



The work is licensed under the Creative Commons Attribution-Non-Commercial-Share Alike 2.0 UK: England and Wales Licence. To view a copy of this licence, visit creativecommons.org/licenses/by-nc-sa/2.0/uk/

Guide to good practice: micro data handling and security

PUBLIC VERSION

31 AUGUST 2010

T +44 (0)1206 872001
E info@data-archive.ac.uk
www.data-archive.ac.uk



UK DATA ARCHIVE

UNIVERSITY OF ESSEX
WIVENHOE PARK
COLCHESTER
ESSEX, CO4 3SQ

WE ARE SUPPORTED BY THE **UNIVERSITY OF ESSEX**, THE **ECONOMIC AND SOCIAL RESEARCH COUNCIL**, AND THE **JOINT INFORMATION SYSTEMS COMMITTEE**

Document Control

Created by:	Karen Dennison
Maintained by:	Karen Dennison
Version:	03.00
Controlled document:	Not applicable
Last reviewed:	31 AUGUST 2010
Review due date:	31 AUGUST 2011

Contents

1. Licence framework	3
1.1. End User Licence data	3
1.2. Special conditions	3
2. Accessing data	4
2.1. Re-use of data	4
2.2. Research projects and teams	4
2.3. Teaching purposes	5
2.4. Security	5
3. Data storage security	5
3.1. End User Licence data	5
3.2. Special Licence data	5
3.3. Secure Data Service data	6
3.4. Passwords and pass-phrases	7
3.5. Audit of confidentiality and security procedures	7
4. Statistical disclosure	7
4.1. Direct disclosure	7
4.2. Indirect disclosure	8
5. Reporting publications	9
6. When research is complete	9
6.1. Guidelines on destroying Special Licence data	10
7. Institutional responsibilities	10
7.1. Special Licence and Secure Data Service data	11
8. Breach procedures	11
9. Help and feedback	12

Introduction

This guide is for users of micro data accessed from the UK Data Archive (the Archive) through its Economic and Social Data Service (ESDS). In particular, all users who obtain Special Licence or Secure Data Service (SDS) data, including data that require Accreditation as an Approved Researcher, are required to read it under the terms of access.

1. Licence framework

The Archive is the guardian of data for the data owners. The conditions under which data may be accessed are specified under licence with the data owners. These conditions include providing the data only to users who have registered with the ESDS and agreed to an End User Licence (EUL). Researchers accessing the data have responsibilities to preserve data confidentiality and to observe the ethical and legal obligations pertaining to the data. In particular, researchers must maintain the commitments made to survey respondents to preserve the confidentiality of the data provided.

1.1. End User Licence data

Use of the data is governed by a legally-binding EUL which forms part of the registration process. Each individual who requires access to data has to register and will need a UK Federation login. Users who have no way of obtaining a UK Federation login can apply to the Archive.

Under the terms of the EUL, users agree:

- not to use the data for any commercial purpose (except with prior permission/under an appropriate commercial licence agreement);
- to preserve the confidentiality of, and not attempt to identify, individuals, households or organisations in the data;
- to use the recommended methods of citation and acknowledgement in publications;
- to supply the bibliographic details of any published work based on the data collections;
- to ensure that the means of access to the data (such as passwords) are kept secure and not disclosed to anyone else;
- to abide by any further 'special conditions'.

1.2. Special conditions

Additional legally-binding conditions to those of the EUL may be specified by the data owners for particular data collections. Where data pose a higher risk of disclosure, special conditions may take the form of a Special Licence that requires the completion of an additional application form, the signature(s) of the researcher(s), and the explicit permission of the data owners to release the data to the researcher(s). Access to Special Licence data may be restricted to certain users (for example, to UK applicants only).

Access to download ONS Special Licence data via the Archive is available through the legal framework set out in the Statistics and Registration Service Act which came into force 1 April 2008. Any registered user wanting to download ONS Special Licence data will have to be

accredited by the UK Statistics Authority as an Approved Researcher. To apply for accreditation a user will complete (i) forms that will require evidence that he/she is a fit and proper person and details about the purpose of the research (ii) an online order for the data and (iii) a signed declaration that he/she understands the confidentiality obligations owed to those data including its physical security.

