MRC Epidemiology Unit (EPID), University of Cambridge, School of Clinical Medicine

Data Access & Sharing Policy Version 2.3

Contents

1.0 Introduction

1.1 Background

This document is the Data Access and Sharing Policy for the MRC Epidemiology Unit (EPID), University of Cambridge School of Clinical Medicine. The purpose of the policy is to define the principles and processes for accessing and sharing our data. Appendix A details a glossary of specialised terms used, and appendix B has relevant key contact details.

The main research aims of EPID are to study the genetic, developmental and environmental determinants of type 2 diabetes and related metabolic disorders and to develop and evaluate strategies for their prevention. The Unit's research is supported by datasets developed through a range of observational studies; randomised controlled trials; natural experiments and surveys.

This policy has been based on the key principles that are applicable to publicly funded research related to population and patient studies[1]. It covers all studies where the University of Cambridge is the Sponsor on behalf of EPID or where EPID has primary responsibility for the data. The data represent a substantial commitment by our study participants, researchers, support teams and funders. As such, the policy outlines the processes in place to ensure the data is appropriately used, including safeguarding our participants, protecting the confidentiality of the data and maintaining the reputation of our studies whilst enabling maximum use of the data for public benefit.

1.2 Information on the data resources available

The scale of datasets is diverse, ranging from studies with under 20 participants to large epidemiological studies with thousands of participants. The complexity and the amount of data can be significant. Examples of this include genetic based studies where there are millions of markers per individual and the use of tri-axial accelerometer and spatial tracking data sets where storage requirements approach petabyte size.

The data types that we have are both quantitative and qualitative. Examples of quantitative data include questionnaires, model state files, research measures and images (e.g. objective physical activity, anthropometry, DEXA scans), data derived from biological samples (e.g. clinical chemistry, genetic data) and social and economic data. Qualitative data include questionnaires, field notes and audio recordings of interviews, transcripts and images.

The data has been generated for our seven MRC core-funded research programmes and a number of departmental programmes that are supported by a range of funders. Further information on the EPID research programmes is provided on the Research pages of our website. All studies which we are primarily responsible for are presented on the Unit website; study descriptions are provided and there are links on how to request data or access study metadata.

In addition to supporting direct requests for data, the Unit contributes to national initiatives to further enhance data access and re-use. For example, we provide access to EPIC-Norfolk data via the Dementias Platform UK Data Portal, which is a free-to-use remote platform where researchers can analyse multiple cohorts' data together. Researchers can apply for access to data from different cohort studies in a single application and receive a response in a target of 28 days, with online access within a target of 90 days if the application is approved.

We are part of the UK Longitudinal Linkage Consortium (UKLLC) and are contributing both EPIC-Norfolk and Fenland data. The UKLLC is a national research resource that is bringing together study data from more than twenty of the UK's top longitudinal studies linking it with NHS COVID-19 data, education data, occupation data and information related to where people live. Approved researchers may apply to access the data via the UKLLC website.

1.3 What is the scope of data access and sharing and why do we need a policy?

All study data collected are of sufficient scale or uniqueness to be of potential value to the wider research community. Our aim is to maximise the use of our research data for the benefit of the public. The value of the information collected develops as it goes through the 'data lifecycle' of being assembled, quality controlled, analysed and made accessible to other researchers. To ensure that our research data is both effectively and appropriately used, we have put in place processes and practices to ensure that the ethical, legal and any security constraints are adhered to. This Data Access and Sharing policy is based on:

- The requirement to **protect participants** within the scope of their informed consent.
- Ensuring **compliance** with the UK's General Data Protection Regulation 2018[2] and Data Protection Act 2018[3] and that data confidentiality and security is consistent with the rules of the University of Cambridge and any funder requirements whilst also meeting legal requirements and best practice[1].
- Fostering **high-quality** research using robust and equitable systems for data access and sharing.
- The **governance** of access is appropriate and proportionate to the nature and scale of the study and associated risk(s).

