Microdata Handling and Security

Guide to Good Practice

PUBLIC VERSION
09 February 2017
Version: 06.00

T +44 (0)1206 872832
E wardm@essex.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**

Contents

Scope

This guide is for users of microdata accessed via the UK Data Archive (the Archive) through its online services provided by the UK Data Service. In particular, all users who obtain Special Licence data or data considered to be more disclosive than Special Licence data (known as 'secure data') are required to read this document under the terms of access.

1. Licence framework

The Archive does not own the data in its collection but is licensed by the data owners to curate and share the data on their behalf. The conditions under which data may be accessed are specified in the deposit licence. These conditions include providing the data only to users who have registered with the Archive's online services and agreed to an End User Licence (EUL). Users accessing the data have responsibilities to preserve data confidentiality and to observe the ethical and legal obligations pertaining to the data. In particular, users 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 with the UK Data Service and will need a UK Access Management Federation (UKAMF) login. Users who are part of the UK Higher/Further Education (UK HE/FE) sector will automatically be issued with these details by their organisation. Users who have no other way of obtaining a UKAMF 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 citation and acknowledgement information provided by the Archive, in publications;
- to supply to the Archive, 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 user(s), and the explicit permission of the data owners to release the data to the user(s). Access to Special Licence data may be restricted to certain users (for example, to UK applicants only).

Secure data are only available through a virtual secure environment, the UK Data Service Secure Lab. Any registered user requiring access to secure data will have to (i) be accredited as an ESRC and/or UK Statistics Authority Approved Researcher, (ii) complete training and (iii) agree to a Secure Access User 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 can only be accessed by approved individuals for a specified usage and for a specified time;
- Secure data can only be analysed remotely within the Archive's Secure Lab and outputs are only released to users subject to statistical disclosure control by Archive 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 research projects via their online UK Data Service account. A research project can be linked to each user on the project. The period of access is set by the data owner for Special Licence and secure data. Users should contact the UK Data Service Helpdesk if they wish to extend a project.

Where a user joins a research team that is using Special Licence or secure data:

- the new user must place an online order for the data and complete the necessary forms;
- permission must be sought and gained before the new user can access the data;
- the user must complete training to access secure 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 must be returned to the Archive.

Special Licence and secure data cannot be used for teaching.

2.4. Security

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

3. Data storage security

3.1. End User Licence data

All data provided by the Archive must be stored under conditions that meet the undertakings given in the EUL (see section 7 for organisational responsibilities) and those listed below:

- access to PCs on which data are held must have personal authentication (i.e. protected by a secure username and password/pass-phrase);
- if data are placed in a shared directory or on a Local Area Network (LAN), access must 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) must be kept secure;
- data must be stored securely;
- data on portable media (e.g. a back-up on CD) must be protected using a secure password/pass-phrase;
- users must be aware of, and follow, any additional information security guidelines provided by their organisation;
- data must be deleted upon project completion as set out in section 6.1.

3.2. Special Licence data

In addition to the responsibilities under the EUL, Special Licence data:

- must only be accessed, in an organisational setting, via a stand-alone PC, laptop, other portable device or a closely controlled LAN with restricted access and must not be accessed at a private residence;
- must be protected, where possible, using pass-phrases instead of passwords;
- must be protected by a screen-saver with an interval of no more than five minutes and that requires a secure password/pass-phrase to unlock it;
- must be stored in physically secure conditions (e.g. any portable or printed copies must be stored in a locked cabinet with restricted access);
- must be stored on a PC, laptop or other device in a room which is NOT accessible to the general public;
- must be stored on a PC, laptop or other device in a locked office when unattended;
- if stored on a laptop, or other portable device, can only be accessed in an organisational setting, i.e. the device must be treated as though it is a PC and not removed from the organisation;
- 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 must contact the UK Data Service Helpdesk with details of the system in place);
- must be accessed on a site which has security standards that meet the guidelines in this guide;
- must be auditable;
- must only be accessed at an appropriate site according to the access requirements set by the data owner;
- must be deleted upon project completion as set out in section 6.1.

Stand-alone PCs, laptops, other devices 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. Those which have internet access via a 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.

3.3. Access to data held within the Secure Lab

The UK Data Service provides two methods of access to confidential data via the Secure Lab: remotely from the user's organisation and from the safe room at the UK Data Archive.