Data considered to be more disclosive than Special Licence data may be available through a virtual secure environment to UK academics via the ESRC-funded UK Data Archive Secure Data Service. Any registered user requiring access to SDS will have to (i) be accredited as an ESRC and/or UK Statistics Authority Approved Researcher (ii) complete SDS Training and (iii) agree to an SDS Service Agreement.

2. Accessing data

Data can only be accessed under certain conditions:

- under the EUL, data can only be accessed by registered users;
- data supplied under special conditions can only be accessed by those who have accepted these conditions;
- Special Licence and Secure Data Service data can only be accessed by approved individuals for a specified usage;
- Secure Data Service data can only be analysed remotely within the SDS virtual secure environment and outputs are only released to researchers subject to statistical disclosure control by SDS staff.

2.1. Re-use of data

To re-use data already supplied, but for a different purpose, it is necessary to re-apply for access. For example, if depositor permission is required, this will need to be obtained again.

2.2. Research projects and teams

Users are required to register all usages of data, and to add other users working on the same usage to the usage details in their online account. Each usage is given a period of access that cannot exceed two years for EUL data and three years for Special Licence and Secure Data Service data. Users should contact the help desk to extend the expiry date of a project.

Where a researcher joins a research team that is using Special Licence or Secure Data Service data:

- the new researcher must place an online order for the data and complete the necessary forms;
- permission must be sought and gained before the new researcher can access the data;
- the researcher must complete SDS training to access Secure Data Service data;
- the Archive will provide advice on the process to be followed.
-

2.3. Teaching purposes

When using data for teaching all students must be registered or have signed an access agreement for teaching, which should be returned to the Archive.

Special Licence and Secure Data Service data cannot be used for teaching.

2.4. Security

Passwords and pass-phrases should never be disclosed to anyone else. Data should not be left on a computer that might enable unauthorised access.

3. Data storage security

3.1. End User Licence data

All data provided by the Archive needs to be stored under conditions that meet the undertakings given in the EUL (see section 7 for institutional responsibilities):

- access to PCs on which data are held should require personal authentication (secure username and password/pass-phrase);
- if data are placed in a shared directory or on a Local Area Network (LAN), access should only be available, via personal authentication, to those permitted to use the data;
- means of access to the data (such as passwords or pass-phrases) should be kept secure;
- data must be stored securely with data on portable media (e.g. a back-up on CD) protected using a secure password/pass-phrase;
- users should be aware of, and follow, any additional information security guidelines provided by their institution/organisation.

3.2. Special Licence data

Special Licence data should additionally:

- be protected, where possible, using pass-phrases instead of passwords;
- be stored in physically secure conditions (e.g. any portable or printed copies should be stored in a locked cabinet with restricted access);
- be stored on a PC in a room which is NOT accessible to the general public;
- be stored on a PC in a locked office when unattended;
- be protected by a screen-saver with an interval of five minutes and that requires a secure password/pass-phrase to unlock it;
- only be accessed, in an institutional setting, via a stand-alone PC or a closely controlled LAN

with restricted access - data must not be accessed at a private residence;

- must not have live internet links while the data are unencrypted on the machine unless access is through a secure organisational provider, such as JANET: (If there is any uncertainty as to whether an organisational provider is 'secure' users should contact the help desk with details of the system in place.)
 - stand-alone PCs and LANs, which have internet access via broadband (and not through a secure organisational provider e.g. JANET) must be disconnected from the Internet and the broadband cable must be physically disconnected from the PC;
 - stand-alone PCs and LANs, which have internet access via dial-up telephone connection (and not through a secure organisational provider e.g. JANET), must not have live internet links while the data are unencrypted on the machine;
- be accessed on a site which has security standards that meet the guidelines in this guide;
- be auditable;
- be accessed at a site within the UK as Special Licence data will not be provided to licence holders who are sited at, and thus would intend to access data, at an institution that is not within the UK;
- be deleted upon project completion:
 - this must be confirmed by the licence holder;
 - copies of data (including derived datasets) must be deleted from all PCs used;
 - any printed or electronic copies, including back-ups, temporary or intermediate files, must be destroyed;
 - any portable copies (e.g. CDs) must be destroyed or returned to the Archive;
 - Section 6.1 has detailed guidelines.

