UKCRC Registered Clinical Trials Units

UKCRC Registered CTU Network – Considerations for Data Retention & Re-use

UK CLINICAL RESEARCH COLLABORATION

igniting our potential

Considerations for Data Retention and Re-use

Authors:

Jaclyn Bailey – Warwick Clinical Trials Unit

Tim Chater – Liverpool Clinical Trials Centre

Andrea Corkhill – Southampton Clinical Trials Unit

William Cragg – Leeds Clinical Trials Unit

Lucy Culliford – Bristol Trials Centre

Carrol Gamble – Liverpool Clinical Trials Centre

Catherine Hewitt – York Trials Unit

Amanda Loban – Sheffield Clinical Trials Research Unit

Sharon Love – MRC Clinical Trials Unit at UCL

Natalie Rowland – Birmingham Clinical Trials Unit

Claire Snowdon – Institute of Cancer Research Clinical Trials & Statistics Unit

John Turgoose – Hull Health Trials Unit

Helen Evans – UKCRC Registered CTU Network

Acknowledgements:

We also acknowledge contributions from former Chairs and Deputy Chairs who have now left the Network: Amanda Bravery, Claire Johnson, and Denise Forshaw.

Contents

Considerations for Data Retention and Re-use 2
1. Background 4
2. Definitions 4
3. Purpose 5
4. Scope 5
5. Data re-use within the original trial team 6
6. Considerations when preparing for data retention and re-use 7
6.1. What data will be retained, and for how long? 7
6.2. Costings 8
6.3. Protocol development 8
6.4. Contracts/agreements 8
6.5. Trials including healthcare systems data or other third-party data 9
6.6. Preparing core datasets in advance of data re-use requests 9
6.7. Staff training and awareness 10
6.8. Participant information 10
6.9. Informed consent 11
6.10. Withdrawal of consent 12
6.11. Data storage and custodianship arrangements post-trial 13
7. At the end of the clinical trial data retention period(s) 13
8. References 15
9. Acronyms 16
10. Other Definitions 16
Appendix 1 – Further guidance regarding process considerations for data re-use within the original trial team 17

1. Background

It is an ethical obligation, and often a requirement of clinical trial funders, publishers and others, that optimal use is made of clinical trial data. This must be done within the confines of the permissions with which it was given or obtained, and available resources.

Clinical trials units (CTUs) therefore need to be adequately prepared to respond to reasonable requests for re-use and sharing of the data beyond the uses specified in individual trial protocols. This includes requests arising from external parties and from within the CTU itself.

2. Definitions

Clinical trial data: all data collected by a CTU required to support the objectives of a trial protocol. This includes data that will be directly used in planned analyses, and data used for broader management of the trial (e.g. data about processes around trial monitoring, or informed consent).

Personal data: as per the UK General Data Protection Regulation (GDPR). The assumption in this document is that clinical trial data will invariably include at least some personal data. Any clinical trial data that is not personal data will not need to be handled in line with the UK GDPR.

Data retention: keeping clinical trial data until the end of the planned archiving period (or longer if justifiable; see section on the end of the data retention period(s), below), maintained in a way that facilitates potential re-use and/or sharing.

Data re-use: sometimes called 'secondary use', this includes any use of the clinical trial data beyond what is required to support the objectives of the original trial protocol. It includes re-use by the original trial team (however it is defined within the CTU) or others. It includes research purposes and other purposes (e.g. providing information to NICE to inform their recommendations or re-using trial data within a CTU for management/operational purposes).

Data sharing: transferring data or granting access to individuals or groups outside of the original trial team (however it is defined within the CTU, see below). Not all data re-use involves data sharing.

Original trial team: it can be relevant to data re-use and sharing processes to know if data is proposed to be re-used by members of the original trial team, or someone else. In this guidance, we do not define 'original trial team' but instead encourage CTUs to decide on their own definition so that they can design their processes accordingly. See "Data re-use within the original trial team", below.

Data custodian: this term is used in this document to refer to the organisation that is 'looking after' any clinical trial data and has primary responsibility for managing access to that data and/or making decisions about re-using or sharing data. This may be the data controller (in the sense used in the UK General Data Protection Regulation [GDPR][1]) or it may be delegated to a data processor (again as defined in the GDPR) or delegated to department level within a data controller (e.g. a CTU).

