MRC Regulatory Support Centre: Retention framework for research data and records

Background and scope

You may be asked to make decisions about the retention of data / records resulting from research. This guidance provides a framework for making risk proportionate decisions about retention. It also outlines relevant MRC Policy and where you can access further guidance so that you can make these decisions with appropriate support.

This framework is designed primarily for MRC staff (i.e. researchers and those who have research governance responsibilities within MRC research units and institutes) and for those whose research is sponsored by the MRC. It may be used to inform policy within MRC university units and MRC centres. It is *not* designed for researchers to make decisions about data / records retention on their own; but rather for these decisions to be made in partnership, with the involvement of both the researcher and their organisation (Data Protection Officers, sponsor representative, etc.).

Preservation of data from research is important for a number of reasons:

- it supports Open Science and promotes reproducibility (important for transparency);
- it provides an audit trail (important for good governance);
- it enables future research opportunities through data sharing.

However, data / records retention and long term archiving do not come without financial cost. Therefore it's important to weigh up the risks of retaining data / records versus the risks of their destruction and then take well-informed, risk proportionate decisions about retention and long-term archiving based on this information. Efforts can then be concentrated on managing data / records retention for research posing 'high risk' and taking a light touch approach for all other research.

Examples of research posing 'high risk' might include:

- Research studies where there is a greater likelihood that scientists will be asked to provide data / records to enable reproduction of findings as their research findings are high profile / high impact or in some way contentious; or
- Research studies where there is a greater likelihood that participants might suffer harm as a result of the research.

MRC Policy

MRC's Retention Policy is outlined in Good Research Practice: Principles and Guidelines and applies to all MRC funded research (i.e. for MRC research units and institutes, MRC university units, MRC centres and MRC grant holders). Good Research Practice is intended to complement statutory or regulatory requirements (e.g. the Clinical Trials Regulations, Data Protection Act,[1] etc.).

There is additional guidance on the MRC's intranet (MRC Hub) for the management of corporate records in MRC's Head Office and units in the MRC Records Management Policy and MRC Records Retention Schedule. The Concordat on Open Research Data, RCUK's Policy on Open Access, MRC's Data Sharing Policy and The Home Office may also be relevant.

For *basic research* - Research data and related material should be retained for a minimum of 10 years after the study has been completed.

[1] In terms of compliance with the Data Protection Act, the research exemption allows researchers to keep data *indefinitely* as long as it is still useful and there is an intention to use them further for research. If research data / records are no longer useful, then you should consider their disposal.

MRC Regulatory Support Centre: Retention framework for research data and records

For *population health and clinical studies* - The MRC expects that research data / records be retained for 20 years after the study has been completed.

Studies which propose retention periods beyond 20 years must include valid justification.

However, longer retention periods for both *basic research* and *population health and clinical studies* may be appropriate in some cases. For example:

For *basic research* – Retention periods of 10 years+ may be more appropriate where there is the potential for Intellectual Property to arise (e.g. laboratory notebooks could be retained indefinitely).

For *clinical studies* – It has been recommended that a minimum of 25 years may be more appropriate for some studies, for example:

- Where a clinical trial of an investigational medicinal product (CTIMP) involved a participant who became pregnant / whose partner became pregnant during the course of that trial, any claim for damages may arise from either the participant or their child. For children the 3 year statute of limitations does not begin until they reach the age of 18 and then lasts for up to 4 years (3+). Therefore any claim with respect to the child in the above example could arise over 20 years later and hence a 25 year retention period might be a sensible approach in these cases. For further information on risk in Clinical Trials see Guideline on strategies to identify and mitigate risks for first-in-human clinical trials and Risk-adapted Approaches to the Management of Clinical Trials of Investigational Medicinal Products.
- For those who lack capacity to consent there is no statute of limitations so a potential claim could arise at any time. Between 20-25 years would seem a sensible point to review what is being retained and then determine whether records should be kept longer.
- You may wish to consider retention periods beyond 20 years for research data relating to studies where there is a greater likelihood that scientists will be asked about reproducibility (i.e. because their research findings are high profile/high impact or in some way contentious).
- Research data relating to studies which directly inform national policymaking should be considered for permanent preservation in an appropriate archive (e.g. the MRC have permanently preserved proposals relating to the Human Genome Mapping Project in The National Archives).

Organisational policies

In order for organisations to carry out their responsibilities under the Data Protection Act as well as research governance responsibilities[2] all will have their own policies on data / records retention. Similarly, funders, journals and others may have their own requirements. You will need to understand and meet the policies of *all* of the relevant organisation(s) involved in your research.