3.3. Secure Data Service data

Secure Data Service data will:

- only be accessed remotely via the SDS virtual secure environment;
- only be accessed in an institutional setting and within the UK - data must not be accessed at a private residence;
- be accessed on a site which has security standards that meet the guidelines in this guide;
- only be accessed in a room which is NOT accessible to the general public and that is locked when unattended;
- (depending upon the data owner requirements and the disclosiveness of the data) be accessed using one of the following:
 - o the researcher's desktop PC;
 - o secure machine provided by the SDS;
 - o a designated SDS safe room;

- be protected by a screen-saver with an interval of five minutes and that requires a secure password/pass-phrase to unlock it.

Data may not be removed from the SDS virtual secure environment under any circumstances and outputs will only be released to the researcher subject to statistical disclosure control checks by SDS staff. Under no circumstances should users attempt to take away/copy outputs that have not been checked and released to them by the SDS. SDS data and unauthorised outputs should not be printed or be seen on the user's computer screen by unauthorised individuals

Researchers who have been approved to work together on the same project may only share unchecked outputs from that project with each other in the relevant shared project area of the SDS virtual secure environment. Temporary or duplicate files and disclosive outputs should be deleted by the researcher(s) from the SDS virtual secure environment.

Use of the SDS virtual secure environment will be monitored including any suspicious activity and keystrokes.

3.4. Passwords and pass-phrases

Pass-phrases differ from passwords in format and in length. Pass-phrases are usually much longer - up to 100 characters or more and contain spaces. The greater length and format of pass-phrases makes them more secure.

A password should contain a combination of at least eight alphanumeric and symbolic characters. Quotes should not be used as pass-phrase characters.

Passwords and pass-phrases should:

- not be disclosed to anyone else;
- not be written down;
- be changed at least every three months;
- not be easily guessable;

The Secure Data Service will provide users with personal logins. Passwords will be renewed every three months and logins will expire at the end of the project.

3.5. Audit of confidentiality and security procedures

The depositor of Special Licence or Secure Data Service data may reserve the right to conduct an onsite audit of the licence holder's confidentiality and security procedures and practices, or to require a report of such an audit. For the purpose of conducting an audit, the depositor may reserve the right of entry to the premises where the data are stored and/or accessed. (Also see section 7.1).

4. Statistical disclosure

4.1. Direct disclosure

The EUL requires an undertaking not to attempt to identify any individual, household or organisation or claim to have done so. Where EUL data are matched with external data sources this must not be for the purposes of identification.

For Special Licence data, it is forbidden to attempt to match individual, household, or organisation records to any other data, including data from other Special Licence data series, at the level of individual, household or organisation. Only area-level descriptors or other group-level classifications may be matched for analysis purposes.

Whilst the Secure Data Service provides a secure environment in which data could be linked (for example with another dataset in the SDS collection or with the researcher's own data) this is strictly subject to the approval of the data owners. Users will only be able to access those datasets approved for a particular research project – it will not be possible to subsequently add new data without a new application and approval. Where a user is approved to link data not in the SDS collection, the data will be uploaded to the user's area subject to checks by SDS.

ONS business data may be linked using the anonymised reference numbers (known as IDBR references). A user may be able to produce a larger 'combined' dataset, with many variables providing characteristics that will directly identify an organisation. While this is an acceptable risk within the confines of the SDS virtual secure environment, users should be aware that output requests containing information that will identify an organisation, will be rejected.

4.2. Indirect disclosure

4.2.1. Special Licence data

Outputs from Special Licence data must be subjected to disclosure control. The guidance below is general advice but users should also refer to the full details of the procedures to be used in the Government Statistical Service (GSS) guidance available from the [Office for National Statistics Access Arrangements web page](#).

Tables that contain very small numbers in some cells may be disclosive. Tables should not report numbers or percentages in cells based on only one or two cases. Cells based on one or two cases should be combined with other cells or, where this is not appropriate, reported as zero per cent.

