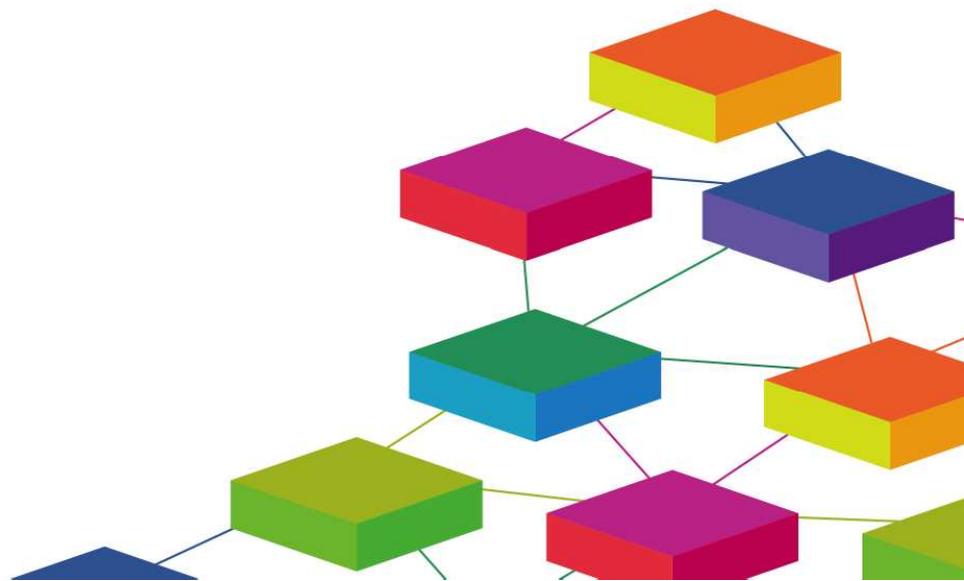


CONSUMER DATA RESEARCH CENTRE DATA SERVICE USER GUIDE

Version: 5.0



Introduction

The Consumer Data Research Centre (CDRC or Centre) is an academic led, multi-institution laboratory which discovers, mines, analyses and synthesises consumer-related datasets from around the UK. The CDRC forms part of the ESRC-funded Big Data network and offers a data service aimed at providing researchers with access to a wide range of consumer data to address many societal challenges. CDRC's key areas of interest include retail, transport, health, crime, housing, energy, mobility and sustainable consumption. We support the acquisition and analysis of data in these areas and others to achieve benefits for the 'public good'.

The purpose of this guide is to describe the Centre's data services and how researchers can access them. It identifies the different types of data the Centre holds and the service tiers through which these data sets are available. For data that are not publicly available, the guide details how researchers can register or apply for access and the kinds of support that is available to them.

CDRC Data Services

The CDRC provides data with three different levels of access. These correspond to the data levels described in the UK Data Service's three tier access policy:

- *Open data*: data which are freely available to all for any purpose – examples might include geodemographic data derived from the Census. Open data are accessed through the CDRC service via basic registration and download.
- *Safeguarded data*: data to which access is restricted due to licence conditions, but where data are not considered 'personally-identifiable' or otherwise sensitive – an example might include data from retail companies on store turnover. Access to safeguarded CDRC data is via a remote service that requires users to submit a project proposal. This proposal must receive approval from the Centre's Research Approvals Group (RAG) (see below) before access to the data will be authorised. Users are able to retrieve data after authentication and authorisation by the service.
- *Controlled data*: data which need to be held under the most secure conditions with more stringent access restrictions, including data which are 'personally-identifiable' and therefore subject to Data Protection legislation or are considered commercially sensitive. Examples might include data on individual consumer purchases. Access to CDRC controlled data is provided through the CDRC-secure service. This service requires that individuals gain project approval through the RAG and visit one of our secure facilities at either the University College London, University of Leeds or University of Liverpool.

Finding Data

All data available through the CDRC are accompanied by metadata that enable both attributes and geography to be searched.

Research Approvals Process

Access to both safeguarded and secure data requires a process by which individuals submit project proposals for assessment and approval. The approval process is overseen by the Research Approvals Group (RAG) which comprises archive representation from the UK Data Service (UKDS), an independent academic representative, the Chair of the CDRC Advisory Board, a CDRC data scientist, plus a data partner representative. This composition ensures that the RAG has expertise in research design and analysis, in research ethics and legal issues and commercial sensitivity as well as having representatives of policy interests.