Good Research Practice: Principles and Guidelines (GRP) outlines retention periods for all MRC sponsored studies. Where MRC is the sponsor or employer, the retention periods stated above are expected to be followed. Where MRC is purely the funder, your relevant organisational policies will need to be followed. MRC retention periods can be used as a guide to inform other organisations' requirements.

For further guidance please see the Retention decision-making flowchart in Annex 1.

[2] These are outlined in the Health Departments' Research Governance Frameworks for health and social/community care research. In addition the Medicines for Human Use (Clinical Trials) Regulations outline requirements for Clinical Trials of Investigational Medicinal Products.

MRC Regulatory Support Centre: Retention framework for research data and records

Expectations of other organisations

It is important to consider any external organisations involved in your research (e.g. those providing regulatory oversight, your collaborators, funders or journals) and what their expectations may be with respect to data / records retention. (In some cases these expectations may be outlined in an agreement).

For example any organisation providing regulatory oversight may either audit and/or inspect your premises (e.g. Health & Safety Executive, Home Office, Human Tissue Authority, Medicines and Healthcare products Regulatory Agency, etc.). You should be aware of the research data / records that they are likely to want to see and whether you can provide access to these.

If your research involves external collaboration with the NHS they will have their own expectations and may check compliance. You should be aware of these expectations, be confident that you are in a position to deliver them and consider any long-term requirements (i.e. how long do they expect you to retain any data / records and who should pay for this?)

What to keep?

During the lifecycle of any research project a whole range of data / records will be generated, recorded, collected or used (e.g. information recorded in laboratory notebooks, images, machine generated print-outs, home office licences, regulatory and ethics approvals, questionnaires, participant information sheets, consent forms, etc.).

What to keep will very much depend upon the type of research that you do and what you need in order to enable an understanding of what was done when, how and why. You should not make these decisions alone, your organisation should have a policy and/or process on how any decisions about what to keep are to be made (similar to the MRC's Records Management Policy and Records Retention Schedule). This process should be led by someone who is familiar with the research data / records (e.g. the Principal Investigator or Trial Steering Committee) making decisions in line with their organisation's policy and sponsor expectations (where appropriate).

For some *basic research* the information recorded in laboratory notebooks may provide sufficient detail to enable an understanding of what was done when, how and why.

Whilst for some *clinical studies* a whole raft of data / records will be required in order to enable that same understanding (e.g. research protocols, participant information sheets, consent forms including details of version control / when each version was being used, individual signed consent forms, details of randomisation and what drug each participant received, any changes to the drug regime over the lifecycle of the study – including documentation which provides details and timings of the review / recommended change: Committee minutes, etc.).

Decisions about what to keep need to be made on a case-by-case basis with respect to which data / records are required in order to meet the aims of data preservation (e.g. to support open science and reproducibility of findings, to provide an audit trail and to enable data sharing) and where appropriate these must be in line with the expectations of participants. For example if a participant were to withdraw consent from your study, then you would retain their data / records in line with what was agreed with respect to withdrawal (e.g. the participant may have agreed that you can retain the data / records that you already hold, but that you won't accrue any further data).

A risk proportionate approach should be taken to decisions about data / records retention and long-term archiving. There should be some form of risk assessment for all studies to determine those which are 'high risk' and require additional management. For example you may decide to manage 'high risk' studies by retaining data / records for a longer period of time and/or in their original format. That's not to say that *all* data / records from 'high risk' studies (or indeed any study) should be kept for longer and/or in their original form as the scenarios below demonstrate:

MRC Regulatory Support Centre: Retention framework for research data and records

What to keep – scenario 1

For some basic research what to keep may be influenced by the outcome. A record of what was done when, how and why will always be kept in laboratory notebooks. If an experiment didn't work (rather than produced negative results) then the original output may not be as valuable to preserve and may not necessarily be kept (although this decision should be clearly documented and signed off). You may also consider retention in the short term for protocol optimisation reasons.

Some outputs take up a lot of physical space (e.g. plates and cell cultures). Normally plates are captured via a plate reader and the images printed and/or saved electronically. Cell lines can also be preserved for future research use or for reproducibility reasons. As such the original cell cultures and plates would usually only be kept if deemed useful and/or unlikely to deteriorate.

What to keep – scenario 2

Some outputs can deteriorate over time. Gels and immuno blots are normally photographed and then discarded as they are too fragile to maintain. Stained microscope slides may be retained in the short term but may not be worth keeping after analysis as staining can fade over time (and slides can take up a lot of physical space).

