



Qualidata Process Guide

Qualitative Data Processing

CONTENTS

Introduction	1
1 Overview	2-5
1.1 What is data processing?	2
1.2 Definitions of data types	2
1.3 Format of data	4
1.4 Definitions of metadata types	4
2 Data processing procedures	6-17
2.1 Acquisitions review and evaluation of data	6
2.2 Processing standards	7
2.3 Format conversion	10
2.4 Guidelines on anonymisation of data	12
2.5 Digitisation	14
2.6 Creating Data Listings	15
3. Guidelines for the physical preparation of paper and audio/video-cassettes	18

LIST OF TABLES

Table 1.	Principal types of qualitative data
Table 2.	Principal types of contextual data: research and methodological information
Table 3.	Assessment of data for evaluation purposes
Table 4.	Summary of definitions of outgoing standards
Table 5.	“At a glance” definitions of processing standards
Table 6.	Ingest, preservation and issueable formats for qualitative data.
Table 7.	Mandatory and optional fields recommended for Data listing

Table 8. Example of a Data Listing for a set of individual interviews

Table 9. Example of a Data Listing for a set of focus groups

Table 10. Example of a Data Listing for a set interviews with members of families

Introduction

What is in this Guide?

This guide defines the data processing standards and procedures used by Qualidata for preparing qualitative data for preservation and dissemination via the UK Data Archive. This document falls into 3 main sections.

The first section is designed to give an introductory overview of the types and formats of data and metadata that Qualidata process.

The second section defines the processing standards that determine how datasets are processed, and provides guidelines on the procedures for format conversion, digitisation and anonymisation. On the whole, attention is paid to text, although discussion touches on processing digital audio materials.

The final section offers guidelines on how non-digital materials should be prepared and processed. It offers guidance for those processing data collections that will not be converted to digital format by Qualidata, but are to be preserved in a traditional repository. This work is not carried out in-house.

What is not covered by this Guide?

Procedures for TEI mark-up of text (to A*) are not documented in this Guide, but will be documented in the *TEI Qualitative Data Mark-up Guide* (forthcoming 2002). The procedures employed for reviewing and evaluating data, file format conversion, creating user guides, and cataloguing and indexing are documented in separate UKDA Guides: *Acquisitions Review Process*, *Translating Data Files into Different Formats*, *Documentation Processing*, *Cataloguing and Indexing Guidelines* and the *UK Data Archive Preservation Policy*.

1. Overview

1.1 What is data processing?

The way datasets are processed usually can be split into two main activities: checking and converting the dataset and generating metadata. Checking the dataset consists of carrying out activities such as checking the completeness and quality of data, the relationships between data items (e.g. interviews, fields and audio recordings), and anonymisation. Converting data consists of transferring data to a format suitable for both preservation and dissemination.

The level of processing assigned by Qualidata governs the thoroughness and complexity of the checking, converting and metadata creation activities. Category 'A*' and 'A' receive the same stringent standards of checking, but A* requires additional work to prepare them data for dissemination. Category 'B' datasets receive intermediate checks and conversion work, and Category 'C' receive only basic checks and no conversion.

The second main processing activity, generating metadata, refers to the contextual information generated during processing, such as the creation of Data Listing of interviewee biographical details that enables transcripts to be identified, or ensuring that speaker and interviewer tags and questions or topic guide headers are added. A key function of the output generated during processing is to enable users to locate transcripts or specific items in a data collection most relevant to their research.

To understand how data are processed, it is necessary to understand the principal types of data that Qualidata receives. This topic is covered in the following section.

1.2 Definitions of data types

Qualidata is concerned with research data arising from the range of social science disciplines, including sociology, social policy, anthropology, social and economic history, political science, social and human geography and social psychology. Qualitative data are data collected using a qualitative methodology - defined by openness and inclusiveness, aiming to capture participants' lived experiences of the world and the meanings they attach to these experiences from their own

perspectives. Moreover, a qualitative perspective encompasses a diversity of methods and tools rather than a single one. As a result data types extend to: in-depth or unstructured interviews, field and observation notes, unstructured diaries, personal documents, photographs and so on. The methods chosen depend on aim of the study, the nature of the sample, and the discipline. Finally, qualitative research often involves producing a large amount of raw data (audio, text, photographs for example) although the methods typically employ relatively small sample sizes.

Qualidata deals with all formats of data: in digital format, paper format (typed and hand-written), audio, video and photographic. Much qualitative data nowadays is 'born' digital in the sense that the text is word-processed and audio-visual material is in a digitally recorded form. The main types of qualitative data are set out in Table 1.

