Interviews with residents of Cycling City and Towns (2010-2011): Anonymisation principles

Anonymity is generally a central clause in the establishment of informed consent for research participants. With quantitative data, anonymisation may often be a relatively straightforward task, as participants' personal information and identifying characteristics can be transformed into a set of numbers/codes, which retain much of their meaning, but lose their individual identity. With qualitative data however, this task is more challenging, as often a greater depth of personal information and context is recorded, and much of the meaning in qualitative research is derived from its subjectivity. Therefore the difficult balance to be struck in anonymising qualitative data is one between preserving the context and subjectivity of the data, whilst making the participant themselves unidentifiable.

There are several different types of identifier found in qualitative data which individually or collectively could lead to identification of participants. The main types of identifier are listed below:

- Names (of participants, family members, friends, colleagues, acquaintances)
- Place names (of own/family's/friends'/colleagues'/acquaintances': residences, local streets, local shops/amenities, schools/colleges/universities, landmarks, estates, workplaces)
- Personal information (age, nationality, home village/town/city, previous addresses)
- Other idiosyncratic details

The UK Data Archive (2015) explains that whilst it is necessary to find and anonymise these identifiers, nonetheless this task should be carried out in a delicate manner to retain as much of the meaning in the qualitative data as possible. This involves replacement in place of deletion where possible, and a reasoned approach to achieving a robust (but not counterproductive) level of anonymity:

"When anonymising qualitative material, such as transcribed interviews, identifiers should not be crudely removed or aggregated, as this can distort the data or even make them unusable. Instead pseudonyms, replacement terms or vaguer descriptors should be used. The objective should be to achieve a reasonable level of anonymisation, avoiding unrealistic or overly harsh editing, whilst maintaining maximum content"

Below is outlined the approach taken in anonymising qualitative interviews undertaken with residents of 12 Cycling Cities and Towns (CCTs). This outlines how each of the identifiers above was addressed and discusses other issues which were encountered in the process.

<u>Note:</u> In all cases of removal or replacement, it was standard to denote a changed term using square brackets (e.g. "I said to [husband] that he should cycle more..."). Maintaining this practice ensures that it is always clear exactly where changes have been made.

Names (of participants, family members, friends, colleagues, acquaintances)

In this case, names that could identify individuals were removed and replaced with a descriptor.

- For family members this was possible through using [husband], [wife], [son], [grandmother], etc. In other cases it was appropriate to use terms such as [male friend], [female colleague], etc.
- This form of anonymisation presented few problems, however in the rare instances that it was not possible to ascertain who was being discussed; a more generic descriptor such as [adult male] was used instead.

Place names (of own/family/friend(s)/colleague(s)/acquaintance(s): residences, local streets, local shops/amenities, schools/colleges/universities, local landmarks, estates, workplaces)

Greater discretion was necessary in making judgements about the appropriate level of anonymity for place names. The research was exploring cycling behaviour and in this context it was important

to retain as much geographic detail as possible to understand where people cycled and did not cycle.

Place names were routinely left intact when referred to by participants *in general terms*. As an example, if a participant specifically discussed themselves or others working at or attending a hospital in their hometown, the hospital's information was removed; however if they more generally discussed the same hospital as a big local employer, or as a cause of town-centre traffic congestion at rush hour, its details were retained.

- The neighbourhood of residence (i.e. suburbs in cities and towns or villages in rural areas) was not removed if it was of reasonable size. Streets and locations identified as being in the immediate proximity of participants' home, workplace or school, etc. were anonymised. Streets outside the immediate proximity were retained as they were considered useful in interpreting the cycling behaviour of the participant. The identify of locations outside the immediate proximity were retained, unless they were very specific locations not used by larger numbers of people in which case they were anonymised with additional detail as appropriate (for example: [café]).
- Large place names were almost always left in (for example: London, Chester, or Cambridge).
- Workplaces and employer names were anonymised, sometimes with additional detail as appropriate when the additional context was useful (for example: [healthcare provider]).
- Local schools and colleges which participants or other named people attended were replaced with descriptors. University names were sometimes replaced and sometimes left in, dependent on context. When it was likely that a village, containing a school that the participant or a family member attended, would contain only that one school, the village, as well as the school, was anonymised.

Personal information (age, nationality, home village/town/city, previous addresses)

Personal information including the age, nationality, and previous addresses of participants and other people mentioned were routinely replaced with an appropriate descriptor.

- In the case of age, individuals' ages were replaced with an age range; the type of range used was varied by context. Generally ages were termed as infant, pre-primary school age, primary school age, secondary school age, late teens, 20s, 30s, 40s, etc.
- Previous places where participants had lived were retained, as they were considered to be important context to understand the cycling history of the participant, but this was at the level of the town or city and not the address, street or neighbourhood.

Other idiosyncratic details

This is the category in which contextuality played the greatest part, and discretion and a reasoned approach were essential. Idiosyncratic details encountered in this case included unusual medical conditions, unique combinations of household vehicles (for example three classic cars, a camper van and a tandem bicycle), and unusual hobbies and pastimes. These details were dealt with on a case-by-case basis. Medical conditions were often termed [illness] or [medical condition], due to their sensitive nature.

Finally, combinations of the information revealed in each transcript were considered. This is because details, whilst perhaps relatively innocuous information *by themselves*, hold the potential to be identifiers in conjunction with other information. The general principle was to reduce the risk that someone reading the transcript would be able to use the combination of personal information along with other sources of information to be able to track down the participant.

Reference

UK Data Archive (2015) Anonymisation / Qualitative. Accessed at <u>http://www.data-archive.ac.uk/create-manage/consent-ethics/anonymisation?index=2</u> (24 September 2015)