Data accessed remotely via the Secure Lab must:

- only be accessed by a user following attendance at a training course;
- only be accessed remotely from a secure environment;
- only be accessed in an organisational 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 document;
- only be accessed in a room which is NOT accessible to the general public and that is locked when unattended;
- be accessed using one of the following methods (depending upon the data owner's requirements and the sensitivity of the data):
    - the user's desktop PC or Laptop;
    - a secure Thin Client supplied by the Archive;
    - a designated safe room;
- be protected by a screen-saver with an interval of five minutes that also requires a secure password/pass-phrase to unlock it.

When accessing data via the Secure Lab, it is not technically possible for a user to transfer or download or copy and paste any data to a local computer, or to print to a local computer. Users must not copy screenshots to a local computer.

Outputs are only released to the user subject to statistical disclosure control checks by Archive staff. Users are strictly forbidden from copying anything from the screen. Secure Lab data and outputs must not be seen on the user's computer screen by unauthorised individuals.

Users 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 within the Secure Lab. Temporary or duplicate files should be deleted by the user(s) from the Secure Lab.

Use of the Secure Lab will be monitored to provide the Archive with information about any suspicious activity and keystrokes.

3.3.1. Access via the UK Data Archive's safe room

The conditions under which data are accessed via the UK Data Archive's safe room are similar to those for accessing the Secure Lab remotely. However, access via the safe room differs from remote access in that:

- access is only available from within the room;
- thin-client terminals are used to access the Servers where the data are held;
- users must abide by the procedures, listed in the *safe room procedures* document (currently known as CD226-SafeRoomProcedures)*;*
- users are required to visit the UK Data Archive to carry out their research, and to undertake a special training programme to ensure they are aware of how to safely use the Secure Lab.

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

Passwords and pass-phrases must:

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

The Archive will provide users with personal logins to access the Secure Lab. Users are required to change their password on first logon and to renew it every three months.

3.5. Audit of confidentiality and security procedures

The Archive and the depositors of secure data 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 Archive and the depositors 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. Data matching

Data matching can lead to disclosure and is therefore only permitted under certain circumstances.

4.1.1. End User Licence data

Where EUL data are matched with external data sources this must not be for the purposes of identification.

4.1.2. Special Licence data

Any plans to match or attempt to match individual or household records to any other data source at the level of the individual or household must be declared and can only be undertaken with the permission of the data depositor and the owners of the data sources.

4.1.3. Secure data

Whilst the Secure Lab provides an area in which secure data could be linked (e.g. with another dataset in the secure collection or with the user'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 of this.

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 Secure Lab, users must be aware that output requests containing information that will identify an organisation, will be rejected.

4.2. Outputs

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 must also refer to the full details of the procedures to be used in the UK Statistics Authority Code of Practice for Official Statistics (https://www.statisticsauthority.gov.uk/monitoring-and-assessment/code-of-practice/) and the GSS Statistical Disclosure Control (https://gss.civilservice.gov.uk/statistics/methodology-2/statistical-disclosure-control/) for tables produced from surveys.

Tables that contain very small numbers in some cells may be disclosive. Tables must not report numbers or percentages in cells based on only one or two cases. Cells based on one or two cases may 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, e.g. between GOR and local authority, then the user must consider whether there is a risk of disclosure. Where there is any doubt, the user must contact the UK Data Service via the Helpdesk to obtain confirmation of the acceptability of publication of the output if the geography is 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 must be taken to ensure that individuals, households or organisations cannot be identified. In particular, results based on very small numbers must be avoided. Any output that refers to unit records, e.g. a maximum or minimum value, must be avoided. Models must not report actual values for residuals.

Graphical outputs must be based on non-disclosive data. Particular care must be taken not to report extreme outliers.

4.2.2. Secure data

The Statistical Disclosure Control (SDC) requirements for secure data differ from those mentioned above for Special Licence data.

Access to secure data is only through the Secure Lab. Users must conduct all their analysis, and produce outputs (such as papers, presentations etc.) within this area. Therefore secure data will not be released under any circumstance. Outputs will be returned to users following a full manual SDC examination by a trained member of Archive staff. It is the policy of the Archive to only release ‘final results’ which are those considered ready for publication. This is because users working on the same project in the Secure Lab can easily share their intermediate findings through shared project folders.