Table 1. Principal types of qualitative data

<ul style="list-style-type: none">▪ In-depth/unstructured interviews (audio/videorecordings /transcripts/verbatim recorded/summaries)▪ Semi-structured interview (audio/videorecordings /transcripts/ verbatim recorded/summaries)▪ Structured interview questionnaires containing substantial open comments (Interview notes)▪ Group discussion (audio/videorecordings /transcripts/ verbatim recorded/summaries)▪ Thematically organised interview/group discussion materials or field notes▪ Unstructured or semi-structured diaries▪ Structured/time budget diaries containing substantial open comments▪ Participant observation field notes/ technical fieldwork notes▪ Kinship diagrams/other anthropological materials▪ Case study notes▪ Minutes of meetings▪ Observational recordings (e.g. in psychology and education)▪ Psychological test data▪ Personal documents (e.g. letters, personal diaries, correspondence)▪ Press clippings▪ Photographs▪ Naturally occurring speech/conversation (audio/video/transcripts)

The most common type of data that Qualidata deals with are in depth, semi-structured and focus group transcripts. The type of data and format govern how the data are processed, as discussed in the following section.

1.3 Format of data

An increasing proportion of datasets arrive at Qualidata in word-processed format. These may be accompanied by audio-recordings, which generally are not in digital format. Other digital formats received are, for example, extracts of more structured interview material in the form of databases or spreadsheets. Whole projects are sometimes deposited in a Computer Assisted Qualitative Data Analysis Software (CAQDAS) format, for example, NUD*IST or WinMAX, although depositors are advised to export the data into a non-proprietary format. Non-digital formats include paper, audio or video cassette, newspaper clippings and photographs.

1.4 Definitions of metadata types

In order to help guide researchers through the raw data, a *Catalogue Record*, a *User Guide*, and a *Data Listing* are the three most important types of metadata prepared by Qualidata. Procedures for creating a catalogue record are detailed in the UKDA *Cataloguing and Indexing Guidelines*.

User Guides, freely available online via the catalogue record, contain useful and often unique research and methodological information about the study. They provide a context to the data and the research investigation, and if detailed enough can help the raw qualitative data be more usable by secondary analysts who have not previously been directly involved with the project.

With the creation of *user guides* in mind, depositors are asked to provide the following kinds of research related documents set out in Table 2:

Table 2. Principal types of contextual data: research and methodological information

- Original grant application
- End of award report
- Published description of methodology
- Interview schedule(s)/topic guide
- Questionnaire
- Diary format
- Observation checklist
- Interviewer instructions/ prompt cards
- Details of investigation and/or interviewers
- Communication with informants relating to confidentiality
- Written consent forms
- Matrices
- Tree diagrams
- Other background information
- Details of missing information
- Correspondence

- Speaker markers in text, typically associated with internal metadata; question or thematic markers in text; cross-reference of text to audio material.

- References to publications and reports based on the study

Researchers are encouraged to document their research and the material produced during fieldwork from the earliest stages of the project and throughout analysis. Procedures for creating *User Guides* are detailed in the UKDA *Documentation Processing Process Guide*.

The *Data Listing*, detailing the key characteristics of the data, is constructed to help users to identify particular types of interviews or transcripts (such as women of a particular age in the sample). For example, in the case of data that are based on a sample of interviews with individuals, the list would contain their date of birth, gender, employment, and perhaps geographical region, although any other key characteristic that initially defined the sample should be listed. In many ways these defining characteristics are analogous to 'variables' in quantitative datasets. The *Data Listing* also provides information about names of files and length of transcripts. Depositors are asked to collate and supply this information themselves, which, many have already done in the course of conducting fieldwork or analysing data. Procedures for preparing the *Data Listing* are set out in Section 2.6.

2. Data processing procedures

2.1 Acquisitions review and evaluation of data

Data collections at the processing stage will have already undergone a formal evaluation and had a series of initial checks. The nature of the evaluation process of data being offered to the UKDA is detailed in the UKDA *Acquisitions Review Process*. For the purposes of qualitative data, the Acquisitions Review Committee is made up of Qualidata staff plus some academic staff from the University of Essex. In summary, data are assessed for a number of issues, as set out in Table 3:

Table 3. Assessment of data for evaluation purposes

Issues Assessed	Digital data				Non-digital data		
	Text	Images	Databases/ spreadsheets	Audio/video	Paper	Audio/Video	Image
Relevance to Collections Development Policy	Y	Y	Y	Y	Y	Y	Y
Completeness	Y	Y	Y	Y	Y	Y	Y
Physical condition	Y	Y	Y	Y	Y	Y	Y
Software format	Y	Y	Y	Y			
Readable/audible	Y	Y	Y	Y	Y	Y	Y
Viruses	Y	Y	Y	Y			
Documentation is adequate to enable processing/re-use	Y	Y	Y	Y	Y	Y	Y
Confidentiality and anonymity	Y	Y	Y	Y SAMPLE	Y	Y SAMPLE	Y SAMPLE
Deposit and Licence forms signed and completed	Y	Y	Y	Y	Y	Y	Y
Suitability for digitisation					Y	Y	Y

At this stage any problems concerning confidentiality, access and re-use are logged and the depositor is re-contacted for clarification, if necessary, either by the Acquisitions staff or the Processing Officers depending on the nature of the query.

