Keeping Data Confidential
Anonymising Records

'Personal data' are any information about living people who can be identified by that information, or from a combination of the data and other information that the person in control of the data has, or is likely to have in the future.

'Coded data' are identifiable personal information in which the details that could identify someone are concealed in a code, but which can readily be decoded by those using the data. They are not anonymised data.

'Anonymised data' are data prepared from personal information but from which the person cannot be identified by the recipient of the information.

'Linked anonymised data' are anonymous to the people who receive and hold it (e.g. a research team) but contain information or codes that would allow the suppliers of the data, such as Social Services, to identify people from it.

'Unlinked anonymised data' contain no information that could reasonably be used by anyone to identify people. The link to individuals must be irreversibly broken. As a minimum, unlinked anonymised data must not contain any of the following, or codes traceable by you for the following:

- name, address, phone/fax number, email address, full postcode
- NHS number, any other identifying reference number
- photograph, names of relatives

Linked data are typically used when it may be necessary to refer back to the original records for further information, or for verification, or if it is planned to provide feedback to participants or service providers. Unlinked data usually ensures confidentiality but prevents follow-up, verification or feedback, may not be compatible with the aims of the project and may not be in the interests of the individuals or service providers.

With both linked and unlinked anonymised data it is sometimes possible to deduce an individual's identity through combinations of information. The most important identifiers are:

- family structure – eg has Deaf twins; the only Deaf child in a family of 6 children; married to a hearing Portuguese man
- rare disease or treatment, especially if an easily noticed health problem/disability is involved
- partial postcode or partial address
- location of the interview/meeting or the name of the educational or social services professional responsible for care
- rare occupation or place of work
- combinations of birth date, ethnicity, place of birth and date of death

Frequently Asked Questions ***(specific issues about video at the end)***

Why anonymise personal data in research projects?

Respect for confidentiality is essential to maintain trust between the public and researchers. There is a strong public interest in maintaining confidentiality so that individuals will be encouraged, for example, to seek appropriate treatment and share information relevant to it. If members of the public become suspicious of researchers, they may choose not to take part in research in future.

Wherever possible research should use unlinked, truly anonymised data. If this is not possible, the amount of personal data stored by researchers should be kept to the minimum necessary to achieve the purpose of the study. The law states that data kept should be 'adequate, relevant, and not excessive' in relation to the project involved. Personal data should be modified as early as possible in the processing of data so that some or all of those who might see it cannot identify individuals. While anonymisation may introduce delays and risks of error, even a basic coding system can provide a safeguard against accidental or mischievous release of confidential information. Sharing of identifiable data should be limited to those who have a demonstrable need to know it as part of their role in the research project.

Researchers should always consider when planning a project, when giving data to and receiving data from others and before publishing information, whether their research data may lead to the identification of individuals or very small groups. Exactly what information is potentially identifiable can only be decided on a case-by-case basis, taking into account the sample size, the way the data will be published, and all the other circumstances of the study.

I am receiving data from another organisation to use in my research. Who is the best person to anonymise the data?

Ideally, the organisation providing the data should anonymise it before giving it to you. This means you have received unlinked data, reducing (but not entirely removing) the risk that the data will be identifiable. Where this is not possible, it is better for the research team to anonymise the records than to use identifiable information.

If I remove the subject's name have I anonymised the record?

Probably not. Usually, anonymising records does not just involve removing the subject's name. If data are stored as individual data sets there is a risk that the data set could be linked to a data subject by age, postcode or medical condition. The more information included in each data set, the greater the risk of identification. Replacing a name with a pseudonym would not necessarily remove this risk.

Will removing the name and address be sufficient?

That will depend on the number of people involved in your study and where they are. If it is a countrywide study using many thousands of records this may be acceptable. However, in small communities it may still be possible to identify an individual even without their name and address, by a combination of other obvious characteristics such as ethnic origin, gender, disability, health

issues, postcode (in Britain postcodes contain, on average, 14 contiguous addresses, but some postcodes cover only a few addresses), or even gender. Similarly, cross-tabulation of data in a study with a small number of subjects could identify individuals. In general, the more characteristics there are in a personal record and the fewer people there are sharing those characteristics, the easier it is to identify individuals.

If I replace names and addresses with codes have I fully anonymised the data?

The Information Commissioner advises that any personal data that has been encoded remains personal data as defined by the 1998 Data Protection Act as long as the key for decoding it remains in existence. So if the key is in the possession of the University then you cannot be said to have anonymised the data. However, if you have destroyed the key, or another organisation is holding it and will never give you access to it, then the University believes that you have taken suitable steps to anonymise the data, provided you have taken into account the advice given in these guidelines.

Is it possible to anonymise images of faces?

Traditionally, blacking out the eyes has been employed to anonymise photographs of faces. However, the International Committee of Medical Journal Editors advises that it is highly unlikely that this successfully disguises identity. Similarly, while digital imaging can distort features, it is entirely possible that a subject could be identified by friends or family. Since complete anonymity of faces is almost impossible to achieve, **informed consent** should always be sought if there is any doubt.

What about other images?

Apparently insignificant features distinguishing marks, such as tattoos, body piercings, posture and gait may still be capable of identifying a patient to others. Informed consent, therefore, should always be obtained before taking and using pictures of individuals for the purpose of teaching, research and publication.

Do I need to worry about anonymising records belonging to people who have died?

Data Protection law does not apply to information about people who have died before their data are disclosed. However, it is possible for information about a dead person to betray information about their living friends and relatives, for example if the individual had a hereditary medical condition or transmissible disease. Care should be taken to ensure that this does not happen.

Are there specific issues about the use of video?

Yes, this is a major issue in work with Deaf People for the simple reason that sign language is visual and any record of sign language has to include a moving picture of the signer.

Where interviews are videorecorded, then

- the video records should be maintained securely,
- the consent form should specifically indicate who will view the videotape,

- the uses of the videorecording beyond the interview should be indicated (where there is to be use at conferences, seminars etc, then specific consent has to be obtained.
- The 'sell by date' should be indicated – ie when the data will no longer be used and will be deleted.

None of this, of course, anonymises the data. In order to do so, the interview should be transcribed to English (and then anonymised as above). The data can then be used in transcribed form or if video is needed, it has to be re-recorded in sign language using a Deaf model.
In the case of linguistic examples – ie where the sign language articulation is the critical aspect – then the item, phrase, sentence has to be re-recorded with a Deaf model.

Sources
This paper has been modified from one available in the Department of Exercise and Health Science web site. Their paper in turn draws upon:

International Committee of Medical Journal Editors: http://www.icmje.org/
Institute of Medical Illustrators: http://www.imi.org.uk/lawethics.htm
Videos, photographs and patient consent' by Catherine A Hood, Tony Hope, Phillip Dove: http://bmj.com/cgi/content/full/316/7136/1009
Medical Research Council: 'Personal Information in Medical Research'
'Informed consent in medical research: Journals should not publish research to which patients have not given fully informed consent - with three exceptions', Len Doyal, BMJ 1997;314;1107 (12 April): http://bmj.com/cgi/content/abstract/314/7087/1107