Thermal print outs (such as ECG print outs) fade and degrade over time so it may be more appropriate to take the necessary data from these print outs and transpose into electronic form. Once these data have been subject to some form of quality assurance process (ensuring data accuracy and validity), a process which should be clearly documented and signed off, the original thermal print outs can then be destroyed.

What to keep – scenario 3

It's worth considering whether keeping a paper record in its original form is necessary in the longer term. If you were to recruit healthy volunteers to a volunteer database, you might ask them to complete a form to capture some personal details (i.e. name, address, contact telephone number) as well as the types of research in which they might be interested in participating. All of these details could then be entered into a database and the original form stored for reference.

Is it worth keeping these forms in the longer term? The information that you need is now stored in the database. Therefore as long as there is some form of quality assurance process to ensure that data entered into the database are accurate and valid (clearly documented and signed off), then the original healthy volunteer forms can be destroyed after an agreed period of time.

It's worth considering the risk of keeping these healthy volunteer forms in the longer term and balancing this against the risk of destruction. The forms contain 'personal data' and are therefore subject to Data Protection Act requirements. Storing these in the longer term, particularly if this will mean storing them with an external archiving organisation, presents a potential risk to your volunteers' privacy and a cost burden. Additionally, people move and so the original data may become inaccurate.

What format - Paper versus electronic?

When data are created they will either be created in hard copy (i.e. paper, images, etc.), in electronic format or both. Where both are created a decision should be made with respect to whether both formats will be retained in the longer term. These decisions should made by someone who is familiar with the research data / records (e.g. the Principal Investigator or Trial Steering Committee) in line with organisational policy and sponsor expectations (where appropriate).

Good Research Practice: Principles and Guidelines allows for the conversion of original paper records into electronic format (e.g. by inputting data into software or by scanning). In principle, both of these options can be a good solution to reducing long-term paper archiving costs.

However neither comes without cost, both are time consuming and there is the potential for error as explored in the following scenario.

Scanning scenario: An MRC Unit wanted to stop retaining the original hard copies of a screening form and consent form for people undergoing an MRI scan, and switch instead to retaining an electronic scanned copy. In line with Good Research Practice, as long as there is a system in place which serves to check that these screening forms and consent forms have been scanned to an appropriate standard, are held electronically and that this process has been signed off and documented, then it would demonstrate that these data were: "… validated using quality assurance procedures" and therefore destruction of the paper copies could be justified.

It's about balancing risk:

- how likely is it that you will need to provide these forms to someone in the future (for whatever reason – it could be for an audit, complaint from someone who was scanned, etc.)?
  It is more likely that you will to be asked to provide these forms for research considered 'high risk' (i.e. for research which is high profile, high impact or in some way contentious; or where there is a greater likelihood that participants might suffer harm).
- if you had to provide these forms would this be easier to manage if they were in paper versus electronic format?
- what are the risks of retaining records on paper versus electronic format? Paper can take up more physical space than its electronic counterpart. Whilst the potential for things to be destroyed (e.g. fire, flood) or go missing / be accidently deleted are probably fairly similar in each case, at least with electronic format you have the option of making a back-up copy. However, it's important to note that storing electronic records is not without cost; in terms of scanning time, server space and in ensuring that records are maintained / remain accessible over time (i.e. maintaining audit trails / keeping track of back-ups, ensuring security of back-ups, and that electronic files can be read as technology advances, etc.).
- what are the risks of transferring paper records to electronic format – again, this is where a process of checking or validation is absolutely vital (it's no use archiving scanned records if they haven't scanned as intended and you've got incomplete copies).

Regardless of whether data / records are retained in paper or electronic format, there should be a clear and transparent system in place for the management and retrieval of data.

- In terms of management, there is the possibility for scanned records to be amended (either deliberately or accidently). Therefore access to scanned records should be controlled by an appropriate person (e.g. a librarian, Information / Data Manager, Archivist or equivalent) and with appropriate audit systems in place.
- In terms of retrieval, it should be easy for those authorised to access these data, to be able to find the relevant data and/or records that they need (i.e. via meta-data and/or data documentation for electronic format or an inventory for hard copy records). This is particularly important for data sharing.

Where to keep?

Initially research data / records will be collected and retained under the supervision of the Principal Investigator or Trial Steering Committee (i.e. in the Principal Investigator's employing organisation, Clinical Trials Unit(s) and/or at any other site(s) also involved in the research).

At some point a decision may be taken to curate the data / records centrally; if so, your organisation should have a policy or process on when they should be curated centrally and where.

Archiving identifiable information: Usually in multi-site studies participants are told that all identifiable information relating to them will not leave the site at which they were recruited and/or treated, and that non-identifying information will be kept by the coordinating centre (which may be the sponsor). Whatever arrangements are appropriate in a specific study, participants should be informed of these, and with their agreement, the necessary arrangements can be put in place.