Increasingly depositors are following guidelines advocated by Qualidata so that a degree of processing has taken place before the material arrives. In all cases Qualidata processing staff go to great lengths to ensure that processing reflects accurately the nature of the original data. In cases where data arrives in a poorly-organised condition, Qualidata's processing aim is to re-instate order and meaning to the data. Ideally, researchers should, at the grant application stage, cost in where appropriate the expenses associated with preparing material for archiving: including transcription or summaries of in-depth interviews; anonymisation of data; and the preparation of a data list. This helps to ensure that data will be offered in an appropriate form, and in turn promotes the idea that the archiving and sharing of qualitative data are part of good research technique and practice.

Processing of data involves a range of activities that covers: checking, digitising and Optical Character Recognition (OCR), anonymising, organising and listing the material so that it can be accessed for re-use. The level of processing of a given study is determined by the processing standard assigned to it by the Qualitative Acquisitions Review Committee.

2.2 Processing standards

A processing strategy is agreed within the Processing Section and a single member of staff is assigned to the study. This strategy broadly defines the level of data processing appropriate to the collection via an appraisal of the condition in which it is acquired. The importance of the collection and its priority as a processing task is also taken into account.

Qualidata has defined a number of core processing standards, with A* being the highest level of processing and C, the lowest. Tables 4 and 5 below provide a summary of the outgoing standard and “at a glance” more detailed definitions of the processing tasks and expectations for the data and metadata.

Table 4. Summary of definitions of outgoing standards

A*	Data are fully digitised, anonymised, and fully web accessible Metadata are fully digitised, anonymised, rich and fully web accessible
A	Data are fully digitised, anonymised and available via download service. Metadata are fully digitised, anonymised, and fully web accessible
B	Data are digitised, anonymised and available on download service. Where data cannot be fully anonymised, may only be accessible with restricted access. Selected metadata are fully web accessible; some only accessible with data orders
C	<u>Digital collections</u> are anonymised but not available on download service. Where data cannot be fully anonymised, may only be accessible with restricted access. Selected metadata are digitised and accessible as user guides <u>Non digital collections</u> are not anonymised or digitised but transferred to another repository. Selected data/metadata may be digitised and accessible as user guides

Table 5. “At a glance” definitions of processing standards

	A*	A	B	C
Dataset dimension checks	Number of individual level files and other data types checked against documentation Missing data noted in <i>read</i> and <i>note</i> files	Number of transcripts/individual level files checked against documentation Missing data noted in <i>read</i> and <i>note</i> files	Number of transcripts/individual level files checked against documentation Missing data noted in <i>read</i> and <i>note</i> files	No checks made
Format conversion (where not digital)	Conversion to digital format Text: full OCRing with editing Audio/video: partially indexed where appropriate	Conversion to digital format Text: partial OCRing, with some error resolution Audio/video: not indexed	Conversion of text to digital format for small collections; 20% sample for large collections. No conversion of audio/video to digital format. Partial automated OCRing, but no error resolution	No format conversion. Data collection passed to other repository
(continued)				

(continued)	A*	A	B	C
			Problems encountered corrected if possible, but if not possible and the effect on data quality is not critical, uncorrected discrepancies should be clearly documented in the <i>read</i> and <i>note</i> files	
Format conversion (where digital)	Conversion to preservation and issueable formats	Conversion to preservation and issueable formats	Conversion to preservation and issueable formats	No conversion, but any format problems noted in <i>read</i> and <i>note</i> files
File naming and labelling	In accordance with the UKDA file naming and labelling guidelines	In accordance with the UKDA file naming and labelling guidelines	In accordance with the UKDA file naming and labelling guidelines	No renaming /labelling carried out unless file names contain real names
Quality - image or audio	Images checked for picture quality and cleaned up where possible. Problems documented accordingly in the <i>read</i> and <i>note</i> files Audio checked for sound quality and problems documented accordingly in the <i>read</i> and <i>note</i> files	Images checked for picture quality. Problems documented accordingly in the <i>read</i> and <i>note</i> files Audio checked for sound quality and problems documented accordingly in the <i>read</i> and <i>note</i> files	A sample of 5% of images checked for picture quality. Problems documented accordingly in the <i>read</i> and <i>note</i> files A sample of 5% audio material checked for sound quality and problems documented accordingly in the <i>read</i> and <i>note</i> files	No checks
Confidentiality	Data fully anonymised, as per Section 2.4	Data fully anonymised, as per Section 2.4	Data fully anonymised. Some data only released with restricted access	Checks made and documented. No changes made. Data only released with restricted access
Transcription checks	Text checked. Errors and all typographic errors amended	Text checked for errors and problems documented accordingly in <i>read</i> and <i>note</i> files	A sample of 10% of text checked of errors or incomplete transcription and problems documented accordingly in the <i>read</i> and <i>note</i> files	No checks made
Mark-up of digital interview material	Text marked up by theme, questions and speaker (using Qualidata TEI DTD) Questions from topic guide input where appropriate Digital audio-files linked formally but not synchronised to text	Interview text marked up with speakers and questions (using Qualidata TEI DTD)	Interview text left as deposited	No mark-up
(continued)				