EPID aims to allow as wide access to the data as is possible in a timely manner. As soon as the consistency and validity of datasets has been assured, they can be made available in a de-identified form where risk of participant identification has been minimised. No embargo periods for data sharing are envisaged unless required for the protection of the participants or intellectual property rights. If any such period is required, it will be of appropriate duration related to the needed protection and such need will be determined by the Study Investigators or Study Steering Committee and Senior Data Manager, in discussion with the Unit Director.

2.0 The governance of data access and sharing

This section of the policy covers the systems that are in place to enable active sharing of research data and to ensure that our overarching management systems are appropriate and proportionate.

2.1 Management

The management of EPID actively supports the sharing of data and continues to invest resources and funding into facilitating access and sharing of our research data. The responsibilities of the EPID management team are to:

- Review and approve the Data Access and Sharing policy.
- Give clear direction and assignment of specific roles and responsibilities for data access and sharing procedures.
- Provide a clear direction for future sharing initiatives.
- Help provide the resources and funding required for enabling continued data access and sharing.
- Highlight any unmet funding or strategic needs to research funders.

For ongoing studies with an active Chief Investigator (CI), Principal Investigator (PI) or Study Committee, the CI/PI or Committee will be responsible for deciding on applications for access and sharing data in line with this overarching policy. For a study that is no longer active and does not have a CI/PI or Committee, the Unit Director will be considered the responsible custodian and decide on data sharing requests.

2.2 Data Access and Sharing Reporting

The Unit records details of all data requests that are made and their outcomes and periodically reviews this.

2.3 Data Management and Governance

The Senior Data Manager is responsible for ensuring that the arrangements for the data access and sharing policy are functional. The primary tasks of the Senior Data Manager are to:

- Provide guidance on the data access and sharing arrangements for requestors.
- Ensure that the operational aspects of data access and sharing are maintained, monitored and improved in line with the University of Cambridge, Medical Research Council (MRC) and best practice guidance.
- Monitor and track the location of all data collections.
- Keep the staff informed of relevant matters relating to data access and sharing.
- Provide guidance to staff on the data sharing and access policy

3.0 Data Sharing

The Data Sharing section describes the process for a requester to follow to access data.

3.1 Data request queries

Information on our studies can be found through our study pages. For all general questions about data sharing or if the study data you are interested in is not listed please contact datasharing@mrc-epid.cam.ac.uk.

3.2 Submission of data request

A study data request form can either be found on the study website or, for some studies, provided by the study or data management team on request. The completed data request form should have a title for the proposed analysis, the research question being asked including an objective and the planned outputs. A list of variables is needed including outcome variables, exposure variables and any covariates requested.

The completed study data request form should be submitted to the relevant Study Committee, CI or PI for their consideration using datasharing@mrc-epid.cam.ac.uk. For any completed studies that do not have a named custodian the Senior Data Manager will ensure that the proposal is considered.

The study data request form for the EPIC-Norfolk study is available from the EPIC-Norfolk website, along with detailed information for researchers.

3.3 Deciding on data sharing requests

The relevant CI, PI, Study Committee (or in the case where there is no named custodian the Unit Director) will decide whether a proposal is accepted or not. The following criteria will inform the decision:

Capability and capacity for high quality research

- The proposed research is bona fide; the requestor is a bona fide researcher working within a bona fide research organisation.
- The research is for public benefit and is not being carried out for personal or commercial gain.
- The research is methodologically sound and the request would provide the appropriate data to address the research question.
- There is a commitment to publishing the findings in an openly accessible and peer-reviewed publication in a timely manner.
- Newly derived variables from original study data will be returned to EPID so that they are available for other research in the public interest.
- The proposed research is consistent with the study ethical approval and terms of participant consent.
- No barriers to ethical and regulatory approvals are envisaged.
- Appropriate mechanisms can be put in place to maintain the confidentiality of participants' identity and meet all data protection and data security requirements.
- There is no significant risk to the reputation of the study, or the continued involvement and retention of study participants.

Funding requirements

The proposed research is compatible with the requirements of the MRC or other funders of EPID research to deliver an existing or planned programme of research.

Commercial use

Requests for commercial use will be dealt with on a case-by-case basis in liaison with Cambridge Enterprise.