Completed consent forms are identifying and therefore are usually kept at each local site. Sites may incur archiving costs as a consequence of storing such documentation. These costs should be identified early and arrangements made as to how they will be covered.

When archiving identifiable information away from the controlled environment in which they were originally collected and collated, the Principal Investigator or the Trial Steering Committee should consider any additional risk. With guidance from the sponsor and local Data Protection Officer, Principal Investigators or the Trial Steering Committee should ensure that only the identifiable information that needs to be kept is kept (i.e. a risk-based decision) and that data security can be maintained as required. This may necessitate additional anonymisation or removal of identifiers from records before long-term storage and the physical separation of consent documentation from health and other sensitive information during storage.

All data: Wherever data / records are kept, their security must be ensured and they must be maintained / stored in such a way that they can be easily identified and located to support future activities (including access for audit purposes, data sharing, etc.). If you are considering storing data / records with an external service provider you must ensure that they can do this.

If you would like further guidance on approved providers of external archiving, please contact the MRC Head Office Records Management Team.

Reviewing what you keep

Long-term storage should not be the end of the story, your organisation should have some policy / process on reviewing what is kept with a view to either further retention or destruction of records as appropriate. Records should only be kept as long as they are still useful (e.g. to support your own or others research in future, to defend any scientific challenge for research which is high profile / high impact or in some way contentious, to investigate any instances of participant harm, etc.). If research data / records are no longer useful, then you should consider their disposal. For further guidance please see the Retention decision-making flowchart in Annex 1.

An Information Asset Register should be kept, listing all key datasets including relevant information such as storage location, time of data collection and type of data (see MRC Hub for template). As part of this there should be an identified person(s) who holds responsibility to review data / records and decide whether these should be retained or destroyed beyond the initial retention period. Any decisions taken relating to a review of data / records should be clearly documented and signed off. i.e. Any further retention or destruction of records should be justified by someone who is familiar with the research data / records (e.g. the Principal Investigator or Trial Steering Committee if still appropriate; or a librarian, Information / Data Manager, Archivist or equivalent) and then signed off by the Unit Director / Head of Department or sponsor representative.

The process of review could be initiated by the impending departure of a member of staff from an organisation (e.g. a PhD student moving on, or a programme leader retiring, etc.). When leaving an organisation, researchers (including students), must leave original copies of research records with the organisation. If records are to be retained beyond the initial retention period then perhaps further decisions will need to be made with respect to where these will be kept, who will take responsibility for these and who will pick up all of the associated costs.

Good Research Practice: Principles and Guidelines recommends that research data relating to studies which directly inform national policymaking should be considered for permanent preservation. In some cases the potential impact on policy may be a clear aim of the study, while in others the significance may only come to light later. It may therefore be necessary to consider the impact of the study at several stages during its lifecycle, particularly for studies which have long-term goals and which may be running for many years. Once decisions have been made with regards to what will be kept and in what format, it is important to review what is held on an ongoing basis. The table below suggests some timelines:

MRC Regulatory Support Centre: Retention framework for research data and records

Retention period	Decision?
Less than 5 years	Retain paper records / data in their original form. Unless doing so renders them illegible, e.g. thermal print outs which fade and degrade over time (such as ECG print outs).
Between 5-20 years	Make decisions with respect to what will be retained and in what format. For example you may wish to consider scanning paper records.
More than 20 years	Consider destruction unless the records / data are extremely valuable e.g. inform national policy making or are from a "risky" participant group (e.g. 'high risk' study involving children, adults without capacity or contentious research outcome).

Data curation and further use
It's important to consider how data will be stored, maintained and accessed to ensure it remains useable and retrievable for future research. The Research Councils UK (RCUK) Concordat on Open Research Data promotes openly available research data for use by others wherever possible.

