

Timescapes Anonymisation Guidelines

These guidelines are intended to be used by all members of the Timescapes programme and all affiliates of that programme. They have been created from several sources: lessons learned from previous projects (particular thanks to all members of Making the Long View) that are now integrated with Timescapes, procedures and recommendations evolving at the UK Data Archive, consultation with other projects (e.g., Real Life Methods), and feedback and comments from Timescapes members and Advisory Network Members. Every effort has been made to settle on guidelines that will remain in effect for the duration of the Timescapes projects. However, because Timescapes involves the innovative integration of research and preparation for archiving, further revisions can not be ruled out. Everything possible will be done to minimise the burden of past and future changes to individual projects. Details are provided below.

File handling procedures

- Make a copy of the unanonymised file and put it in a secure location.
- Begin anonymisation on the copy.
- Files should be named clearly so the anonymised version can be identified.
- Records should be kept of who performed the anonymisation.
- Use consistent procedures especially within a single file.
- All unanonymised materials (texts, images, audio, tracking tables that link real names with pseudonyms, etc.) must be held securely, and tracking tables should be held apart from corresponding data.

What to anonymise

Exactly what content needs to be anonymised is variable across projects and must be determined by the research teams in consultation with the archiving team. Some examples are provided below for the most common forms of information that are usually anonymised. These guidelines are being used by MLV:

Names of people: describe according to significance to the respondent: ‘female / male friend’, ‘mother’, ‘father’, ‘teacher’ etc. When many different friends are mentioned it is best to assign each person mentioned a pseudonym as they can then be followed throughout the series of interviews. Letting people pick their own pseudonym can engage participants and avoid people feeling mis-identified by an unsuitable name.

Names of towns/cities/villages: describe according to the significance of the place to the respondent’s life: ‘city she grew up in’, ‘town he moved to’, ‘small neighbouring village’. In some circumstances, ‘London’ could remain (e.g., when they are thinking of going to university there? Have a day out?). Places or streets should be changed according to their relevance in the text. For example the name of a housing estate may in some cases be changed to ‘local housing estate’ but in another instance the same place may require a change which indicates its political or community significance – e.g. ‘local loyalist housing estate.’

Name of country: describe according to part of world or perhaps ‘country regularly visited by backpackers’, if relevant.

Name of school/college: ‘high school she attended’ etc.

Name of workplace: describe more generally e.g. ‘fast food restaurant’, ‘pub’, ‘shop’, ‘factory’ etc. When it comes to descriptions of departments in companies or particular sections of a workplace, use ‘department’ or ‘section’.

Nationalities: in close knit communities, talk of a different and specific nationality may easily disclose identity e.g. ‘the Americans who took over the pub’. Suggest these be changed to ‘people’.

School subjects – to be changed at discretion of anonymiser and possibly in consultation with respondents, especially young persons.

Original	Example of possible change
Street names and names of local areas, places visited	Either changed to a fictional name or described as ‘local area’ or ‘city centre street’ housing estate housing development
School names	Changed to ‘local secondary school’ or changed to pseudonym
Parent’s occupation	Possibly changed to something similar.
Specific businesses, places of work e.g. McDonalds	Changed to fast food outlet/restaurant
Places travelled to/ visited/ worked in.	e.g. Italy change to Northern European country
Names of family and friends	Either changed to a pseudonym or referred to as ‘younger brother’ ‘female friend’
Names of football clubs	May be changed e.g. to ‘English football club’ or local football club
Name of youth club etc	May be changed to pseudonym or referred to as local youth club/centre
Models of cars driven	May be changed to different model of car

Revised system for marking sensitive and anonymised text-PLEASE READ

These guidelines document an important shift from the previous (18 April version) for marking anonymised text. The previous version called for use of an XML tag “<seg>”. That system is no longer recommended and a new system has replaced it.

Timescapes recommends using the following system to indicate anonymised text. At the start of the text to be anonymised, use the punctuation marks @@. At the end of the text, use the marks ##. The reason for using these characters is that they are highly unlikely to appear for

any other reason in the text. Also, it is important to have different marks at the beginning and end of a segment to prevent the marked and unmarked sections of text from getting out of sequence should the closing marks be forgotten.

In addition to marking anonymised text, some projects may also want a system for marking sensitive text that might need to be anonymised at a later date. It is important to use a system that enables all readers (transcribers, data users and others) to distinguish text that needs to be anonymised, but has not been changed, from text that has been anonymised. For this, we propose using \$\$ to mark the start of sensitive text and ## for the conclusion of this text. This system makes it easy to distinguish sensitive from anonymised text. Thus, an example would look like this:

Sensitive: “Then I went to \$\$West Leeds High School## where I met new friends.”

Anonymised: “Then I went to @@Haisley High School## where I met new friends.”

This system adopts a high level of transparency. That is, it deliberately shows what has been changed and does not attempt to blur changed and unchanged text. This transparency will be obvious where replacement phrases are used, such as “childhood friend” (note: do not include quotations marks in the transcription), and this system makes all such changes visible. This is deliberate choice and any comment on this choice is welcomed.