3. Purpose

This document aims to provide guidance and considerations around data retention and re-use, usually, but not only, following the end of the trial (as defined in the trial protocol). It is intended to complement existing guidance on data archiving[2], data sharing[3-6] and anonymisation[7]. It is aimed at CTUs developing or reviewing procedures in this area, and at trial teams (e.g. trial management, data management, statistics, and others) working through practicalities for their individual trials. CTUs should also consider how they comply with the FAIR Guiding Principles for scientific data management and stewardship[8].

The guidance aims to help UKCRC Registered CTUs:

- Plan and set up trials so that preparation for data retention and re-use is considered from the outset.
- Ensure potential retention and re-use needs are adequately resourced and supported by suitable technical and other infrastructure.
- Ensure plans for data retention and re-use are transparent for participants/data subjects so that their rights can be upheld.
- Understand how changes in consent or other factors might mean data cannot be retained or re-used as planned.
- Ensure there is clarity for all trial staff about what data may be retained and for how long.
- Ensure suitable processes are in place to be able to manage all planned, required and requested data re-uses.
- Decide on a suitable approach and a clear understanding of what constitutes re-use of data by the original trial team and what constitutes data sharing beyond the original team, and the implications of this decision.
- Ensure all applicable legal obligations can be met.

Sponsors, funders, journal editors, regulators and others may have their own policies and requirements around data retention and re-use. CTUs should ensure they adhere to these requirements where applicable. CTUs might also consider involving patients and the public when deciding on their policies in this area.

4. Scope

In scope is any clinical trial data that a CTU is 'data custodian' for (see above for definitions).

Processes around data sharing for secondary research, including managing data sharing requests, assessing data prior to sharing and monitoring post-sharing activities are covered by existing guidance (see references in Purpose, above) and therefore not in scope of this guidance.

This guidance also does not address archiving in a Good Clinical Practice (GCP) sense – see guidance from MHRA[9-10] and data archiving guidance from the UKCRC Registered CTU Network[2] on this topic.

Some trial data might be retained long-term for a specific, planned purpose, such as contacting trial participants to invite them to take part in further research projects. This guidance is focussed on re-use (as defined above) and therefore uses like this that are planned as part of existing trials are not in scope.

5. Data re-use within the original trial team

There may be different considerations for data re-use when it is being undertaken within the original trial team, compared to when it is being shared (as per the definitions above). For example, the number and nature of approvals required prior to re-use may be different, and the rationale for de-identification may differ, if no new data access would need to be granted. It may be appropriate, and more efficient for CTUs, to take a risk-proportionate approach to re-use within the original trial team.

CTUs should take a position on what constitutes re-use within the original team (seeking advice and/or following the instructions of sponsors or others as needed), and whether this will influence their processes around data re-use. Possible positions on this could include:

- All re-use of data could follow the same overall process, regardless of whether data is being shared outside the original team or not.

- Re-use of data by the original team within the CTU can proceed with a reduced, proportionate process (still with appropriate checks, controls and approvals), but not re-use by members of the original team who are not based at the CTU, or members of the CTU who were not in the original trial team.

- Re-use of trial data for non-research purposes can proceed with a reduced, proportionate process (still with appropriate checks, controls and approvals), but not be re-used for research purposes. Non-research purposes might include re-use to understand something about trial processes to inform CTU functioning, without any intention to publish or otherwise share the results beyond the CTU.

CTUs' position on what constitutes 'the original trial team' should specify whether additional staff can be added to the trial team in order to carry out the work (on the basis that staff may routinely be moved between trial teams within a CTU anyway), or whether this strictly means staff who are listed in the protocol, or similar. CTUs might also consider whether it makes a difference if members of the original trial team would be the ones re-using the trial data, but in a different capacity (e.g. a trial statistician who now wants to re-use trial data for their PhD).

CTUs should consider whether to make any specific allowances for cross-CTU, non-research re-use purposes – for example to look at some aspect of trial management across all trials – where otherwise trying to follow the standard 'data sharing' process for all trials involved would be burdensome or infeasible. CTUs should decide exactly what they consider 'research purposes', including for activities such as re-using trial data to prepare new trial grants.

Unlike some re-uses where clinical trial data is shared with a third party, re-use of individual-level data within a CTU is very likely to constitute processing of personal data under the UK

GDPR[1], whether it is within the original trial team or not. The GDPR makes allowance for further processing for research purposes (Article 1(b) on the 'purpose limitation' principle of data protection confirms that 'further processing for...research purposes...shall...not be considered to be incompatible with the initial purposes [of the processing]'). However, CTUs should take necessary advice to ensure their position on defining and managing re-use within the original trial team matches that of any local host institution policies.