Decisions by the RAG will be based upon publicly available evaluation criteria (see RAG Terms of Reference www.cdrc.ac.uk/data-services/using-our-data/).

Criteria for Approval

The RAG will develop its assessment criteria in conjunction with CDRC staff. These criteria will align with CDRC objectives and cover the following:

- **Scientific advancement** – how the project has the potential to advance scientific knowledge, understanding and/or methods using consumer data;
- **Public good** – how the project has the potential to provide insight and/or solutions that could benefit society;
- **Privacy and ethics** – the potential privacy impacts or risks, and wider ethical considerations relating to the project
- **Project Design and Methods** – how the project will be conducted and who will be involved with a focus on demonstrating project feasibility.
- **Cost and resources issues** – what impact the project is likely to have on CDRC resources, including CDRC staff time and use of infrastructure, as well as any data acquisition costs. Resource requirements should be justified.

The RAG typically considers applications remotely and is designed to be lightweight but robust, enabling timely decisions on user applications.

Approval will not be granted without evidence that the user has acquired ethical approval for the research through their institution, or supplied evidence that it is not applicable. For non-academic projects, where there is no approval process in place the CDRC will assist the user with acquiring this.

Safe Researcher Training and Training and Development

Safe Research Training

Users, both academic and non-academic stakeholders, wishing to access safe-guarded and/or controlled data are required to have completed a national accredited safe researcher training course within the last two years, as offered by the Administrative Data Research Network (ADRN), HM Revenue and Customs (HMRC), Office for National Statistics (ONS) or the UK Data Service (UKDS), prior to submission of a project proposal. Evidence of completing this accredited training will be required. If the user has not previously completed such training the CDRC will offer access to training courses.

Training and Development

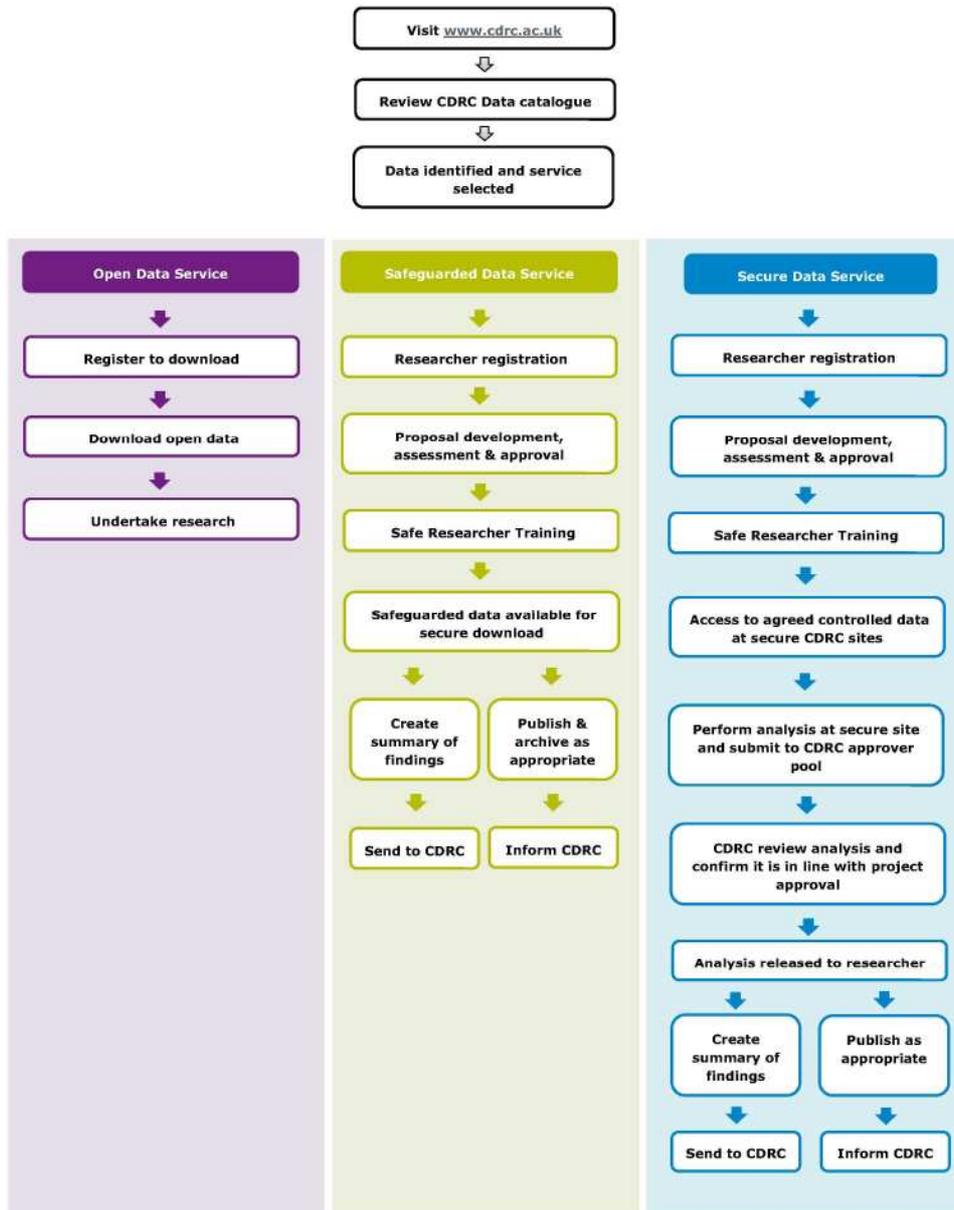
In addition to providing data services, the CDRC has a range of training courses and materials available. Many of these will be of benefit to those who wish to use our facilities, as they are aimed at enhancing capacity in data analytics and data visualisation methods. Full details of the training available can be found at www.cdrc.ac.uk/training-capacity-building/. Our programme includes training in the following areas:

- Working on Big Data: introductory courses that explain the growing importance of Big Data; the importance of analytics and protocols; and standards for data management.
- Introductory and advanced courses in data analysis and visualisation, including courses in R.
- Introductory and advanced courses in Geographical Information Systems, including ArcGIS and Q-GIS.
- Advanced courses in microsimulation and geo-temporal demographics.
- Courses on how insights from Big Data analytics can enhance business.
- Visualisation.

Charges for CDRC Services

While a service will be provided to the academic community and stakeholders free of charge, researchers may need to apply for funding to cover the costs of additional data acquisitions, or be charged for access to certain, licensed software.

CDRC Services Overview and User Journey



CDRC Website: A Single Point of Entry into the CDRC Data Services

The CDRC website, www.cdrc.ac.uk, is designed to provide a single point of entry into our services and these are clearly linked from the homepage.

CDRC Data

CDRC Data is a remote service that enables the dissemination of data, both open and safeguarded. Open data will include non-sensitive and aggregated data and derivative products produced by the CDRC. Safeguarded data will include deposits that have some sensitivity/disclosure implications, albeit not to the extent that access merits a restricted secure access environment.

All users of this service are required to register, to enable monitoring of resource use. For those data that are safeguarded, users will also have to have completed safe researcher training (see the section Safe Researcher Training) prior to submitting a project proposal. Proposed projects will need to go through the research approvals process where an independent Research Approvals Group (RAG) will evaluate applications for access to these data. Following approval, the user will be emailed details on how to download the encrypted data file, and will then be required to contact the Centre for a password to enable unzipping and access to the data.

Accessing data from CDRC Data <https://data.cdrc.ac.uk/>

Open Data:

Access to these data requires:

1) Registration

Users will be required to provide contact details including a valid email address prior to download. This is to enable the CDRC to monitor the use of the resource. Data will then be available to the user to download for unrestricted use.

Safeguarded Data:

Access to this service requires that users obtain formal approval.

1) Initial Proposal

An approach is made to the CDRC by the user through completion of an online form, www.cdrc.ac.uk/data-services/using-our-data/. This initial proposal is processed and assessed by a CDRC data scientist to see if it fits within the remit of the Centre. If not, the proposal may be referred to another Centre in the Big Data Network. Proposals that do not fit into either of these categories will be turned down at this stage.

2) Proposal Development

If the initial proposal fits within the Centre's remit, the user is supplied with a full proposal document, where they are also assigned to a CDRC data scientist who can advise on the technical aspects of the formal application. The aim is to co-produce an acceptable project proposal. Proposals will comprise:

- a) The Project scope
 - i) Motivation and purpose

- ii) Public perceptions and research ethics
 - iii) Data required
 - iv) Outputs planned and timings
 - v) Impact Plan
- b) Timing
- i) The number of days the applicant intends to access the downloaded data
- c) End uses of the analytical work
- d) Ethical approval from user's institution¹
- e) Safe researcher accreditation. The user will be required to have completed accredited safe researcher training within the last two years and provide proof of accreditation, see **Safe Researcher Training** section.