The differences between SDC management for Special Licence and secure are twofold: firstly, the secure data are more sensitive, and contain variables that directly identify survey respondents; and secondly, the secure data include business data, for which a large number of additional methods (other than social science techniques) may be adopted by users, and therefore present additional disclosure concerns that must 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. These are addressed in detail as part of the mandatory training for use of this service.

Users must remain wary of producing outputs containing low cell counts and to maintain familiarity with SDC. The guidelines for outputs that are found in the ESSnet ‘Guidelines for the checking of output based on microdata research’ are recommended.

If users are unsure about SDC when they produce outputs, we recommend that they speak to an Archive support officer as soon as possible, and certainly before any outputs are submitted for checking. This will avoid disappointment if a user writes an entire paper within the secure environment, 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 and the full citation of the publication should be supplied via the UK Data Service Helpdesk.

Depositors of Special Licence data reserve the right to ask to see drafts of publications based on those 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.

Users of the Secure Lab are not permitted to publish outputs unless they have been checked and released to them by the Archive.

6. When research is complete

It is recommended that users always retain a well-documented copy of the syntax used to prepare a paper or report.

When a project has been completed users should 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.

It is essential that all copies of Special Licence data held by users are destroyed and the Archive notified via the completion of a Data Destruction form.

Users using secure data must ensure that their syntax files are stored in the 'Syntax Folder'. All other files will be deleted at the end of the project. Syntax files can be removed from the Secure Lab subject to clearance checks by Archive staff.

6.1. Guidelines on destroying data

The following are guidelines for destroying data:

- data must be deleted from the system on which it has been stored using a secure erasure programme, such as Eraser (http://www.heidi.ie/eraser/index.php) 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 must be emptied, preferably to be immediately followed by running a secure erasure programme; portable media holding any data must be returned to the Archive or destroyed and disposed of in a secure manner;
- backup tapes must either be completely overwritten and degaussed (demagnetised) before being re-used or disposed of;
- paper copies must be destroyed by shredding, preferably using a cross-cut shredder;
- before the PC, laptop or other device used for data storage leaves the possession of the organisation or individual (for destruction or second hand sale, etc.), the hard disk must be completely erased using a secure erasure programme;
- destruction of Special Licence data must be confirmed to the Archive by the user. A Data Destruction form will be sent to the user one month before the project expires.

7. Organisational responsibilities

UK Institutes of higher or further education (HE/FE) are bound by JANET policies (https://community.ja.net/library/janet-policies/security-policy), 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 must also follow JISC guidance on information security, including handling information legally (https://www.jisc.ac.uk/guides/security-mobile-devices-and-data-protection).

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 to comply with the BS 7799 security standard as part of their Implementing Electronic Government (IEG) requirements.

7.1. Special Licence and Secure Lab data

For a user accessing Special Licence data, 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 the Secure Lab must provide a Secure Access User Agreement which should be signed by an authorised signatory of the user's organisation.

Users of Secure Lab 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 3.5).

Access to Special Licence and Secure Lab data may require the user to provide the contact details of a senior member of staff at their organisation 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. Non-compliance procedures

The user is required to report promptly any non-compliance with any of the terms of the EUL, Special Licence or Secure Lab rules (this includes any non-compliance by someone else that the user becomes aware of). Failure to disclose an act of non-compliance is a non-compliance with the licence.

Non-compliance with 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 the UK Data Service either permanently or temporarily;
- legal action being taken against the individual who has not complied with the terms of the EUL;
- withdrawal of access to all Archive and UK Data Service services either permanently or temporarily to the user's organisation.

Additionally, any non-compliance with the terms of access for Special Licence or secure data:

- will result in the immediate termination of the user's access to the data and the termination of the licence; depending upon the seriousness of the non-compliance, the termination of access may be permanent;
- may result in sanctions being sought against the user by the data owner;
- will, for ONS Secure Lab data under the Statistics and Registration Services Act 2007, incur penalties as specified in S39 of the Act, which may include a fine and/or imprisonment;
- for Secure Lab data, penalties could also include individual or organisational sanctions including withdrawal of ESRC funding and organisational suspension from all ESRC data services.

Users will be provided with detailed guidance on non-compliance and penalties when undertaking the Secure Lab 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 UK Data Service Helpdesk via our 'Get-in-touch' web page: http://ukdataservice.ac.uk/help/get-in-touch.aspx