Where individual-level clinical trial data will be re-used within a CTU but outside the trial team, it should be de-identified to minimise the disclosure of confidential information[11] and as a matter of good information governance. However, from a regulatory point of view, it cannot be rendered non-personal data to those outside the trial team.[12]

See **Appendix 1** for some additional process considerations regarding data re-use within the original trial team.

6. Considerations when preparing for data retention and re-use

The guidance in this section refers to the trial set-up stage, unless otherwise stated.

6.1. What data will be retained, and for how long?

Trial teams should have a clear understanding of all the data they will collect in the trial, and what the retention/re-use considerations are for each type ('type' here means data collected from the same source that is subject to the same conditions and restrictions). This includes how long each type of data may be retained for and any potential restrictions on re-use or sharing. Existing data protection mechanisms, such as Data Protection Impact Assessments[13] or Record of Processing Activities[14], can help with this scoping task.

Clinical trial data will usually include data obtained from trial sites and may include data from routine data suppliers, from other institutions or directly from participants.

Teams should consider at an early stage whether to create specific datasets for long-term retention or for specific re-use purposes, e.g. for secondary research (see section on preparing core datasets, below). This can include datasets that will be retained indefinitely, if they are confirmed to fall outside of any applicable retention restrictions (for example, because they do not constitute personal data under the UK GDPR, or do not come within the remit of any data sharing agreements).

In establishing the data retention period for each type of data, teams should consider, amongst other things:

- The data protection principle that data is retained in identifiable form only for as long as necessary for the purposes of the data processing.
- Other regulatory requirements, for example those applying to clinical trials of investigational medicinal products.
- Local data retention policies.
- Contractual requirements.

Data retention periods should be clearly and consistently documented in relevant trial documents including the protocol, participant information and the corresponding questions in the trial's IRAS application.

6.2. Costings

At the grant development stage, the cost for the time involved in preparing data for retention and re-use should be factored in. As a minimum, trial teams should consider including costs to prepare core datasets (and any associated data packs (see below)) in case of any data re-use or sharing requests in the future. The possible complexity of the work may need to be considered too, for example if data from various sources will need to be prepared.

For trials where there is reason to expect there may be numerous requests for data with varying or unpredictable requirements (e.g. due to a known level of interest in a trial intervention or research question), more costs should ideally be planned to allow for this. If possible, CTUs should also plan costs for any ad hoc requests that are not directly covered by trial grants.

Things to consider include the costs of data storage or maintenance of records, data processing activity and disposal, and any particular requirements (such as fees for data retention) arising from healthcare systems data or other third-party data (see below).

6.3. Protocol development

The plans for data retention and sharing should be covered in the protocol, including what data will be made available for sharing/re-use and when it would be available i.e. a milestone such as after publication of the primary outcome, rather than including a specific date.

6.4. Contracts/agreements

Ensure all contracts and agreements cover plans for data re-use where applicable, and are clear about the roles and responsibilities, custodianship of the data (including who makes decisions about re-use and sharing), ownership of intellectual property and any limitations on data retention and re-use. Contracts may allow for re-use with data sharing, or just re-use within the original trial team. Aim to understand early during contract development if any data or service providers may have restrictive terms around data re-use and retention (including, for example, providers of data collection tools) so that these issues can either be mitigated through negotiation with the planned supplier or avoided through using different suppliers.

Ensure the requirements and expectations of all relevant external parties regarding data retention and re-use are clear, including trial collaborators and trial-specific sponsors.

When data will be re-used within the same organisation, a legally binding contract is not possible. Instead, trial teams should consider alternative methods for agreeing the terms of the re-use and data sharing in writing.

CTUs should consider how to ensure contracts and agreements will allow for the data storage and custodianship options they might want to use (see more on custodianship post-trial, below).

6.5. Trials including healthcare systems data or other third-party data

When clinical trial data and healthcare systems data (HSD) (e.g. data from NHS England) or other third-party data are combined to make the complete clinical trial data, retention and re-use of data from each source needs to be considered separately.

The relevant contract/agreement governing the terms of data import should be clear about which data can be retained beyond initial trial purposes (e.g. trial analysis), and how long the data will be retained for. Contracts/agreements should also be clear about whether data may be re-used or shared, and any conditions or limitations applying to the re-use and sharing of the data (including re-use within the original team). Where contractual terms are mandated by the data provider, CTUs should have processes in place to ensure they can and will comply with those terms.

