Data sharing glossary

Definitions for terms used throughout the MRC data policy and related guidance.

**Data** are qualitative or quantitative information created or collected in the course of research. Sources may include experimental measurements, clinical measurements, observations and information obtained via survey questionnaires, interviews, non-research documents (for example, letters) or focus groups.

**Metadata** are ‘data about data’, providing concise information about the content, quality, condition, and other characteristics of datasets. Metadata are vital for understanding and enabling the efficient use of shared data.

**Derived data** are data values that have been obtained as a result of modifying or combining the data values that were originally acquired about study participants, the raw data. They should be documented and where they have value for new, they should be discoverable.

**A study** is a funded programme of research using data from and about individuals that is representative of a defined human population.

**An MRC study** refers to a research programme or project wholly or partially funded by the UK Medical Research Council (MRC). The roles and responsibilities of the study Director or Principal Investigator and of a data requester are explained in detail within the policy.

**The study team** comprises the individuals who have responsibilities for delivering the MRC-funded research programme, as identified by the Director / PI and their management team.

**Lifetime value** recognises that a study has a life cycle, from study design, through data collection and processing, to sharing, publication and preservation. Some study life cycles involve a series of repeated data collection; others a sequence in which different kinds of data are accrued; and others a mixture of both. The point(s) at which value is derived by sharing differs according to the study design and the properties of the data.

**Study participants** are the individuals from a defined human population, about whom data are collected and interpreted, with their consent. They may comprise a cohort.

**A cohort** is used in the epidemiological sense of a defined group of healthy participants or patients that is the subject of a programme of observational research, often longitudinally with periodic measurements over time.

**A study dataset** comprises all of the data from and about the participants of that study, data derived from those data, and metadata about the collection, governance and interpretation of those data. In practice, complex studies will organise all this as several datasets. The study dataset might include some data that are not appropriate or permitted to share.

**A dataset is considered a community resource** if it has been established primarily as a resource for a research community, rather than to address a specific investigator-driven set of questions. The conditions of its funding stipulate that the intention is the widest possible access and good use beyond the study team that collect the data.
The MRC Research Data Gateway is a new online resource developed in 2011 to enable the discovery of MRC-funded studies and the variables they have collected. It supports initial enquiry about data-sharing and access to studies’ research data. Via the gateway users can find, select and export variables to support data requests from individual studies. The gateway contains the Population and Patient Research Data Directory.

The Population and Patient Research Data Directory is a directory of MRC-funded studies and their variables. The directory does not contain research data themselves, only metadata (information about the data). The directory can be searched via the MRC Research Data Gateway.

Data-sharing covers a range of activities whereby data collected for a specific research purpose, or as a common resource, are used for new research purposes, thus increasing the value of the data and the eventual benefits for patients and the public. This description is consistent with the use of the term by other research funders. It differs from the ONS definition, which includes the concept of “joint ownership but different uses for the data” and which is not necessarily relevant here.

A research study has shared data when a researcher who is not a member of the funded study team (i.e. an external researcher) has been able to undertake new research using the study dataset (derived data or raw data).

A data-sharing request is a documented request by a potential new user (the data requester) for access for new research. For an illustrative workflow see the section on access governance within the policy.

Data access focuses on the interaction between a potential new user and an MRC study. The interaction generally includes the requester’s initial enquiry, the governance of the request for data (or for analyses), the selection and the processing of the data. It extends also to the formalisation of the relationship between the study and new user through a formal data-sharing agreement. For the purposes of this guidance, data access includes both transfer of data from the study to another party for analysis and analyses conducted by the study itself on behalf of, or with, the new users.

New uses are additional to the purposes for which the study is funded, usually research, and always bona fide. They may be similar to and complement the purposes for which the study is funded, or very different. They can involve data linkage, analyses of previously unanalysed data, reanalysis and validation of previous findings and replication studies, meta-analyses, modelling, pilot work and so on. New uses may also involve educational, knowledge transfer and other activities. In terms of scale they can encompass large, national and international collaborations to small, individual student projects.

New users are parties who use data for new uses. It follows that new users can include both (i) new parties that are not existing co-investigators, and (ii) study researchers and collaborators already working on the study’s funder-approved programme.

External users are researchers who are not members of the study team.

Personal data are data which relate to an individual. This is a broader definition than that of the Information Commissioner’s Office (opens in new window) where the
emphasis is on a living individual (Data Protection Act 1998). In either case, such data may be used to identify that person directly or indirectly through association with other data or information.