[www.farrinstitute.org](http://www.farrinstitute.org) [www.herc.ac.uk](http://www.herc.ac.uk)

# HeRC Trustworthy Research Environment (TRE) – Information for data owners

# Document History

|  |  |  |  |
| --- | --- | --- | --- |
| **Editor(s)** | **Date** | **Summary** | **Version** |
| Niels Peek | 19.11.2015 | Approved | 1.2 |
| Emily Griffiths | 06.12.2017 | Redrafted | 2.0 |
| Ben Green | 15.12.2017 | Revision | 2.1 |

1. Introduction

The HeRC Trustworthy Research Environment (TRE) provides a secure, controlled facility certified to ISO27001 to enable researchers to access health data while protecting confidentiality of staff and patients. The HeRC TRE does this by providing a range of services for individual researchers, research projects and programmes of research to manage and analyse data within the TRE. The TRE seeks to minimise the risks of data leakage by adopting the latest security mechanisms, restricting the importing and exporting of data from the TRE, and the extent of data processing.

Data owners who decide to share their data through the HeRC TRE get a transparent, controlled process for managing data access requests, supported by a meta-data catalogue and services for data provision and archiving. Added advantages are oversight of the users and uses of their data, retaining control over access, and ensuring a collective knowledge of their dataset and resultant efficiency through the analysis platform.

1. Overview

The HeRC TRE provides a data analytics facility for personal healthcare data that may be any combination of sensitive, personal and identifiable. Researchers are granted access through standardised and transparent procedures whilst minimising the proliferation of datasets and the frequency of data transfer. In addition, the TRE aims to retain user productivity and confidence by providing a familiar desktop environment for data analysis.

Figure 1 gives an overview of the HeRC TRE architecture. The TRE enables participation with NHS organisations by providing data transfers from the NHS N3 network. The HeRC TRE is also certified to IG Toolkit Level 2. Authorised users have remote access to the TRE. Data within the TRE is controlled by:

1. Prohibiting users from importing data or taking copies of data out;
2. Giving users access to only the data they are permitted to access;
3. Knowing the location of every copy of every dataset in the TRE;
4. Knowing who has been granted access to each copy of each dataset in the TRE;
5. Knowing the purpose for which each access right has been granted;
6. Having the means to instantly remove any access right from any user.

It is the responsibility of data owners to provide sufficient information about their datasets to support data curation and the TRE meta-data catalogue; to define the access rules for datasets that they have provided, and to give final approval to access requests and requests to export datasets from the TRE.



Fig. 1. HeRC TRE architecture

The TRE is physically located at the University of Manchester Oxford Road Campus, in Vaughan House, Portsmouth Street, Manchester, M13 9GB.

1. Data Import and Lifecycle

The HeRC TRE service team is committed to providing a solution for data owners who need assurance that their data will be safely transferred and cared for throughout its lifecycle within the TRE.

The HeRC TRE operates a robust data import process where data is transferred directly into the TRE using the optimum security mechanisms and curatorial processes. The integrity of the source data is monitored continually during transfer and throughout its lifecycle within the TRE. This means at any time during the project, the researcher will always have access to the exact same dataset originally provided by the data owner.

At the end of a project, or at any time during the lifecycle of a dataset hosted within the TRE, the data owner is entitled to request that their data is removed from the TRE and formally destroyed. The TRE service is able to provide evidence of data destruction if required.

If the data owner wishes their data to the retained within the TRE so that it can be made available to other projects in the future, the TRE service is able to provide long-term storage and data archive resources, and can make the data discoverable via the TRE data catalogue. The TRE service can also help make the datasets it hosts citable in publications and other bibliographic references.

1. Access procedure

Access to data that is stored in the HeRC TRE can proceed in two ways depending on whether the data are a new extract for a dedicated research project, or re-use of a previous project’s data:

New extracts

The researcher arranges research ethics and other project documentation to apply for the data from the relevant provider. If approved the researcher then completes a TRE Project Application Form.

Data re-use

The researcher browses the online meta-data catalogue of the TRE on the publicly accessible HeRC website to see which datasets are available and which items they hold, and selects the dataset(s) of interest and any relevant code for reproduction of analyses. The researcher completes a TRE Project Application Form.

After each application

1. The TRE Project Board meet every two weeks and review all applications. Approval criteria may differ between datasets and will be decided by the data owner for their dataset.
2. If approved, a service manager of the TRE will verify the identity of the researcher, and ask the researcher to complete a training course and sign a confidentiality agreement. The researcher then obtains authorisation to access the dataset in their dedicated work environment in the TRE.
3. The researcher gets access, through the HeRC TRE and for a specified period of time, to a virtual machine that contains the requested dataset as well as software tools needed to conduct the analysis. On request research outputs can be exported from the TRE after consulting the data owner and scrutiny of a trained output checker who ensures no disclosive information are included. Outputs are then emailed to the researcher for distribution as needed, for example for publication, conference presentation or submission to a regulator or study sponsor.
4. After completion of the data analysis and publication of the results, a backup of the virtual machine is created and access is discontinued.