Some HSD suppliers may need to be contacted periodically (e.g. every 1-3 years) to allow continued retention of the HSD data. This needs to be considered at trial set up and actioned during the retention period.

6.6. Preparing core datasets in advance of data re-use requests

Following the final data cleaning and analysis as per the protocol, while trial staff are still in post on the trial and if there are funds available to support the work, it is suggested that a set of core datasets be prepared to facilitate future requests for re-use of trial data. Additional datasets could still be developed later, if needed and if funding allows, to meet the specific needs of received requests. However, developing core datasets up front can be beneficial as they are prepared at the optimal time, when resources are still available and before trial staff have moved onto other work. They should be developed to meet the needs of most anticipated re-use requests, therefore reducing the work required to respond to data re-use requests received in future.

The datasets should be suitably prepared so that they would not be identifiable if shared outside of the original trial team (though further data minimisation may often be needed prior to sharing data, in line with the principle of only sharing data required for the purpose specified).

It will usually include datasets relating to the following, with an identifier to link the datasets, and any additional trial-specific data relating to the primary outcome and any other key data (e.g. relating to safety outcomes):

- Baseline data, such as age at randomisation (rather than date of birth, and possibly presented in groups e.g. 20-29 years), sex, gender, ethnicity (with small number suppression), and other trial specific baseline data, randomisation date (as days, with participant 1 being day 0), treatment allocation.

- Primary outcome data – depending on the outcome this could be:

    - Single variables, e.g. blood pressure, or a derived questionnaire score.

    - Individual questionnaire data for each participant.

    - Derived variables, e.g. time to event outcomes based on days from randomisation rather than a date.

- Assessment number and date (in days from randomisation), plus any other trial-specific data relating to the primary outcome.

- Secondary outcome data – if there are a limited number of secondary outcomes, then a decision may be made to prepare datasets for each secondary outcome. Alternatively, CTUs may prefer to focus on key secondary outcomes only, with only raw, unprepared data (or stored data analysis datasets) for other secondary outcomes retained for future potential re-use.

Trial teams might also consider producing datasets for indefinite retention in advance of the end of the overall appliable clinical trial data retention period (see more on the end of the clinical trial data retention period, below). These would be suitably prepared so that, once all other data from that trial is securely destroyed, the CTU could retain just these datasets without it constituting personal data under the UK GDPR. Achieving this effect of holding non-personal data requires more than just de-identifying datasets (e.g. it might be necessary to ensure datasets held by collaborators are also destroyed) but producing these datasets in advance might make the process easier at the end of the data retention period.

Consider developing a 'data pack' along with the minimum datasets. This should include a data dictionary and any other key trial documents necessary for re-using and correctly interpreting the data, such as the trial protocol, annotated data collection materials (e.g. Case Report Forms, or equivalents), the Statistical Analysis Plan and other documents. This data pack could also be useful to retain in perpetuity for CTU records.

6.7. Staff training and awareness

All staff involved in data retention and re-use should be aware of applicable data protection laws and regulations and understand their responsibilities in ensuring compliance with these laws.

Clear data retention policies and standard operating procedures should outline the types of data that should be retained and for how long. Staff should be aware of the procedures for securely storing and disposing of data in accordance with these policies.

Staff should be trained on best practice for securing data such as using strong passwords, encrypting sensitive information, implementing access controls to prevent unauthorised access and handling data breaches.

Consider using training to highlight the process for data re-use within the original trial team (see more on re-use within the original trial team, above), including any differences with the process for data sharing beyond this group.

6.8. Participant information

Participant information should be clear about planned retention and re-use of personal data. It is acceptable to allow some uncertainty around these, i.e. to say data will be retained for 'at least' a period (with some indication about how the decision to stop retaining data will be made), and for the planned re-uses to be broad.

Participant information should be clear that data might be shared with external researchers for further research purposes. It should also be clear that trial data may be re-used within the original trial team. It is helpful for CTUs to decide and document internally what they mean

by 'the original trial team' (see more on re-use within the original trial team, above), but this does not need to be specified in this much detail in the participant information.

Wording of these points could mention the benefits of re-using data in this way, i.e. that collecting trial data can take a lot of effort, and so it is good if we can get the most out of it, for the eventual benefit of patients and the public.

