOnline tool offers the following features:

- Allows plans to be shared between multiple editors.
- Allows plans to be saved and returned to.
- Can export completed plans in a variety of common file formats.
- Provides context sensitive guidance from the funding body, DCC and researchers.

4.2. Mention in the Study Protocol

If a funding body requires a data sharing plan, the study protocol must state data will be shared as outlined in the Data Sharing Plan and Patient Information Sheet. This policy should be referenced along with the funding bodies own data sharing policy, if provided.

4.3. Features of the data sharing plan

The data sharing plan is normally expected to cover the following topics:

- A description of the dataset.
- Data security and information governance.
- Timescales for dataset release.
- Dataset discoverability.
- Data access governance – the Data request process.
- Provisions for not sharing.
- Preparation of a Data Sharing Pack.
4.3.1. Dataset Description
This is documentation which describes the dataset and allows end users to access and interpret the data. The dataset description should include:

- Format and scale of the data.
- A data dictionary and/or codebook.
- Additional metadata such as annotated case report forms.
- Reference to any data standards used.
- Transformations applied to the data, such as anonymisation of IPD or aggregation.
- Dataset and documentation formats.

4.3.2. Data Security
Measures must be taken to ensure datasets are adequately protected from unauthorised access and the disclosure of private and confidential personal data is prevented.

4.3.3. Data Access Governance
The plan should document the criteria and procedures followed by CCP for controlling access to the data.

4.3.4. Timescales for release or Period of exclusive use
CCP researchers should allow sufficient time for the research team to complete and publish all planned primary analyses. The Institute of Medicine recommends:

> It is reasonable to expect clinical trial data that will not be part of a regulatory application to be available for sharing no later than 18 months after study completion. This may not be realistic in some trials where trial participants are followed up beyond the time of primary analysis.[71]

The RCUK Knowledge Exchange Principles include a provision for privileged use stating:

> To ensure that research teams get appropriate recognition for the effort involved in collecting and analysing data, those who undertake Research Council funded work may be entitled to a limited period of privileged use of the data they have collected to enable them to publish the results of their research.[80]

The default CCP policy should be to make research data available within an 18 month time frame.

Once the period of privileged use has expired the dataset should be discoverable to external persons via the mechanism described in the data management plan.

4.3.4.1. Exceptions to the release timescale
Exceptions to the 18 month release timescale can be made when:

1. The Data Management Plan specifies a different period, for example as a requirement following the funding bodies own data sharing policy or prior agreement made with the funding body.
2. Participants are followed up beyond the time of primary analysis. In this case the Data Management Plan may provide arrangements for partial study data to be made available if the provision of such datasets will not adversely affect the trial team’s analysis of follow-up data.

4.3.4.1. Extending the release timescale
The research team may reserve the right to seek to extend the period of privilege use with a suitable justification.
The research team should be aware that the funding body for the research may need to be notified of any such extension and their agreement sought.

4.3.5. Discoverability and the mechanism for sharing
The plan should describe the mechanism for data sharing and include:

- how dataset requesters can discover the dataset
- the procedure for requesting access to a dataset
- the procedure for providing data for an approved access request

4.3.6. Limitations on data sharing
The data sharing plan should document any limitations on the data to be shared, including:

- Patient identifiable data will not be shared.
- Narrative data may not be shared.
- Data will be transformed to prevent re-identification of participants.
- CCP is not the data controller for the data or part of the data.

4.3.7. Data Sharing Pack
A data sharing pack is a suite of dataset and documentation which should be prepared once data collection and primary analyses are complete. The data sharing plan should outline the expected contents of the data pack as a manifest. The data pack would be expected to contain the data which is to be made available for sharing plus supporting documentation. The documentation should be to a sufficient level that a researcher can use the dataset with minimal guidance from CCP.

5. Data Access Governance
Data Access Governance covers the procedures for:

1. Dataset requesters to discover CCP data.
2. Dataset requesters to request CCP data.
3. CCP to assess and either approve or reject dataset requests.
4. Dataset requesters accessing CCP data for approved requests.
5. Logging requests and their outcomes.

5.1. Controlled Access
CCP has adopted a controlled access model for requesting access to its data assets. This model requires data requesters to provide information to support their request. This information is then reviewed by an expert panel at CCP in order to decide whether or not to approve or reject the request.

