

Notification of Change of Serial Numbers NCDS, BCS70 and MCS February 2008

This document is to inform researchers that all the serial numbers that are used to uniquely identify individuals, attached to data from the 1958, 1970 and Millennium Cohort Studies, are being changed. This change of serial numbers will apply to all datasets from the cohort studies that are already in the UK Data Archive as well as any new datasets that are deposited. The reason for this is to improve the security of the data and increase safeguards on the confidentiality of cohort members. The aim is to complete these changes during March for MCS and by the end of June 2008 for NCDS and BCS70. Further information about the background to the decision to change serial numbers is provided below.

The aim of this note is to provide practical information and guidance to researchers who have datasets based on the cohort studies that have been downloaded from the UK Data Archive or who have received special datasets of cohort data directly from CLS (e.g. for beta testing).

All datasets that carry the existing/old serial numbers should normally be deleted within three months of receipt of the revised data by the researcher. CLS requires that all End User Licence holders delete any earlier versions of the data that they hold. However, where a user has downloaded earlier versions of the data and is using them in a current project, they may continue to use the data and any outputs derived from their use until the project is completed. On completion of the project, users are then required to delete the original dataset/s. CLS does not endorse the use of any previous versions of the datasets unless they are being used for a current project. Any future projects should use the new versions of the data. Any users having concerns about this should contact the CLS.

Further guidelines on destroying data are provided below and these are also available as part of guidelines on data handling and security on the ESDS website at

<http://www.esds.ac.uk/news/publications/microDataHandlingandSecurity.pdf>

Researchers who have created their own datasets based on the deposited cohort data but which include variables which they have derived, recoded or should normally be in a position to recreate these datasets by applying their existing SPSS syntax files (or STATA Do files) to the revised data deposited at the archive. However, it is recognised that this will not always be possible, and researchers who need assistance in modifying existing derived datasets so that they are indexed by the new serial numbers should get in touch with ESDS Longitudinal ([contact details here](#)). In the majority of cases, users will then be referred to CLS who will be able to translate the old serial numbers into the new serial numbers and re-issue users with a copy of any existing datasets but with the new research serial number attached.

Study	Variable name for individual unique identifier	Examples of old unique identifiers	Format of new unique identifiers
NCDS (1958 cohort)	serial	010001K 010002M 010003P 010004R 010006V	N00000X
BCS70 (1970 cohort)	key	180 190 200 240	B00000X
MCS	famsrno	A10503W A10504X A10505Y A10506A	M00000X

Guidelines on destroying data

See: <http://www.esds.ac.uk/news/publications/microDataHandlingandSecurity.pdf>

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 UKDA 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;

Note that destruction of any special datasets supplied under Special Licence must be confirmed to the UKDA by the licence holder.

Background to the change of serial numbers: CLS Confidentiality and Data Security Review

For some months now there has been concern at CLS about the need to improve data security to ensure the continued anonymity of cohort members. Discussions have been held both at the CLS Scientific Committee and at the National Longitudinal Strategy Committee about the need to recognise the following concerns.

- a) Some of the data held on cohort members and available to researchers via the archive are of a sensitive nature.
- b) The nature of the sample of cohort members (all those born in a single week or sampled in particular wards) coupled with the longitudinal nature of the data means that the risk of disclosure is somewhat higher than for other similar studies.

To ensure confidentiality and reduce the risk of disclosure, names and addresses of cohort members are held on an address database on a secure isolated network within CLS that is completely separate from the research data in the archive. However, a recent ESRC-funded project which has enabled CLS to scan the huge archive of materials related to the cohort studies in their entirety, has highlighted that there are a large number of documents (such as archived address record forms) that do link the names and addresses of individuals to the serial numbers that are used as identifiers on the research data.

Given the increased concerns about identity theft, and the publicity and concern surrounding the loss by HMRC of addresses of all those families claiming Child Benefit in November 2007, it was decided that there was a pressing need to ensure that all practical steps are taken to ensure the anonymity of cohort members. In conjunction with its host institution, the Institute of Education, CLS is therefore conducting a thoroughgoing review of data security. Some of the first practical measures that are being taken as a result of the initial phase of this review are as follows:

- 1) All research data, based on the cohort studies, deposited at the data archive will be issued with a new 'research' serial number that will allow for longitudinal linking of research datasets, but is not directly associated with the name or contact details of any cohort member. All datasets carrying the old serial numbers will be removed.
- 2) All future surveys will use survey-specific serial numbers for field work so that the 'research' serial number that is used as an identifier in the data archive never appears on the same document as the name or address of a cohort member.

It is recognised that the replacement of serial numbers described under (1) above will have implications for some existing users of the cohort data where researchers have 'cleaned' data using serial numbers. CLS believe that it is of paramount importance to safeguard the interests of cohort members but do not want to hinder the important scientific research that is already underway using data from the studies.

CLS will therefore recall all old cohort datasets and where necessary will reissue users with their data with the old serial number replaced with the new research serial number.