This system also leaves open the decision about final presentation of the text, that is, whether or not the punctuation markers are left in versions to be disseminated outside of the Timescapes team. It will be possible to retain or remove these marks at a later date.

Procedures and costs for changing to the new system

This change has not been made lightly and we are aware of the costs imposed on projects of shifting to new procedures, re-training transcribers, and so on. We understand that there have been multiple versions of guidelines and that some transcriptions will have been done in accordance with those versions. First, if you have started a file using a different system (such as <seg>, then **continue that file in the same format.** When you begin the next file (transcript), please change and follow these guidelines. It is relatively straightforward to convert files between Word and XML formats, but it is best if there is consistency within a file.

For files that have already been edited using an older system, we (Ben, Isla and Libby) will work with each of the project teams to convert your files to the new system. There are two options: sending the files to Leeds where the rework will be done, or arranging for the work to be done at the project site with costs reimbursed. Details of these options need to be worked out. Please contact Libby to discuss your situation.

Note on using “search and replace”

Using S&R is recommended and can save a great deal of time in the tedious work of, for example, adding pseudonyms. However, it should never be done without further review and

proofreading. The MLV team, who have done extensive anonymisation, do not recommend using “replace all”. Employing a broad ‘replace all’ of an interviewee’s name with a pseudonym may mistakenly replace the names of another person with the same name as interviewee, with interviewee’s pseudonym.

When to anonymise

When to do anonymisation is also highly project specific. For much cross-sectional work (single interview at one point in time, no follow up contact), it can be efficient to anonymise at the time of doing transcription. No further disclosive information will be acquired so anonymisation can be done simultaneously with transcription. Even in this case, however, we would strongly recommend that the full interview be listened to before attempting anonymisation in case information is revealed late in the interview that might affect what content should be anonymised.

For longitudinal work, including Timescapes, it is advisable to consider doing anonymisation in phases, with more of the work happening later in data collection or even after all data have been collected. Projects may need to develop strategies to suit their specific forms and phases of data collection.

One approach would be to use minimal anonymisation for the first phase, including selected names and places. Then a second phase would be more comprehensive, after information is fully assembled and it is easier to determine what additional anonymisation is required. The MLV team recommend anonymising texts of multiple interviews after data collection has been completed. In some instances the significance of some data only emerges after a number of interviews. The more information gathered, the easier it is to make appropriate anonymising decisions.

Projects will vary in how this work is distributed among researchers, transcribers and administrative staff. These guidelines are intended to indicate how the files should look when they are ready to be submitted to the archive. Other techniques can be used within projects prior to this stage.

Tracking anonymised changes

Some level of tracking changes is essential, but the level of detail can vary. The tracking tables systems explained below was used by the Making the Long View team. Tracking tables are particularly useful if a number of interviewees come from the same site and refer to the same people, places, streets, schools, venues or events in the course of their interviews. Commonly mentioned places and people can be given one pseudonym which can be recorded in the tracking table and used across same site interviews. This can provide useful data for analysis on responses to different people or places. It makes it easier for a person reading anonymised text to keep track of frequently mentioned people or places across time as well as within sites. Where many different friends or siblings are mentioned, it also makes the text easier to read if pseudonyms are used as opposed to Friend1, Friend2, and so on. Tracking tables provide an easily accessible record of pseudonyms used and avoid confusing duplication within sites or interviews. To reduce the labour involved in this (which is fairly

minimal) the team decided to record the pseudonym change once for every page it appears on regardless of how often it appear on the page.

An acceptable alternative that is less time-intensive method is to create tables that record all changed terms, but do not record the location of changes. This is an adequate solution as it captures the key information of identifying all text changes made during anonymisation.

Example of tracking table from MLV

Anonymisation Tracking Table for 411241/ Int 3 (SAVED IN RTF)

Interview/Page	Original	Changed to
Int3 (ff2)/page		
3/1	Spain	European country
3/1	Salou	Holiday clubbing resort
3/1	20 th June	June
3/1	Kenny	Ian
3/1	Julie	Mandy

Images, audio and video

In the first instance, we are not going to anonymise these file formats. Ben is investigating options for digital alternative of audio and image files and will distribute a report when his research is complete. While this is technically feasible, there is a high risk of so damaging data quality as to render it unusable for many research purposes. For the time being, we will hope to make these file formats available either by gaining consent for their use or by placing appropriate levels of access restrictions on them.

Final note – reasonable anonymisation

In summary, it is important to reach an appropriate level of anonymity, whilst trying to maintain maximum meaningful information in the research data. Information should not be crudely removed or blanked-out, but rather pseudonyms, replacement terms or vaguer descriptors should be used. Some data that combine many difficult features: geographically specific references, sensitive and potentially harmful content, longitudinal detail that increase disclosiveness, will be difficult or impossible to anonymise in a manner that both protects the quality of the data and the confidentiality of participants. Other strategies will be necessary for such data, for example, the anonymisation of a small subset of data for illustrative purposes and might be highly valuable for methodological insights.

To summarise: the objective for all data is to achieve a reasonable level of anonymisation which is then combined with other strategies, namely consent agreements and access controls, in order to maintain confidentiality.