Notification of Change of Serial Numbers Next Steps (First Longitudinal Study of Young People in England) May 2017

This document is to inform researchers that all the serial numbers that are used to uniquely identify individuals, attached to data from the Next Steps cohort study, are being changed. This change of serial numbers will apply to all datasets (waves 1 to 7) from this cohort study that are already in the UK Data Archive as well as any new datasets that are deposited. It also applies to the linked NPD data currently available on Secure Access. 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 May at which time sweep 8 of Next Steps will be deposited at the UKDS.

The aim of this note is to provide practical information and guidance to researchers who have datasets based on this cohort study that have been downloaded from the UK Data Archive. All datasets that carry the existing/old serial number (surveyid) 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 ongoing 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 CLS. Further guidelines on destroying data are provided below.

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 there may be cases where this will not be possible, therefore researchers who need assistance in modifying existing derived datasets so that they are indexed by the new serial numbers should apply for assistance with UK Data Service via the attached link https: //www.ukdataservice.ac.uk/help/get-in-touch.

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 (NSID) attached.

New identifier

Variable Name for individual identifier	Format Description of variable name	Example of identifier
NSID	8 characters: NS followed by 5 digits followed by a single character	NS21140C

Old identifier

Variable Name for individual identifier	Format Description of variable name	Example of identifier
SURVEYID	5 digits	25533

Guidelines on destroying 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 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.

CLS Confidentiality and Data Security

Please note that the decision to introduce a new research id was made with the need to improve data security to ensure the continued anonymity of cohort members. Sweeps prior to the age 25 sweep were collected by the Department for Education (DfE). The study was transferred from the DfE to CLS in 2013, we will redeposit the earlier sweeps of the data with the new id to ensure that the data can be used longitudinally.

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.

The a new research serial number 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.

As is our current practice, 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 will never appear on the same document as the name or address of a cohort member.

It is recognised that the replacement of serial numbers as described 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.