Consider how to write participant information to ensure it will allow for the data storage and custodianship options you might want to use (see more on custodianship arrangements post-trial, below).

See Heath Research Authority guidance on suitable transparency wording for more information.[15]

6.9. Informed consent

Under the common law duty of confidentiality, consent is required for healthcare providers and sites to disclose identifiable health-related information to the CTU (in the absence of some other lawful mechanism for the disclosure).

It is not a requirement to ask for consent to allow planned data re-uses or data sharing, if those do not involve disclosure of confidential information[11] (and it is well-established in the UK that 'consent' is not a suitable basis for processing of personal data for research[16]). However, it is sometimes considered important for ethical or fairness reasons, by trial teams or by other stakeholders (e.g. ethics committees, public contributors). There are therefore several possible approaches to consent for planned data re-use and sharing, as summarised in the table below.

Where re-use of data will involve a new disclosure of confidential information, consent will be required for that to take place (in the absence of another lawful mechanism). There may be additional considerations in trials with more complicated consent arrangements, for example trials where original consent was given by parents/guardians on behalf of child participants who may be of consenting age by the end of the trial.

Consider how to write consent forms to ensure they will allow for the data storage and custodianship options you might want to use (see more on custodianship arrangements post-trial, below).

Consent approach regarding re-use or sharing of data that could not constitute disclosure of confidential information	**Benefits**	**Drawbacks**
Do not ask for trial participants' consent to share or re-use trial data. Instead, make the participant information clear that the processing may take place if they agree to take part in the trial. If they are not happy with this, then they may feel they cannot agree to take part. This may only be fair in cases where data re-use does not involve sharing personal data/confidential information (i.e. if data would be released, but only in such a way that it could not be identifiable in the recipients' possession, in line with ICO	This approach is simpler to manage in terms of consent withdrawal and should result in being able to re-use data for all participants (with some unusual exceptions).	It may affect trial recruitment or means certain groups of people do not take part. It may be less acceptable to ethics committees, sponsors or other stakeholders. Some potential participants might feel it is not a fair approach, regardless of whether or not personal data will be shared.

Consent approach regarding re-use or sharing of data that could not constitute disclosure of confidential information	Benefits	Drawbacks
guidance[17]), and where the re-use is in the public interest.		
Ask for trial participants' consent to share and re-use trial data as a mandatory part of taking part in the trial, i.e. participants have to say yes to it for their consent to be valid.	This approach helps address any stakeholders' concerns about an ethical need to rely on consent for data re-use.	It may not appear too different to the approach above, from participants' point of view – in that the data re-use is bundled up with trial participation. If consent is given, this then implies it can be withdrawn – a difference to the approach above. This can introduce some challenges in establishing which data can be re-used or shared. See below regarding withdrawal of consent.
Ask for trial participants' consent to share and re-use trial data as an optional part of taking part in the trial.	This addresses ethical concerns and may appear as more of a fair choice for participants.	This approach means withdrawal of consent – and opt-outs at initial consent - need to be managed (see below). This will have a resource implication for CTUs. It also means data may not be re-useable for a proportion of participants, meaning the re-use purposes may be limited and/or impaired. This may be exacerbated if participants overestimate the risk to them of data re-uses and decline consent without fully understanding the implications.

6.10. Withdrawal of consent

Where relying on consent to allow data re-use and sharing, withdrawal of consent needs to be considered.

Often in trials, several forms of consent are given at the same time (e.g. consent to take part in research, consent for human tissue sample storage, consent for disclosure of confidential information). When participants say they want to stop taking part, they are perhaps unlikely to define their decision in terms of exactly which of those original consents they wish to withdraw and which may continue. It is right to implement processes to help participants understand their choices and elicit their wishes. However, these processes cannot be burdensome or act as a barrier to participants stopping taking part, and therefore need to be optional, meaning that in some cases it might still not be clear exactly how they want their participation to change.

It is possible to take a conservative approach to this, meaning that consent to data re-use and sharing is considered to have been withdrawn in most or all cases of general withdrawal of consent. This safeguards participants' right to withdraw but may impair the re-use purposes and might not actually be what participants wanted.

Alternatively, it is possible to design processes around a 'presumed ongoing consent' approach, meaning participants' consent to the data re-use and sharing is taken to persist until such a time that they specifically say they want those data re-uses to stop. There are several conditions to meet for this to be fair and transparent: see guidance on this from the CTU Network's PeRSEVERE project[18].