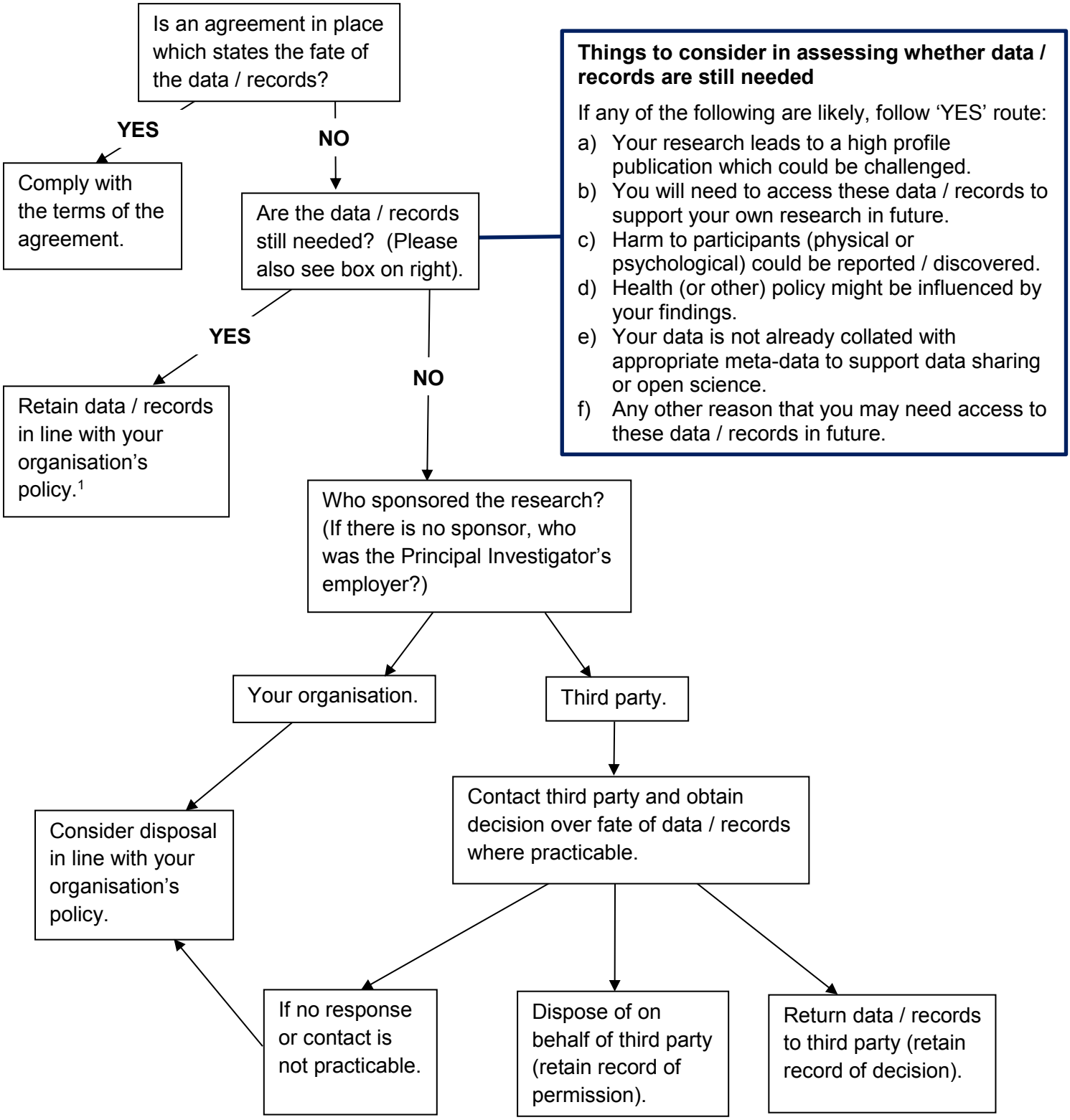
Where will data be kept?
- Will you (the researcher) keep some/all of the data?
- Will it go to a repository?
- If your research will generate clinical information, will the resulting clinical data form part of the patient's medical record?

Who will curate the data?
- Will you (the researcher) maintain the data and any associated meta-data in a useable format?
- Will you have agreements with another organisation to manage the data on your behalf?
- How will access for secondary research or reproducibility reasons be facilitated?

How will access requests be managed?
- Will access be actively promoted, are the data and/or meta-data discoverable by other researchers?
- Are data publicly available or available through a managed/controlled access process? For datasets containing individual participant data MRC supports managed access.
- How are requests for access to data / records reviewed, which criteria are used to make decisions?
- Will you be able to provide an extraction service by those with knowledge of the data and storage systems? Will they help new users understand the data, will it be a collaborative relationship? Long term, how will you retain this expert knowledge of the data?
- In what format will data be sent to the other party? Do you need to ensure secure data transfer?
- How will you oversee usage, e.g. a data sharing agreement with recipients?

For further information on sharing research data please see MRC policy pages; and for sharing individual participant data, please see MRC Methodology Hubs guidance Good practice principles for sharing individual participant data from publicly funded clinical trials.

Annex 1: Retention decision making flowchart for research data / records



[1] Your organisation should have a retention policy similar to the one outlined in the MRC's Good Research Practice: Principles and Guidelines