



Urban  
Big  
Data  
Centre

An ESRC Data  
Investment



# Controlled Data Service Guide for Researchers

---

Version 5.1  
01 February 2017

---



## *Who we are*

The Urban Big Data Centre (UBDC) is a national data service, funded by the Economic and Social Research Council (ESRC) to support data owners, policymakers, researchers and everyday citizens in extracting useful information from urban-related data. We help harness the potential of big data to develop solutions for environmentally sustainable, economically resilient, and socially just cities.

UBDC offers three data services: our Data Collections, Data Sourcing Service and Controlled Data Service. This guide will provide users with information about the Controlled Data Service. Please see *Data Services: Guide for Researchers* for an overview of our services (<http://ubdc.ac.uk/data-services>). This guide describes the Controlled Data Service, available to researchers in the UK, including from academic, public and third sector organisations.

## *Personal data and the Controlled Data Service*

The aim of the Controlled Data Service is to make it easier to conduct research on urban issues using personal data held within administrative systems. We provide secure computing facilities, technical advice on analysis, and knowledge about what data is available and how to go about getting the necessary permissions to make use of it.

The term 'personal data' refers to any data that may identify an individual living person, or which could do so when combined with other information. This could be personal data from administrative datasets (e.g. welfare benefits or health records), data on personal movements (e.g. from mobile phones), or information on personally owned assets (e.g. car registration details). Organisations which hold personal data are governed by the Data Protection Act of 1998 (DPA) and are not generally permitted to sell, share or pass on this data without the permission of the individuals concerned. However, in some circumstances the controllers of personal data can agree to permit others to have access to personal data without consent. Such access is tightly controlled. It is limited to approved uses by approved users and must be organised so that individual confidentiality is protected and risks of disclosure are minimised at all times.

UBDC's Controlled Data Service aims to provide a suitably secure computing environment that makes such data sharing for research purposes possible. UBDC uses an established and highly respected provider, electronic Data Research and Innovation Service (eDRIS – part of NHS National Services Scotland), to provide this service.

## *Accessing the Controlled Data Service*

To access these controlled data, researchers need to go through an approvals process (see Table 1 below) that begins with approaching UBDC via the online research enquiry form: <http://ubdc.ac.uk/data-services/data-services/access-our-services/>. If the project falls within the UBDC's remit and is potentially feasible, we will work with the researcher to develop the project and complete the project proposal form. Together with the researchers, UBDC will also start liaising with the data owners about data availability and potential access. At this stage, researchers should also start preparing for ethical approval from their institution's ethics panel. A researcher that doesn't have access to an ethics board or panel, should let UBDC know so we can explore options with them.

**Table 1: Controlled Data Service applications process**

Process stage	Roles	Actions
<b>Project development</b>	<b>Researcher role:</b>	Online research enquiry form Project development through the project proposal form Liaise with data owners on data availability and access
	<b>UBDC role:</b>	Consult on project proposal form Liaise with data owners on data availability and access UBDC Appraisal Report Consult with eDRIS on project aims and feasibility
<b>Project approval</b>		RAC
<b>Formal agreements and Research</b>	<b>Researcher role:</b>	Redrafting of project (if necessary) Applications to data owners (if necessary) Apply for ethics approval Approved researcher training Data processing agreement with eDRIS Data processing agreement(s) with data owners Privacy Impact Assessment
	<b>UBDC role:</b>	Assist on applications to data owners (if necessary) Consult with eDRIS on applications to data owners and/or research approval panels Assist researcher (if needed) Provide access to eDRIS services Monitor service use with eDRIS
<b>Post-research</b>	<b>Researcher role:</b>	Report on publications and other research output
	<b>UBDC role:</b>	Advertise publications and research output, track and report on impact

Once complete, the project proposal form is submitted to the UBDC's independent Research Approvals Committee (RAC) along with an internal appraisal form from UBDC that details current capacity to support the project. RAC makes the decision on whether to permit use of UBDC facilities for the project. Projects are assessed on academic merit, potential for public benefit and other criteria, as well as researchers' experience and competence for undertaking the work. Projects by postgraduate students will be considered, but supervisors should be listed on the project as lead researchers.

For the research to proceed, individual data owners need to give their approval for the project as well. This may mean that additional applications have to be made to meet data owners' requirements - although the information requested should be very similar to that sought by UBDC and UBDC staff can assist and advise in this process. Projects using controlled data frequently seek to link data from different sources (e.g. health data with that from social care), so several data controllers may be involved. Linkage will be conducted in a way that separates personal identifiers (names, addresses) as early as possible, so that researchers never have access to these.

Once both the RAC and the data owner approve the project, the researcher's institution and the data owner will need to sign a data sharing agreement. This may also involve the completion of a privacy impact assessment (PIA) to outline the potential risks to individual privacy and how these risks are mitigated. Data sharing agreements may effectively make the researcher's institution a data controller for the data, with legal responsibilities under the DPA. Different data owners may have slightly

different requirements. A data processing agreement between the researcher's institution and eDRIS will also need to be signed - permitting the latter to receive the data on behalf of the researcher and transfer it into the Controlled Data Service. Since these processing agreements have significant penalties for infringement, the researcher's institution will want to ensure people with appropriate legal expertise check them. Many universities have data protection offices that can help researchers with these agreements and the PIA.