A compromise between these approaches might be to attach withdrawal of this type of consent (consent to data sharing and re-use) to a more general type of withdrawal, e.g. if participants withdraw from further data collection, this could be taken to include withdrawal from data sharing and re-use. Again, this needs to be done fairly and transparently, from participants' point of view.

6.11. Data storage and custodianship arrangements post-trial

Consider where data will be stored long-term after the end of the trial. Options include:

- CTU retains the data and custodianship, including responsibility for managing data re-use and sharing requests and making the final decision about those requests.
- Datasets move to an approved data repository (following funder, host institution or other requirements), with requests for release managed by the repository (internal to a higher education institution, or external).
- Datasets to be passed to a different organisation for custodianship, e.g. trial sponsor.

The best option might be a combination of these approaches, e.g. a period of CTU retaining data then data moved to a repository. CTUs should decide the best approach for each trial, considering how much control they would like to retain over data re-use versus the amount of resource this will require.

7. At the end of the clinical trial data retention period(s)

While it is critical for all datasets (or groups of datasets) to have a documented data retention period, this does not necessarily have to be binding, i.e. there can be justification to retain data for longer in some cases.

CTUs therefore need a process to monitor data retention period end dates, and to decide what to do at the end of each dataset's retention period. Available choices include:

- Securely destroy the entire dataset.
- Retain some of the dataset or modify then retain – for example to render the data no longer identifiable to the host organisation so that it is no longer considered personal data (and might therefore be retained indefinitely). This can be challenging to do for trials where there are many datasets and where identifying data will still be held by other

organisations (e.g. trial sites or biobanks), particularly if this sort of process was not planned for from the outset of the trial.

- Retain the dataset for a further, defined period (and monitor the revised retention date in line with the process above). There should be a clear justification for doing this, and the justification should be documented. If there is some opportunity for people who the data is about to be informed about the updated data retention period (even if just via a public route such as a website) then this should be considered, if feasible.

In some cases, the required action may be dictated by contractual requirements. Other considerations could include:

- The amount of time since any related publications.
- The amount of time since any requests for re-use.
- The presence or absence of staff with knowledge of the data.
- The level of remaining engagement from the Chief Investigator.
- The quality of supporting documentation.
- The suitability of the data format.
- The capacity and funds within the CTU for any required work.
- The requirements and procedures of the data controller(s) and any other relevant organisations.

The regular monitoring of retention period end dates might be done by trial teams or more broadly across portfolios or all trials. It could be done annually, or on an ongoing basis. CTUs may need a process in place to manage very old trials where data retention periods were not clearly specified.

The decision about what action to take should be informed by information about exactly what participants were told about data retention (e.g. whether a strict or flexible timeline for data retention was given) and by information from the trial team about ongoing usefulness of the data for re-use.

There may sometimes also be ethical considerations, for example depending on the type of participant involved (e.g. children, vulnerable adults) or if datasets include particularly sensitive or intrusive data items.

Where it is agreed to destroy data, this should be done in a timely and secure manner and in line with any contractual or other requirements.