Tables and other outputs must not be published in a form where the level of geography would threaten the confidentiality of the data. To guarantee safety, outputs from Special Licence data should not be published if the geography is lower than UK Government Office Region (GOR).

If there is a requirement to publish outputs from Special Licence data with a lower level of geography i.e. between GOR and local authority, then the licence holder must consider whether there is a risk of disclosure. Where there is any doubt, the licence holder must contact the help desk to gain confirmation of the confidentiality of any outputs for publication with geography below GOR.

No outputs may be published with a geography below local authority.

Although most outputs from models or other statistical analysis will not be disclosive, care should be taken to ensure that individuals, households or organisations cannot be identified. In particular, results based on very small numbers should be avoided. Any output that refers to unit records, e.g. a maximum or minimum value, should be avoided. Models should not report actual values for residuals.

Graphical outputs should be based on non-disclosive data. Particular care should be taken not to report extreme outliers. Graphical outputs should respect all the rules specified in the GSS Disclosure Control Policy.

4.2.2. Secure Data Service data

Access to data held in the SDS may only be accessed within the SDS virtual secure environment. Users must conduct all their analysis, and produce outputs (such as papers, presentations etc) within this area. Hence data may not be released from the SDS under any circumstance. Outputs may be returned to researchers subject to a full manual statistical disclosure control (SDC) by a member of the SDS team.

The SDC requirements for these data differ from those mentioned above for the Special Licence files. There are two reasons for this difference: first, the data are more sensitive, and contain variables that directly identify survey respondents; and secondly, the SDS will make available business data, for which a large number of additional methods (other than social science techniques) may be adopted by researchers, and present additional disclosure concerns that should be considered.

For example, Herfindahl/concentration indices are routinely calculated by industrial economists using business data. Such measures generate additional disclosure concerns which are not addressed by the guidelines for Special Licence data above.

Users should remain wary of producing outputs containing low cell counts. However, there are also other techniques which users should be vigilant about when producing their outputs. Details of the relevant SDC guidelines for outputs generated in the SDS may be found in the ESSnet 'Guidelines for the checking of output based on microdata research' and we recommend that users consult these guidelines.

All users who wish to access data in the SDS will be required to attend a training course where these ESSnet SDC guidelines will be explained. If users are unsure about SDC when they produce outputs, we recommend that they speak to an SDS support officer as soon as possible, and certainly before they intend to request their outputs. For example, a researcher will be disappointed if they write an entire paper within the SDS, only to find that it is not released to them due to SDC problems.

5. Reporting publications

All users of data are required to report publications arising from their research to the Archive. It is good practice to inform the Archive of any publications at the time of publication.

The Archive will provide annual reports to ONS on publications arising from the use of ONS Special Licence and Secure Data Service data.

ONS reserves the right to ask to see drafts of publications based on ONS Special Licence data for the purpose of commenting regarding compliance with the conditions for disclosure protection. If this condition is imposed, users will be notified when their application is processed.

Secure Data Service users are not permitted to publish outputs unless they have been checked and released to them by the SDS.

-

6. When research is complete

It is recommended that researchers always retain a well-documented copy of the syntax used to prepare a paper or report. Derived data should be offered for deposit at the Archive.

When a project has been completed it is good practice for researchers to remove all copies of the data, including derived datasets, back-ups, paper copies, portable copies (including CDs), and all electronic copies from every PC used. Researchers using SDS data should ensure that any unnecessary or duplicate files are removed from the SDS secure virtual environment at the end of the project.

It is essential that all copies of Special Licence data held by researchers are destroyed or, where held on portable media, returned to the Archive at the end of the specified time period. Special Licence derived data can also be returned to the Archive and the Archive can maintain the data on behalf of the user in a secure area for a five-year period. To re-access the data, the researcher should contact the Archive.

Syntax files can be removed from the Secure Data Service virtual secure environment subject to clearance checks by SDS staff. Alternatively the SDS can maintain syntax files on behalf of the user for a five-year period.