3) RAG Assessment and Approval

Once an application has been co-produced it is considered by the RAG against agreed criteria that are published on our website, www.cdrc.ac.uk/data-services/using-our-data/. The number of rejected approvals will be minimised through initial interaction with the data scientists. Where approval is withheld, applications are referred back to the user for revision, and clear guidance will be given regarding those areas requiring clarity. If such amendments are agreeable by RAG, approval will be given. If minor, the user may be asked to make further revisions, however, if issues are still considered to be major the RAG may decide to make a final decision to reject the proposal. Once the project is approved the user and their institution will be required to sign a CDRC User Agreement and will be contractually bound by the licence for the data and the project that is undertaken.

4) Data Access

Following approval, the user is required to agree to the CDRC User Agreement, including those from the relevant data licence. Access to a secure download of the agreed data is then made available. This process requires that users telephone the CDRC to obtain a further password to unlock the encrypted download files. Once the user has downloaded the encrypted file, they are solely responsible for the data and its analysis.

5) Outputs

Users can use results of their analyses in publications, reports and presentations provided they abide by the terms and conditions with particular reference to the data partner publication terms. There is no screening of outputs by CDRC staff.

6) Completion, Reporting and Acknowledgement

Users are required to deposit copies of working papers, peer-reviewed journal articles, logs of impact and other publications for access with the CDRC site wherever copyright permits. Where this is not possible, full references to research outputs are required for CDRC audit purposes. Please email publications@cdrc.ac.uk when publications are ready for deposit or logging. The commitment to produce specified outputs is normally a condition of the data approval process. The terms of service require that published outputs include an acknowledgement stating: "The

¹ If the user's institution does not have a system for data protection and ethics approval then the CDRC host institution will provide access to ethical approval procedures if required.

data for this research have been provided by the Consumer Data Research Centre, an ESRC Data Investment, under project ID xxx.xx (reference to web address of the project outputs and reports)". The acknowledgement will make further reference to the use of specific datasets according to the wishes and needs of individual data partners. After the project end date is reached, the CDRC will contact the user to confirm the destruction of the data and to document any outputs to date. The CDRC will contact users normally at 6 and 12 months after the project end date to request a log of any further publications or impact logs.

CDRC Secure Data

Access to CDRC controlled data is via our Secure Data service at one of three secure facilities located at University College London, the University of Liverpool and the University of Leeds.

The CDRC Secure Data service enables access to data that are considered sensitive or confidential by data providers, or that pertain to anonymised attributes judged to raise potential disclosure issues through linkage. These data are acquired by the CDRC through licenced agreements which limit access to the data for approved projects and by approved users where there is a public interest case to do so. Access is organised in such a way that individual confidentiality is protected and risks of disclosure are minimised as much as possible at all times.

Independent analysis of secure data can be undertaken at all of our secure facilities. If users require bespoke guidance and support with analytics, this service is provided at the University of Leeds only.

Use of the CDRC-Secure service requires registration and project approval, with an additional step of booking into one of the secure facilities and meeting any site specific secure facility requirements. The user will be informed of these once the site to be visited has been selected.

Accessing data from CDRC secure sites

Access to this service requires that users obtain formal approval.

1) Initial Proposal

An approach is made to the CDRC by the user through completion of an online form, www.cdrc.ac.uk/data-services/using-our-data/. This initial proposal is processed and assessed by a CDRC data scientist to see if it fits within the remit of the Centre. If not the proposal may be referred to another Centre in the Big Data Network. Proposals that do not fit into either of these categories will be turned down at this stage.

2) Proposal Development

If the initial proposal fits within the Centre's remit, the user is supplied with a full proposal document, where they are also assigned to a CDRC data scientist who can advise on the technical aspects of the formal application. The aim is to co-produce an acceptable project proposal. Proposals will comprise:

- a) The Project scope
 - i) Motivation and purpose
 - ii) Public perceptions and research ethics

- iii) Outputs planned and timings
- iv) Impact Plan
- b) Timing
 - i) The number of days the applicant intends to access the downloaded data
- c) End uses of the analytical work
- d) Ethical approval from user's institution²
- e) Safe researcher accreditation. The user will be required to have completed accredited safe researcher training within the last two years and provide proof of accreditation, see **Safe Researcher Training** section.