8. References

1. https://www.legislation.gov.uk/eur/2016/679/contents
2. UKCRC Registered CTU Network Data and Information Systems Operations Group - Archiving electronic data: points for consideration, available at https://ukcrc-ctu.org.uk/download/10192/?tmstv=1710338262.
3. Tudur Smith, C., Hopkins, C., Sydes, M.R. et al. How should individual participant data (IPD) from publicly funded clinical trials be shared?. BMC Med 13, 298 (2015). https://doi.org/10.1186/s12916-015-0532-z
4. Ohmann C, Banzi R, Canham S, et alSharing and reuse of individual participant data from clinical trials: principles and recommendationsBMJ Open 2017;7:e018647. doi: 10.1136/bmjopen-2017-018647
5. Good Practice Principles for Sharing Individual Participant Data from Publicly Funded Clinical Trials. Tudur Smith C, Hopkins C, Sydes M, Woolfall K, Clarke M, Murray G, Williamson P. April 2015.
6. UKCRC Registered CTU Network Patient Data Sharing Task and Finish Group - Considerations for a participant data sharing SOP, available at https://ukcrc-ctu.org.uk/download/16843/?tmstv=1718902738.
7. Rodriguez A, Tuck C, Dozier MF, et al. Current recommendations/practices for anonymising data from clinical trials in order to make it available for sharing: A scoping review. Clinical Trials. 2022;19(4):452-463. doi:10.1177/17407745221087469
8. FAIR principles, available at https://www.go-fair.org/fair-principles/
9. Good Clinical Practice Guide ("Grey guide"), Medicines and Healthcare products Regulatory Agency
10. 'GXP' Data Integrity Guidance and Definitions, available at https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/687246/MHRA_GxP_data_integrity_guide_March_edited_Final.pdf.
11. Health Research Authority, Confidential patient information and the regulations, available at https://www.hra.nhs.uk/about-us/committees-and-services/confidentiality-advisory-group/confidential-patient-information-and-regulations/.
12. ICO: How do we ensure anonymisation is effective? Can we anonymise within our organisation? Available at https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/data-sharing/anonymisation/how-do-we-ensure-anonymisation-is-effective/#canweanonymise
13. ICO: Data Protection Impact Assessments, available at https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/accountability-and-governance/data-protection-impact-assessments-dpias/.
14. ICO: Records of processing and lawful basis, available at https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/accountability-and-governance/accountability-framework/records-of-processing-and-lawful-basis/
15. Health Research Authority: Transparency wording for all sponsors, available at https://www.hra.nhs.uk/planning-and-improving-research/policies-standards-legislation/data-protection-and-information-governance/gdpr-guidance/templates/transparency-wording-for-all-sponsors/

16. Health Research Authority: Consent in research, available at https://www.hra.nhs.uk/planning-and-improving-research/policies-standards-legislation/data-protection-and-information-governance/gdpr-guidance/what-law-says/consent-research/

17. ICO: How do we ensure anonymisation is effective? Do we need to consider who else may be able to identify people from the data? Available at https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/data-sharing/anonymisation/how-do-we-ensure-anonymisation-is-effective/#considerwhoelse

18. Do we need PeRSEVERE principle O5: continuing data collection, available at https://perseverepriciples.org/principle-o5-continuing-data-collection/

9. Acronyms

CTU	Clinical Trials Unit
GDPR	UK General Data Protection Regulation
GCP	Good Clinical Practice
HSD	Healthcare Systems Data
MHRA	Medicines and Healthcare products Regulatory Agency
NICE	National Institute for Health and Care Excellence
ROPA	Record of Processing Activities

10. Other Definitions

Anonymisation	A process of preparing data so that it no longer allows identification of individuals by any reasonably likely means
Archiving	This is the longer-term storage of data that is not being actively used
Clinical trial	Any research study that prospectively assigns human participants or groups of humans to one or more health-related interventions to evaluate the effects on health outcomes. Interventions include but are not restricted to drugs, cells and other biological products, surgical procedures, radiological procedures, devices, behavioural treatments, process-of-care changes, preventive care, etc.
External/third-party data (e.g. routine data, registry data)	Data primarily collected by a different organisation than a CTU, possibly for a different purpose than research, meaning that a CTU needs to request the data from that other organisation for use in a trial and use the data only in line with a contract with that other organisation

Appendix 1 – Further guidance regarding process considerations for data re-use within the original trial team

Some further questions are listed below for CTUs to consider when designing or reviewing processes for re-use of data within the original trial team.

- Who needs to approve such re-uses before they can go ahead?

- How to ensure some level of independent oversight of the re-use – for example, might there be a role for a Quality Assurance or Data Management team?

- How should planned and actual re-uses be documented within each trial team and at CTU level? (Consider use of existing mechanisms such as data release logs or Records of Processing Activities [ROPA, as defined in the GDPR])

- What approvals might be needed for the re-use? For example, might the nature of the re-use mean ethical review or some other external review is warranted (e.g. if the re-use might be distressing or intrusive in some way)? What approvals might be needed according to local policy?

- Will there be different processes in place depending on whether the re-use is for research or non-research purposes?

- How should teams account for any restrictions in contracts, or in what was told to participants, for each individual re-use proposal?

- Might re-uses for research purposes need to be made public before they begin (for example on Open Science Framework or other public spaces) for the purposes of research transparency?

- Might there be expectations of patient and public involvement in planning re-uses within the original team?

- Will the usual de-identification steps be needed in all cases, or might there be cases where it can be omitted?

- If controls around this sort of re-use are warranted, how can they be implemented reliably, given the people who would re-use data already have access? i.e. how can you ensure trial teams comply?

- How can processes be flexible enough to account for different scenarios?