Before being permitted to use the Controlled Data Service, researchers will need to be accredited to show that they are fully aware of their legal obligations. This can be done by going through a free online Research Data and Confidentiality course provided by the Medical Research Council.

Once all agreements are in place controlled data will be transferred to, and then held within, a highly secure computing environment where it is possible to monitor closely who works on the data and to ensure that no personal data leaves the system. Appropriate procedures are in place for transferring and linking data from different sources to minimise the risks of disclosing confidential information. Depending on the nature of the data and on the requirements of the data owner or data controller, researchers may have access to the data through a remote login or they may be required to travel to a physical location with a safe setting. It will not be possible for researchers to remove data from the secure environment. All research outputs will be submitted to eDRIS for screening before release to the researcher for public use. Datasets constructed for each project will be destroyed on completion of the research (usually after a reasonable period to allow for publication). Neither UBDC nor the researcher's institution will become the permanent controller for these data.

Researchers should inform UBDC of any research outputs arising from projects that used the service. These would include publications and conference presentations. UBDC may use these to advertise both the research and the UBDC services that facilitated this work. UBDC will maintain a public database of live research projects, with a description of the purposes of the work and the data being used. We may also wish to follow up with projects to develop impact case studies and other resources about the outcomes and benefits of the research.

### *Things to consider when using administrative data*

While administrative data and other types of urban big data may be potentially valuable, unique and interesting to researchers, it should be kept in mind that these data were not collected for research purposes. Furthermore, public bodies - such as local authorities - do not have any obligation to share data with researchers: researchers do not have a default right to data. The data owners themselves do not necessarily have any direct interest in the research or its results. On the contrary, it can take significant time and effort for them to consider and respond to applications, while already working with limited resources. For this reason, researchers are strongly encouraged to consider:

- Whether data sharing could have organisational benefits for the data owners, and whether the research can be conducted in ways that increase the potential for such benefits.
- Whether the research might show the organisation in a negative light and whether steps can be taken to reduce these risks or allay the organisation's concerns.
- Engaging with the public body early in order to understand their needs and concerns to help move the process along. This includes communicating research aims and questions clearly plus

discussing any potential problems, risks and potential research findings in advance with the data owner.

- The 'social licence' for the project: in other words, would the public perceive this as a legitimate use of data? It should be kept in mind that while the use of any data may be legal, this might not be enough for the public to consider the use of these data as legitimate. Concerns about public reactions may also influence how data owners respond.
- Making potential public benefits (impact) explicit and ensure they are described in a clear and detailed manner.

### *Our data: the iMCD project*

**The Integrated Multimedia City Data (iMCD) project** is a Glasgow-wide data collection effort developed at UBDC. It involves a representative household survey and contextual data collected through newspapers, administrative reports, social media, and infrastructure-based sensors such as traffic detectors, environmental and weather sensors. The household survey also includes a sub-sample with GPS traces and images from a lifelogging device worn by survey participants. The GPS traces and lifelogging images are controlled data and can be used for research only through the UBDC application process. See here for more information: <http://ubdc.ac.uk/imcd-data/>.

### *Examples of UBDC projects using personal data*

You can see our approved projects here: <http://ubdc.ac.uk/data-services/current-projects-spotlight/>. Some examples include:

**Looked after children in Scotland.** Since 2000, the rate of children looked after (LA) has increased substantially in Scotland - in 2015 over 15,000 (1.5%) children in Scotland were LA compared to 11,300 (0.9%) in 2000 - but little is known about the health of looked after children. This research will compare the health of school-age LA children to that of other school-age children never LA between 2009/10 and 2015/2016. The research asks whether health, mortality and pregnancy rates among LA and not LA children differ and, if so, for which health-related measures are differences observed. Data for all school-aged children in Scotland (from ScotXed) will be linked to birth and death registrations (National Records of Scotland) and health records (SMR02, SMR01, SMR04; and Prescribing Information System from NHS National Services Scotland).

**Educational Disadvantage and Place.** This project aims to understand the relationship between place and educational disadvantage in the Greater Glasgow and Clyde area. It will use person level administrative data from pupil census records and pupil attainment records and combine these with aggregate information on schools, teachers and the neighbourhood. The combination of these data allows researchers to investigate the relationship between the characteristics of young people, the places where they live and the schools they attend, and analyse how each of these contributes to the educational attainment of young people. The outcomes from this research can inform policy aimed at narrowing the gap in educational attainment experienced by young people from disadvantaged backgrounds.

### *Further resources*

**Medical Research Council** – offers a free online Research Data and Confidentiality course. The course provides researchers with the framework and tools to interpret the legal requirements for research

with personal data. Required from all researchers using UBDC Controlled Data Service.

<http://byglearning.co.uk/mrcrsc-lms/course/category.php?id=1>

**Information Commissioner's Office** – an independent authority set up to uphold information rights in the public interest, promoting openness by public bodies and data privacy for individuals. The ICO provides detailed information on the DPA, data sharing and practical checklists, codes of practice and a templet for privacy impact assessment. <https://ico.org.uk/>