5.2. Discovering a Dataset
CCP must provide an adequate mechanism for other researchers to discover CCP datasets. This can include:

1. Publishing information on a study website.
2. Publishing information to a funder’s dataset discovery portal.
3. Publishing information on the CCP website.

Sufficient discovery metadata should be provided to allow the requester to evaluate the dataset’s appropriateness for their research question.

The data custodians contact information must be provided with instructions for how to submit a request.
5.3. Dataset request process

The purpose of the data request process is to ensure transparency and manage requests in a standardised manner.

The dataset request process consists of the following defined sequence of events:

1. Dataset Requester submits request.
   1.1. Request is logged
2. CCP Data Access Panel (DAP) reviews requests.
   2.1. Outcome is logged
3. DAP approves request.
   3.1. Dataset Requester is informed
   3.2. Data Sharing Use Agreement is signed
   3.3. Data is Released
4. DAP rejects request
   4.1. Dataset Requester is informed

5.3.1. Submitting an Access Request

Dataset requesters may request access by submitting a Data Request Form. The data request form asks for information intended to help facilitate a review of the request. This information includes the research proposal, statistical analysis plan, publication plan and evidence of the research group qualifications.

5.3.2. Reviewing an Access Request

The request should be reviewed by a Dataset Access Panel (DAP).

5.3.2.1. Dataset request assessment criteria

The request must be reviewed and either accepted or rejected on the following criteria:

1. The value of the research proposal to medical science and or patient care
2. The ability of the proposed statistical analysis plan to meet the scientific objectives of the research proposal
3. The publication plan
4. Potential conflicts of interest that may impact on the research proposal and measures to manage these conflicts.
5. The qualifications and expertise of the research team to conduct the proposed research
6. The CCP staff resources required to process the request
7. Risk of patients being re-identified or their privacy and confidentiality being breached
8. Failure of the data requester to demonstrate data can be kept secure
9. For data requesters who have successfully applied for other CCP datasets; failure to abide by the terms of a previous data sharing use agreement
10. Risk of datasets being used to misrepresent CCP research or bring CCP’s scientific credibility into disrepute

5.3.2.2. Dataset Access Panel

The role of the DAP is to review the request against the assessment criteria and either approve or reject on that basis.

The DAP membership should have the joint expertise to assess the scientific merit of the request and the feasibility of providing the requested data. Members are expected to include:

1 Based on the Wellcome Trust criteria used for requests made to https://www.clinicalstudydatarequest.com
The exact make-up of the DAP can vary between studies and may vary during the lifetime of a study.

5.3.2.3. Logging a request

The request should be logged to the DART application by the Records Champion. Research staff must notify the records champion of all requests received.

5.3.2.4. Timeline for acknowledging a request

The request must be acknowledged by the OAP within 20 working days. This is in line with the requirements of the Freedom of Information Act[9].

The OAP acknowledgement should indicate when a decision about the request will be made. This date should be within 3 months of the original request.

Notification must be logged to the DART application.

5.3.2.5. Incomplete Data Request forms

Incomplete data request forms should be returned to the requester within 20 working days with an instruction to complete and resubmit the form. Timelines will be reset once the completed application has been received.

Returned forms must be logged to the DART application.

5.3.2.5. Informal Data Requests

Informal data requests are any requests made using a media other than the standard data request form. Within 20 working days of receiving the request the requester must be informed to complete and submit a data request form. Timelines will be reset once the completed application has been received.

Returned forms must be logged to the DART application.

5.3.2.5. Accepting an Access Request

If the OAP accepts an access request, the data requester must be notified in writing and a timeline provided for when and how the data will be made available.

The outcome must be logged to the DART application.

5.3.2.6. Rejecting an Access Request

If the DAP rejects an access request, the data requester must be notified in writing and be provided with an explanation for why the request was rejected.

The CCP does not have an appeals process for rejected requests, however, based on feedback from the DAP, a data requester can amend their original access request to submit a new access request.

The outcome must be logged to the DART application.

5.4. Maintain a Data Access Request Log

CCP must maintain a log of all data requests using the DART application.