Should an application be declined, the requester can submit a revised application to the Unit Director indicating how the concerns have been addressed. In the event of any dispute the final decision will be with the Unit International Strategic Advisory Group. Any requests that are not supported are recorded and included in the summary data provided to the MRC and the ISAG.

3.4 Terms of access and mechanisms

Terms of access

Data that identifies an individual will not be made available. Data sharing is being facilitated to the extent that it is possible without the risk of disclosing personal information. We also ensure that consent, ethical, data protection and security requirements are met. EPID operates rigorous procedures for de-identifying the data, and the data is made available to researchers in a de-identified form. Any retained link files are held securely and are not made accessible to requestors.

Recipients of the data must agree not to link de-identified data provided with any other data set without the permission of the custodian. There must be no attempt made to identify any individual using the data provided. Data will be provided against a specific set of use rules, data transfer or collaborative agreement.

No preferential or exclusive access will be provided. Intellectual property arrangements will be defined within appropriate data sharing agreements and may vary on a case-by-case basis. Access to linked data will only be provided if permitted by the terms imposed on EPID by the data owners and any data access conditions will be passed on.

The technical nature of data and potential risks to participant anonymity being compromised means that some data may need to be supported in its wider use. Data from such studies will be provided as 'dependently available' with facilitation to ensure appropriate use.

Requesters may be asked to submit progress reports to the custodian or relevant Study Committee. Requesters are asked to provide notification of all scientific publications so that outputs can be collated.

EPID reserves the right to publish relevant information relating to the details of the data access and sharing arrangements. Requesters who do not wish details of their study to be openly available should state this in their application.

Mechanism of access

For most of EPID's studies, data is made available via the University's SRCP platform.

Details of the platform are here and an example of using this resource with a study, in this case EPIC-Norfolk, can be seen here: https://www.epic-norfolk.org.uk/for-researchers/data-sharing/data-access/.

Details on the costs of using SRCP are typically available on the relevant study web page.

When working on the SRCP, requestors will be remotely accessing a virtual system within the Unit to work on the data. They will not be able to copy the data to another location and are not able to remove anything from the system. Only the summary statistics, results and the associated code, not individual level data, can be removed once they have been approved.

For some studies, where the data sets are small and sensitive data access is not required, it may be possible to provide data provided through secure data transfer under the terms of an institutional data transfer agreement.

Alternatively, the most appropriate mechanism may be through the national initiatives described above.

3.5 Research Outputs

It is expected that the work of EPID staff, PI's, collaborators and funders in the setting up, collecting and maintaining study data is formally recognised. Specific arrangements for acknowledgements, publications and for ensuring the confidentiality of the data will be outlined in the access agreements and on the study websites.

3.6 Costs

The requester may be required to cover the costs for the data sharing or access depending on the nature of the arrangements. Estimated costs can be provided after an initial review of the application. Where a requestor is working on SRCP these costs are shown on the relevant study web page.

4.0 References

[1]MRC Policy and Guidance on Sharing of Research Data from Population and Patient Studies

[2]Regulation (EU) 2016/679 of the European Parliament and of the Council of 27th April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation) as it forms part of the law of England and Wales, Scotland and Northern Ireland by virtue of section 3 of the European Union (Withdrawal) Act 2018.

[3]Data Protection Act (2018), United Kingdom of Great Britain and Northern Ireland.

Appendices

Appendix A – Glossary

CI	Chief Investigator
Custodian	The person, organisation or committee responsible for the data (typically this will be a study's Chief or Principal Investigator)
Data life cycle	The process by which data is collected, cleaned, quality controlled analysed, shared and archived
EPID	Abbreviation for the MRC Epidemiology Unit at the University of Cambridge, School of Clinical Medicine
MRC	Medical Research Council – UK government funded medical research organisation
PI	Principal Investigator
Requester	An individual or group of bona fide researchers requiring access to data

Appendix B - Key contacts

Director: Professor Nick Wareham, nick.wareham@mrc-epid.cam.ac.uk

Head of Research Operations: HeadofResearchOperations@mrc-epid.ac.uk

Senior Data Manager: SeniorDataManager@mrc-epid.cam.ac.uk

Data sharing request and general information: datasharing@mrc-epid.cam.ac.uk

Last reviewed 24/01/2025