Administrative Data Research Network

An ESRC Data Investment

Data Retention and Destruction Policy

PUBLIC

01 JULY 2015

Version: 01.00

T +44 (0)1206 873435
E help@adrn.ac.uk
www.adrn.uk

Scope

To support ADRN's commitment to protect the privacy and confidentiality of data subjects a suite of Information Assurance policy documents have been developed to provide a framework for ADRN operations and to guide practice across all ADRCs and the ADS. This document outlines the ADRN policy in relation to data retention and destruction. All ADRCs will in addition have local procedural documents that outline how these policies are operationalised.

Other associated documents include:

ADRN030 Information Security Principles

ADRN031 Trusted Third Party Policy

ADRN032 Secure Environments Policy

ADRN033 Data Movement Policy

Contents

For defintions of terms please see the ADRN glossary at http://www.adrn.ac.uk/using-the-network/documentation

1. Data Retention and Destruction Policy

Data of varying types are created and processed during the research process; Personal Identifiable data (e.g. names, addresses, dates of birth) are used within the Trusted Third Parties for the purpose of matching individuals across datasets and facilitating the creation of de-identified research datasets. The research datasets accessed by Accredited Researchers will not contain person identifiable information. Initially all linked research data will be destroyed at the end of a research project. However, the ADRN recognise that there are potential research benefits and efficiency savings that could accrue from the retention of de-identified research datasets. Therefore this policy will undergo regular review taking into account the views of the public and data controllers, and research needs.

This policy outlines data retention and disposal requirements at different stages of the data journey. The policy also covers data generated by Data Controllers; however this should be seen as advisory only.

A. ADRN

These first are overarching policies that apply to all data and information held by ADRN and partners.

1. When deleting or destroying any data or information ADRN, TTPs, researchers and other partners must ensure that it is done securely and effectively, and in line with CESG or data controller requirements.

2. Local ADRCs, TTPs and ADS must have clear policies that ensure that data are deleted properly and that physical media are destroyed securely (whenever necessary).

B. Trusted Third Party (TTP)

Information held by TTPs and covered by this policy includes, but is not restricted to: the direct identifiers received from data owners; syntax or program configurations used to create matches and project specific index numbers; metadata related to matching rates (to be sent to ADRC); sets of index numbers transferred to the ADRC, and any data returned to the data controllers.

1. TTPs must securely destroy all data on completion of the linkage process.

2. For the lifetime of the ADRN, TTPs will retain metadata on the linkage method, quality and where applicable matching statistics. This will also be supplied to the ADRC, and the ADS who will retain this information.

C. Staging environment for the creation of linked research data

All data linking using the indices created by TTPs takes place in a secure staging environment within an ADRC. The staging area is separate from the safe setting within which the researcher accesses the final research dataset.

Information created within a secure staging area that are covered by this policy includes, but is not restricted to: a copy of the attribute data created by the data controllers; the project specific matched index numbers supplied by the TTP; derived variables created by the ADRC; metadata.

1. Metadata will be retained by the ADRC for the lifetime of the ADRN.

2. Researchers will only receive access to data they need as specified in the application. Any data that has been created as part of the linking process that is not specified in the project application will be destroyed before the data are transferred to the safe setting for research access.

3. Once the research dataset has been created and deemed ready for access all the data within the staging environment will be securely destroyed.

D. Research data

Research data covered by this policy include but are not restricted to: the de-identified research dataset accessed by the Accredited Researcher; any additional copies of research data created by the researcher; processed versions of the data; analyses or statistics that have been created but not approved for release; and syntax (analytic code that has been created in statistical packages such as SPSS, SAS, R, STATA).

1. The research data will be archived by the ADRC for a maximum of 5 years unless the end date stated in the project application is before that or unless the Accredited Researcher makes a valid application for an extension (that is approved by the data controllers), or the Data Controller specifies otherwise.

2. Following the active analysis phase of the project data will be archived for the remainder of the five year period and only made available on receipt of a valid application (via the approvals panel) from the researcher (for example, to enable them to make revisions to a paper to have it accepted for publication).

3. Data controllers will be informed that after the archiving phase the data may still be retained on copies of IT backups. The length of time that IT back-ups will be retained will be described in local ADRC procedure documents.

4. The syntax generated by the research team will be retained by the ADRC for audit purposes for as long as the data are archived (see above). Syntax may also be retained by the Administrative Data Service for future use.

E. Approved statistical results and analytic outputs.

This policy covers all research outputs that have undergone statistical disclosure control and been approved for release into the public domain.

1. Statistical results and analytic outputs will be archived for as long as the ADRN (or similar) is in existence.

F. IT back-ups

This policy covers: all information on the TTP network, staging environment and the ADRC safe environment.

1. The maximum time a backup may be retained in a TTP, an ADRC staging environment must be 48 hours.

2. All other areas data back-ups will be determined by local procedures.

G. Data extracts created by the data controller

This recommendation includes, but is not restricted to: syntax used to create the dataset, data containing the personal identifiers sent to the TTP, and attribute data sent to the ADRC/Secure Environment. Retention decisions for these data are the responsibility of the data controller, however:

1. The ADRN encourages data controllers to retain any code/syntax used to extract data for ADRN projects, ideally for the life time of ADRN.

2. The ADRN encourages data controllers to retain copies of the direct identifiers and the attribute data until the completion of the analysis phase of the project.

3. Where possible, the ADRN encourages data controllers to retain copies of the extracts that could be used to replicate the specific research dataset in the future or re-use the data for a different project.