

MRC/UVRI Uganda Research Unit on AIDS  
Data Sharing Policy


Version	Date	Comments
0.1	30/10/2014	First draft
0.3	05/05/2015	Second draft
0.4	01/12/2015	Final draft

**SIGNATURES OF APPROVAL**

Date:

Version: 1.0

Signatures

 29<sup>th</sup> Feb 2016

---

## MRC/UVRI Uganda Research Unit on AIDS: Data Sharing Policy

### 1. Purpose

The purpose of this document is to summarize the conditions under which data collected by the MRC/UVRI Uganda Research Unit on AIDS will be made available to other *bona fide* researchers, the way in which such researchers can apply to have access to the data, how each application will be considered and how data will be made available if an application for data sharing is approved.

### 2. Introduction

The MRC has adopted a Policy on Research Data Sharing<sup>1</sup> which applies to all MRC-funded research. The rationale for the Policy is that data collected from population and patient studies provide many opportunities for collaboration across diverse research disciplines. The value of data collected from study participants increases as the data are cleaned, organized, analyzed and the outputs of the analysis are made accessible to research data users. Creating the value in the data represents a substantial commitment by all parties involved namely the funders, the researchers and the study participants. Given the effort and resources that have gone into data collection, there is a responsibility to the general public to use the data as widely as possible.

The overarching aim of the MRC for data sharing is “to maximize the life-time value of research data assets for human health and to do so timely, responsibly, with as few restrictions as possible, in a way consistent with the law, regulation and recognized good practice.”

The MRC Policy Document lists a number of benefits of data sharing including:

- Enabling new research questions to be answered using existing data
- Promoting collaboration between different research teams and diverse disciplines
- Sharing of knowledge about best methods for data collection, linkage and analysis
- Ensuring that collected data are cleaned, well documented, with value added
- Independently verifying established research findings
- Development and testing of new research methods
- Using to best effect the gift of data made by study participants

The MRC policy reflects a growing international consensus on the need to preserve and share research data sets in a manner that maximizes their long-term value. For example the Wellcome Trust has a policy on data management and sharing<sup>2</sup> and advocates making research data widely available to the research community in a timely and responsible manner so as to ensure that these data can be verified, built upon and used to advance knowledge and its application to generate improvements in health. They note that the success in maximizing the value of research data depends on fostering a culture in which both data generators and data users adopt good research practice and act with integrity and transparency in managing, using and sharing research data.

In 2008, the Wellcome Trust commissioned an analysis of the barriers to data sharing among epidemiological and public health researchers, and convened an expert workshop. The outcomes of this work were discussed at the global ministerial forum on research for health held in Bamako (Mali) in November 2008. The paper prepared for the Bamako meeting will be referred to as the "Bamako data sharing code of conduct"<sup>3</sup>.

The Bamako code is based on the belief that data collected with public and philanthropic money are "public goods". However the discussion document points to justified concerns that act as barriers to data sharing. Scientists in poorer countries put a lot of effort into collecting data that is then analyzed by scientists from richer countries, and hence there is a need to increase the incentives and capacity to manage and analyze information in developing countries. They also acknowledge the need to protect the researchers who collect the data, and point to the need for a limited amount of time being given to them to explore their findings before sharing the data with other researchers. In addition they point to the need to protect the people participating in research, so there is a requirement for efficient techniques for anonymizing records, and in certain situations access to data may need to be restricted. It has also been noted that conclusions derived from misunderstandings or erroneous analyses of data can harm the reputation of a study and damage the trust of the community in the organization, discouraging further participation in research, so there is a need to screen prospective data users.

This document sets out the data sharing policies of the MRC/UVRI Uganda Research Unit in the light of the prevailing MRC policies and the arguments given above for data sharing, taking into account some of the concerns raised in the Bamako document.

### **3. Models of Data Sharing**

The Bamako document describes four different models of data sharing, which are summarized below with some examples.

#### **3.1 Fully open access**

Data (anonymized where necessary) are made available in machine-readable formats on publicly-accessible websites. An existing example of this is genetic sequencing data.

#### **3.2 Controlled public access**

Data are made available to authorized users after a screening process. An example of this is data from Demographic and Health Surveys.

#### **3.3 Collaborative access among scientists**

Data are made available to other scientists in a collaborative network. An example of this is INDEPTH demographic surveillance site data.

#### **3.4 Exclusive access for primary researchers**

Data are only available to the research team involved in data collection and their institutional partners. This was the norm in health research data collection which the current policies are seeking to change.

#### **3.5 MRC/UVRI Uganda Research Unit data sharing model**

The approach in the policy of the MRC/UVRI Uganda Research Unit will be to follow the collaborative access model as the default model, but to consider the controlled public access model for studies after they have closed. In addition, the researchers in the MRC/UVRI Uganda Unit will have a limited period of exclusive use of the data; for example in the Data Management Plan for the HIVCOMB study, it is stated that the baseline data will be made available six months after the publication of the baseline paper. Study leaders ought to ensure that data sharing is covered by approval from the relevant research ethics committee (REC). In instances where requested data is not covered within the REC's approval; then approval should be sought from the relevant REC.

### **4. Technical Requirements for data sharing**

In order to make data suitable for sharing, the Unit will need to prepare suitable metadata, (or data describing the data) or in the case of a qualitative data set a dataset contents list. For quantitative data the key item of metadata for any given data set is the Data Dictionary, which gives a complete description of each table in the database, with a list of variable names, variable descriptions and codes used. In general there will be a separate table in the database for each questionnaire or Case Report Form (CRF) used in the study. The current model used for data dictionaries is to have an Excel spreadsheet, with one sheet per table. Additional metadata could include a description of the study and how the data were collected. In many cases this could be given by a reference to a key publication describing the study.

