

Guidance regarding archival/retention of research data associated with Trustworthy Research Environment (TRE) projects

Description of the problem

For reasons of transparency it is typically expected that raw research data will be archived for a specified period of time after a project has ended. The purpose of the GM CHC TRE is to provide a secure research environment in which datasets can be held, processed, and analysed while a project is ongoing. It is not intended to be a solution for long-term storage of data and thus does not have archiving capabilities. After a TRE project has ended, any files within the TRE that are associated with that project will be destroyed.

The TRE has an output checking process in place. This means that the only files that can be taken to a less secure environment, for example for publication, are those approved by the output checkers as meeting a certain level of anonymity. In the majority of cases this would not apply to the raw datasets (otherwise there would be no requirement to use the TRE).

Thus, there is a conflict between the need to archive research data, the inability to archive such data within the TRE, and the inability to extract raw data from the TRE. This document aims to give guidance to TRE users regarding requirements for archiving (what needs to be archived and for how long?) to work around this conflict.

Scope

This document applies only to data held within the TRE.

The document is written around routinely-collected data / electronic health records provided by a third party (e.g. NHS Digital, Salford Royal) for the purposes of a specific research question. The guidance may or may not be relevant to other types of data. For example, there are additional regulations regarding clinical trials that you may need to bear in mind, archival facilities for qualitative data may be different, and you may have agreed to manage data in a certain way in the agreements signed to participate in your study.

Research data retention / archiving requirements

It is considered good practice to archive a research project in such a way that the project can be audited or an analysis can be reproduced. Research institutions, funders, and publishers will typically have guidelines or policies covering retention of data after a project has ended.

Guidelines from The University of Manchester¹ state:

“Unless ethical/professional/local or funding body guidance requires otherwise, research results should be archived in a durable form that is immune to subsequent tampering and falsification for a minimum period of 5 years after the date of any publication which is based upon it. It is recommended good practice that evidence for research based on clinical samples or relating to public health should be retained for 15 to 20 years.”

The MRC guidance² recommends:

- “The retention period for primary/raw data and related material from population health or clinical studies will be informed by the relevant regulatory framework, the legal requirements outlined in guidance from the MHRA and any additional requirements identified by ethics committees or professional codes.
- For clinical research undertaken in MRC research units and institutes, the MRC expects research records relating to such studies to be retained for 20 years after the study has been completed to allow an appropriate follow-up period.
- Studies which propose retention periods beyond 20 years must include valid justification, for example, research data relating to longitudinal studies will often be retained indefinitely and archived and managed accordingly.”

Joint guidance from RCUK³ states:

- “Institutional and project specific data management policies and plans should be in accordance with relevant standards and community best practice. Data with acknowledged long-term value should be preserved and remain accessible and usable for future research.
- To enable research data to be discoverable and effectively re-used by others, sufficient metadata should be recorded and made openly available to enable other researchers to understand the research and re-use potential of the data. Published results should always include information on how to access the supporting data.
- UKRI recognises that there are legal, ethical and commercial constraints on release of research data. To ensure that the research process is not damaged by inappropriate release

¹ <http://www.staffnet.manchester.ac.uk/services/rbess/governance/data-management-and-protection/recording-storing-and-archiving-research-data/>

² <http://www.dt-toolkit.ac.uk/researchscenarios/archiving.cfm>

³ <https://www.ukri.org/funding/information-for-award-holders/data-policy/common-principles-on-data-policy/>

of data, research organisation policies and practices should ensure that these are considered at all stages in the research process.”

Data archiving facilities

The University of Manchester offers a DataVault interface to archive your data and manage it once it is archived. Contact ResearchIT for further information.

External archives are available, for example the UK Data Archive at the University of Essex:

<http://www.data-archive.ac.uk/deposit>

Research data destruction requirements

Data providers that provide datasets for TRE projects (e.g. electronic health records datasets) may have strict data destruction requirements regarding the raw data. It may be possible to request an extension to allow completion of an analysis, but this will not extend to long-term archiving.

e.g.1 NHS Digital require the data they provide to be destroyed when the Data Sharing Contract or Agreement has expired. These can be renewed annually, but require re-approval.

e.g.2 CPRD provide data for one year periods. After a year, the researcher must request an extension. At the end of the project the raw data must be destroyed. This applies to raw data but not “analysis” datasets which contain only the variables stated in the application form.

Recommendations

- 1) First and foremost, researchers must comply with the data retention/destruction requirements of their data providers. Be clear about these requirements before starting the project.
- 2) Funder / Institutional guidance above does not necessarily require archiving of raw data. For the types of dataset concerned it is likely that, in theory, the data provider to extract the same dataset upon request for another party. Unless there is approval from the data provider, datasets that do not meet output checking requirements will not be exported from the TRE. Therefore:
 - a. Focus on archiving metadata about the datasets. Carefully document all of the data processing steps (starting from the instructions to the Data Provider) such that your final results could be reproduced by another party with access to the same raw datasets.
 - b. If processed datasets are suitably aggregated these can be exported from the TRE (subject to output checking).
 - c. Plan archiving from the start of the study to ensure that all of the data processing steps and decisions are documented sufficiently
- 3) Once data have been extracted from the TRE then they can be archived in the same manner as any standard research project. Your research institute may be able to offer guidance. One option, provided you have the agreement of your data provider (or, in the case of directly-

collected data, from participants), is to publish aggregated data or data processing information in a repository such as Mendeley Data.

- a. If you are archiving personal data to a repository that is outside the University of Manchester this constitutes processing of data by another legal entity. You need to ensure that you comply with the General Data Protection Legislation, for example have a data processing contract in place and notify data subjects of the transfer.

Bear in mind that a TRE project will be destroyed when the study is completed. It is likely that you will need to export various files before this happens. Plan this in advance and give the TRE team plenty of notice so this can be conducted in a timely manner.

Version History:

Date	Author	Version	Description
2018-07-02	RJ	v0.1	File created
2018-07-31	EG	V0.2	Edits to text
2018-08-02	RJ	v0.3	Add references for guidance
2018-09-07	EG	V0.4	Clean version for circulation