Since the data access request process is intended to be transparent, the log of all approved and rejected requests should be made available for public reference on the CCP website.
**5.5. Data Use Agreement**

The data use agreement is a document intended to ensure data are shared on a basis of mutual trust and understanding between the data provider and the data requester. It describes the conditions and responsibilities which must be agreed to prior to the release of data and should:

1. Prohibit re-identification of participants in anonymised datasets.
2. Detail required actions to be taken should re-identification occur. Required actions should include steps taken to mitigate the re-identification and sanctions taken against the requesting research team.
3. Agree the data to be included in the Data Pack.
4. Prohibit sharing of data outside of the data requester.
5. Detail agreements for CCP research team to provide statistical or data management advice, beyond that which is included in the Data Pack, with respect to the interpretation, analysis and or management of the dataset.
6. Agree to conduct and publish research as outlined in the approved research plan.
7. Detail the retention period for the data and rules for its destruction or return at the end of the research.
8. Agree to acknowledge the original research team originating the dataset in publications or to include them as authors in publications.
9. Agree to reference the dataset in publications.
10. Detail the consequences of non-compliance by the data requester.
11. The DAP may wish to include additional conditions and responsibilities as appropriate and in accordance with any requirements of funder data sharing policies.
12. The data use agreement must be approved and signed by an authorised QMUL signatory and the data requester.

**6. Preparation of the Data Pack**

Data approved for sharing must be released as a prepared data pack.

**6.1. Standard Data Pack**

If the data pack has been previously created, for example as stated in the data sharing plan for a study, then the data pack need only be checked against the dataset manifest before release.

CCP should not expect to incur unexpected resource costs.

**6.2. Custom Data Packs**

**6.2.1. Adequately resourced**

CCP must ensure it has adequate data manager and statistician resources to prepare the Data Pack and validate the dataset.

**6.2.1.1. Costing CCP resources**

CCP may request the data requester funds the cost of preparing the data pack. In line with the Freedom of Information Act Regulation 7(5), CCP staff time is to be charged at the flat rate of £25 per hour, irrespective of whether a high actual rate is incurred.

Charging is discretionary for the budget holder of the staff responsible for preparing the dataset.

**6.2.2. Data Pack Documentation**

An appropriate level of documentation must be provided with the data pack. This should include a manifest of the data pack contents.
6.2.3. Dataset Validation
The extracted dataset and any transformations applied must be validated to ensure the correctness of the dataset.

7. Data Release Process
Release of the data pack must occur as soon as reasonably possible following approval of the request and receipt of the signed data use agreement at CCP.

The data must be released as a Data Pack.

The data pack must be provided via a secure transfer channel acceptable to the ISMT.

References
[6] https://dmponline.dcc.ac.uk/
Appendix 1. Glossary

Aggregated Data
Data relating to multiple participants which has been pooled to display summary values rather than individual values.

Analysis (Derived) Research Dataset
Datasets derived from research datasets. These may be derived from third party datasets.

Anonymisation
The process of rendering data into a form where individuals cannot be identified and re-identification cannot take place.

De-identification
A process of transforming individual information to make re-identification of an individual less likely.

Data Controller
A person who determines the purposes for which and the manner in which any personal data are, or are to be, processed.

Data Dictionary
A document providing technical metadata about a dataset such as data types, constraints (i.e. range checks).

Data Pack
A prepared package of data and metadata to be provided to a data requester by the organisation.

Data Processor
An organisation that processes personal data on behalf of a data controller.

Data Requester
Any individual or research team making a request to the organisation for data.

Data Sharing Agreement
A document which details the conditions of use placed on a dataset shared by a data provider with a data user.

Effective Anonymisation
The process of rendering data into a form where it is highly unlikely individuals can be re-identified.

Individual Participant Data (IPD)
Data where each record in a dataset relates to an individual.

Metadata
This is data about the data. It could include a data dictionary, annotated case report forms, transformation process documentation.

Pseudonymisation
The process of identifying individuals in a dataset using identifiers which do not reveal their identity.

Records Champion
The member of staff nominated at CCP to be responsible for ensuring datasets are appropriately curated and documented.

Re-identification
The process of analysing data alone, or in combination with another dataset with the result that persons become identifiable. Also known as de-anonymisation.

Third Party Research Dataset
Datasets provided by a third party for research and subject to conditions of a data use agreement.
(Dataset) Transformation Process
Any process used to transform a dataset from one form into another. This includes program scripts used to apply transformations and instruction and validation documentation. A transformation process could be changing the dataset file format; an SQL to denormalise a relational database schema; a script to generate an anonymised dataset.