It has been pointed out<sup>4</sup> that providing metadata/qualitative dataset contents lists and proper documentation in particular and data sharing in general, will require additional funds and attention to improving data management skills and making data management a more rewarding career option.

Current funding applications to the MRC require the development of a Data Management Plan, which includes provisions for data sharing. It is thus imperative that going forward, the creation of the required metadata is properly funded. For existing studies as well as studies that have already closed, the MRC policy on data sharing and preservation points out that prospective users could be charged a fee to cover the costs involved in preparing data for sharing, including the preparation of a Data Dictionary and other metadata.

The MRC and Wellcome Trust policy documents point to the desirability of following data documentation standards such as the Data Documentation Initiative (DDI). While this is desirable for the future, currently the Unit does not have the expertise to use this standard, and funds would need to be sought to train data management staff in the requirements if this is to be adopted in the future.

In addition, where study identifiers contain information that could identify participants, these will be anonymized before the data is shared. An example of this is the General Population Cohort, where the identifier includes a village number and a household number within village; the anonymization would replace the village number with a new random village number and similarly replace the household number with a new random household number, so that researchers could still identify participants from the same village and participants from the same household, but would not be able to say which village or which household. This approach is currently being used in data sharing with

the Pangea database. Similarly if the database contains GPS coordinates, these will be anonymized by adding a large fixed constant to each coordinate – this will still allow the investigation of spatial clustering without identifying the location of a given participant.

It should be noted that the anonymisation of qualitative data is often challenging because it requires the removal of all identifiers. This may be difficult to achieve for life history data, for example or reports on particular events or situations. Careful consideration of the data sharing approach for such material will be required.

#### **5. How to apply for data sharing**

Any given dataset within the Unit will be deemed to belong to a single Project. In general the Project will be situated within one of the five Programmes of the Unit, but in some cases the Project may be within a number of Programmes (e.g. the Superinfection and Molecular Epidemiology projects are both within the Basic Science Programme, but also have components within the Epidemiology and Prevention Programme and the Social Science Programme). In order to consider data sharing requests each Project will constitute a Project Management Group, where this does not already exist (for example many Intervention Studies already have Trial Management Groups which would fulfil the requirements). The Project Management Group for the purposes of consideration of Data Sharing applications for **quantitative** data will contain at a minimum the Unit Director or Deputy Director, the Project leader, the Head(s) of Programme(s) in which the Project is housed and a Senior Statistician. The Project Management Group for the purposes of consideration of Data Sharing applications for **qualitative** data will contain at a minimum the Unit Director or Deputy Director, the Project leader, the Head(s) of Programme(s) in which the Project is housed and a Senior Social Scientist. Both a Senior Statistician and a Senior Social Scientist should sit on the Project Management Group for data sets which include both qualitative and quantitative data. In instances where the Trial Management Groups substitute the Project Management Groups; approval from these as well as from the Unit Directors will be required.

All applications for Data Sharing must be made using a standard Data Sharing Form (attached to the policy) and in the first instance addressed to the Project leader who will inform the Project Management Group and convene a meeting to consider the application, if an agreement relating to the request cannot be reached through email.

As part of the Data Sharing application, the applicant should give a detailed description of the purpose for which the data will be used, which could take the form of a proposal for Secondary Data Analysis. The applicant should be a *bona fide* researcher, and in the case where the applicant is not known to the Project Management group, he or she could be asked to submit a brief CV. In the case of postgraduate students who wish to use the data, the supervisor should act as applicant. The applicant should agree to share any prospective publication with the Project Management Group prior to submitting for publication, to ensure that the work of the Unit has not been misrepresented; the Project Management Group should be given a period of four weeks in which to comment on the manuscript, and if they fail to comment this can be taken as confirming that the manuscript could be submitted. In addition applicants should undertake to point out any errors that they find in the data and to acknowledge the Unit for the provision of the data in any publication, and should also give consideration to including Unit staff as collaborators and co-authors on the work where this is possible.

Note that in line with the MRC policy on Data Sharing, there should be a good reason for turning down a request.

Requests should be considered within 4 weeks of being made, and a decision communicated to the applicant as soon as possible thereafter. Note that in some cases the application will be a two stage process as the applicant might need access to the metadata/contents listing for a study in order to decide whether or not to apply for data sharing, and if they do decide to apply for data sharing which variables to request. The data can be provided in a range of formats. Consideration will be given to the preference of the applicant (taking into account that sharing large qualitative data sets from Nvivo for example, may not be practical if the data management system used by the Unit and the applicant are not compatible). Data will be sent as a password protected file together with the data dictionary/contents list.

Note that the MRC policy refers to discovery of databases on the project website. Currently this will not always be possible as not all projects have their own websites or pages on the Unit website. In this case prospective applicants should direct a general enquiry to the project leader.

## References

1. *MRC Policy on data sharing and preservation*. London: Medical Research Council; 2008. Available from:

<http://www.mrc.ac.uk/PolicyGuidance/EthicsAndGovernance/DataSharing/PolicyonDataSharingandPreservation/index.htm>.

2. *Policy on data management and sharing*. Wellcome Trust; 2007. Available from:

<http://www.wellcome.ac.uk/About-us/Policy/Policy-and-position-statements/WTX035043.htm>.

3. *Sharing public health data: a code of conduct*. London: Wellcome Trust; 2008.

Available from: <http://www.wellcome.ac.uk/About-us/Policy/Spotlight-issues/Data-sharing/Public-health-and-epidemiology/index.htm>.

4. Pisani E, Whitworth J, Zaba B, Abou-Zahr C. "Time for fair trade in research data"

Lancet (2009) DOI: 10.1016/S0140-6736(09) 61486-0