6.1. Guidelines on destroying Special Licence data

The following are guidelines for destroying data:

- data should be deleted from the system on which it has been stored using a secure erasure programme, such as Disk Sanitizer (www.east-tec.com/eraser/index.htm) or similar - which repeatedly overwrites files a number of times, until such time as the original data could not be retrieved forensically;
- the recycle/trash bin should be emptied, preferably to be immediately followed by running a secure erasure programme to erase the unused areas of the disk;
- CDs and portable media should be returned to the Archive or cut into many pieces or shredded using a disk shredder and then securely disposed of;
- backup tapes should either be completely overwritten and degaussed (demagnetised) before being re-used or disposed of;
- paper copies should be destroyed by shredding, preferably using a cross-cut shredder;
- before the PC leaves the possession of the organisation (for destruction or second hand sale, etc.) the hard disk should be completely erased using a secure erasure programme;
- destruction of Special Licence data must be confirmed to the Archive by the licence holder.

7. Institutional responsibilities

Institutes of UK higher or further education (HE/FE), are bound by JANET policies (www.ja.net/services/publications/policy/security-policy.pdf), including the JANET Security Policy that places responsibilities on every person and organisation involved in the use or operation of JANET to protect the network against security breaches. UK HE/FE should also follow JISC guidance on information security, including handling information legally (www.jisc.ac.uk/uploaded_documents/ACF63.pdf).

There is a requirement that all central government departments must meet internationally recognised information and security management standards (e.g. ISO/IEC 17799 and BS 7799) for their systems. Local authorities are also obliged, by 2005, to comply with the BS 7799 security standard as part of their Implementing Electronic Government (IEG) requirements.

7.1. Special Licence and Secure Data Service data

Where access to data requires an institutional signature:

- in Universities, the application should be signed by the Head of Department or School, Head of Research Centre, or the chair of the University Ethics Committee;
- in Government departments, or local authorities, the licence should be signed by the statistician with responsibility to represent the organisation or to enter it into a contractual relationship;
- in all cases an institution is required to accept legal responsibility for the user.

For a user accessing ONS Special Licence data through the Approved Researcher mechanism, it is the user's responsibility to ensure that they can store and access the data in a suitably secure physical and electronic environment.

Users of Special Licence and Secure Data Service data undertake to allow the depositor access to the premises where the data are stored and/or accessed for the purpose of conducting an audit, without notice and at any reasonable time. (Also see Section 3.4).

Access to Special Licence and Secure Data Service data may require the user to provide the contact details of a senior member of staff at their institution who can vouch for their suitability for access to the data. The Archive and ONS reserve the right to contact the senior member of staff to ask for a reference.

8. Breach procedures

The licence holder is required to report promptly a breach of any of the terms of the EUL, including the terms of any data supplied under special conditions. Failure to disclose details of a breach constitutes a breach of the licence.

Breach of the terms of the EUL, including any special conditions, may result in the following actions:

- immediate termination of access to all services provided by the Archive and ESDS either permanently or temporarily;
- legal action being taken against the individual who has breached the terms of the EUL;
- withdrawal of access to all Archive and ESDS services either permanently or temporarily to the licence holder's institution.

Additionally, any breach in the terms of access for Special Licence or Secure Data Service data:

- will result in the immediate termination of the licence holder's access to the data and the termination of the licence - depending upon the seriousness of the breach, the termination of access may be permanent;
- may result in sanctions being sought against the licence holder by the data owner;
- for ONS Special Licence and ONS Secure Data Service data, under the Statistics and Registration Services Act 2007, will incur penalties as specified in S39 of the Act - this may include a fine and/or imprisonment;

- for Secure Data Service data penalties could also include individual or institutional sanction from ESRC funding and institutional suspension from all ESRC data services.

Researchers will be provided with detailed guidance on breaches and penalties in the SDS training.

9. Help and feedback

This guide will be regularly updated. For further advice on any of the issues raised, or to provide suggestions or comments, contact the ESDS or SDS help desk.

ESDS help desk

- email: help@esds.ac.uk
- telephone: +44 (0) 1206 872143

SDS help desk

- email: securedata@ukda.ac.uk
- telephone: +44 (0)1206 874968