3) RAG Assessment and Approval

Once an application has been co-produced it is considered by the RAG against agreed criteria that are published on our website www.cdrc.ac.uk/data-services/using-our-data/. The number of rejected approvals will be minimised through initial interaction with the data scientists. Where approval is withheld, applications are referred back to the user for revision, and clear guidance will be given regarding those areas requiring clarity. If such amendments are agreeable by RAG, approval will be given. If minor the user may be asked to make further revisions, however if issues are still considered to be major the RAG may decide to make a final decision to reject the proposal. Once the project is approved the user and their institution will be required to sign a CDRC User Agreement and will be contractually bound by the licence for the data and the project that is undertaken.

4) Data Access

Following approval, the allocated CDRC data scientist arranges access for the registered user. Dates are booked to use the secure facility at either UCL, University of Liverpool or University of Leeds. Users will receive a document informing them of site specific secure facility requirements and instructions of use.

5) Data Analysis

The user works on the data only within the secure environment. If users wish to combine controlled data with other less sensitive data (open or safeguarded), then it will be necessary to have obtained consent for this from RAG as part of the project proposal. This supporting data will then be made available to the user in the secure facility. The same applies to software required for analysis. CDRC staff provide limited support through the advanced analytics service. At the University of Leeds, a supported analytics service is available which provides the user with bespoke guidance and support in both accessing and analysing data.

6) Outputs

All outputs that the user wants to take out of the secure environment must be vetted and cleared by the CDRC before they can be released. Source data do not leave the secure facility. Users can take results of their analyses for use in publications, reports and presentations provided they abide by the terms of the User Agreement and with particular reference to the data partner publication terms.

² If the user's institution does not have a system for data protection and ethics approval then the CDRC host institution will provide advice on acquiring this.

After completion of analysis the user informs the data scientist that the analysis is complete and that their files are now ready for vetting. For full details of the output process please see the CDRC site specific Import/Export Guidelines.

- a) Outputs will be checked by two CDRC data scientists to ensure that they conform to CDRC control criteria.
 - i. Outputs requested should be 'finished outputs' i.e. the finished statistical analyses that you intend to present to the public, must be easy to read and interpret and how they are to be used explained and must be non-disclosive.
 - ii. The CDRC team will ensure that the outputs are the same specification as those agreed in the approved project proposal.
 - iii. The user is informed of the outputs vetting outcome within 5 working days and if successful with details about how the data extracts or analysis will be returned to them.
 - iv. Extracts that match approval are transferred by the CDRC team to a secure server from where outputs can be downloaded under the same arrangements as safeguarded data or transferred to the user on an encrypted hard drive.
 - v. Where extracts are deemed not to match the required criteria, the user is informed.
 - i. Where there are issues with a part of the output, if feasible the user will be allowed to revisit the secure facility to rectify the problem.
 - ii. Major transgressions may be permanently deleted and the remaining output is returned to the CDRC approver pool.
 - vi. Once the user has completed all their analysis or their agreed lab access time has been reached all passes or electronic fobs are returned and access to the secure facility is immediately revoked.

- b) If the user agrees and the licence agreements allow then approved user generated content will be stored securely for 12 months and then archived at the UK Data Archive. If the user and/or the licence agreement does not allow this then the data will be stored securely for 12 months after which time it is destroyed.

7) Completion, Reporting and Acknowledgement

Users are required to deposit copies of working papers, peer-reviewed journal articles, logs of impact and other publications for access with the CDRC site wherever copyright permits. Where this is not possible, full references to research outputs are required for CDRC audit purposes. Please email publications@cdrc.ac.uk when publications are ready for deposit or logging. The commitment to produce specified outputs is normally a condition of the data approval process. The terms of service require that published outputs include an acknowledgement stating "The data for this research have been provided by the Consumer Data Research Centre, an ESRC Data Investment, under project ID xxx.xx (reference to web address of the project outputs and reports)". The acknowledgement will make further reference to the use of specific datasets according to the wishes and needs of individual data partners. After the project end date is reached, the CDRC will contact the user to confirm the destruction of the data and to document any outputs to date. The CDRC will contact users normally at 6 and 12 months after the project end date to request a log of any further publications or impact logs.