(continued)	A*	A	B	C
Metadata checks	Dataset must be comprehensible in itself	Dataset must be comprehensible in association with the documentation given to users.	Visual checks on quality, action taken for systematic problems	No checks
Data Listing	Full data listing linked formally to interview text files and audio-files where appropriate	Full data listing linked formally to interview text files and audio-files where appropriate	Basic Data listing prepared not linked to interview text files	No data listing over and above what has been prepared by depositor. Available only with order for dataset.
Metadata Enhancements	Full catalogue record prepared Extensively bookmarked PDF user guides; additional related resources on dedicated web page	Full catalogue record prepared Extensively bookmarked PDF user guides; additional related resources on dedicated web page	Full catalogue record prepared Production of basic PDF user guide; <i>read</i> and <i>note</i> files with relevant user information	Catalogue record prepared only

Each dataset is assigned both an incoming category, to describe its standard upon receipt, and an outgoing category, to define the level of processing that is required. All subsequent references to the category of a dataset in this document refer to this outgoing category, since this defines the level of validation, conversion and other tasks to be undertaken.

Each item contained in the collection, such as transcript, tape, photograph, set of field notes etc. should carry a unique identifier, that will enable links to be made between these items.

The following section discuss three areas of processing: format conversion; digitisation and anonymisation.

2.3 Format conversion

A large proportion of qualitative datasets arrive at Qualidata as text files in MS Word format. Rich Text Format is a standard UKDA preservation format, and qualitative data are typically distributed in this format. Table 6 gives an overview of various format conversions for the purposes of preservation and dissemination. Note that the preservation strategy is set out fully in a separate document, the *UK Data Archive Preservation Policy*.

Table 6. Ingest, preservation and issueable formats for qualitative data.

Format	Typical ingest format	Preservation formats	Issueable formats
Digital (born digital)			
Text	Word CAQDAS	Rich Text Format (RTF) Plain (unformatted) text XML, SGML	Rich Text Format (RTF) MS Word
Databases	MS Access	Delimited ASCII with data definition statements, Tagged ACSII, XML, SGML	MS Access
Spreadsheets	MS Excel	Delimited ASCII with data definition statements, Tagged ACSII, XML, SGML	MS Excel
CAQDAS format	NUD*IST, WinMax	Rich Text Format (RTF) Plain (unformatted) text, XML	Rich Text Format (RTF) MS Word
Image (documents, photos)	JPEG, GIF	TIFF	Adobe PDF, TIFF
Audio	MS WAV, MPEG-3	MS WAV, MPEG-3	MS WAV, MPEG-3
Video	Apple Quicktime, MPEG-1, MPEG-4	MPEG-1, MPEG-4	Apple Quicktime, MPEG-1, MPEG-4
Non-digital			
Text	Paper	Rich Text Format (RTF) TIFF	Rich Text Format (RTF) Adobe PDF
Audio	Audio cassette	MS WAV, MPEG-3	MS WAV, MPEG-3
Video	Video cassette	MPEG-1, MPEG-4	MPEG-1, MPEG-4
Newspaper clippings	Paper	TIFF	Adobe PDF
Photographs	Photographic paper	TIFF	Adobe PDF

Qualidata seeks to acquire where possible data in digital format. Collections acquired that are in non-digital format assigned for enhanced data processing, will undergo either complete or partial digitisation, depending on the extent and nature of the material. The digitisation process is discussed in section 2.5.

Qualitative software packages, or CAQDAS, such as NUD*IST, ATLAS-ti and WinMax have export facilities that enables one to save a whole 'project' consisting of the raw data, coding tree, coded data and associated memos and notes. For archival purposes the raw data, the final

coding tree and any useful memos should be exported PRIOR to acquisition. Coded data are not preserved, as they cannot as yet be exported in a common non-proprietary format. Qualidata is working to encourage the development of an export standard using XML. At present, coded data are not in demand, mainly because the coding process is subjective, often geared towards specific themes, and therefore may not be applicable to the secondary analyst's topic of investigation. For larger studies, there is a stronger case for retaining coded data, in order to aid searching within and navigation through voluminous bodies of text.

2.4 Guidelines on anonymisation of data

Confidentiality checks are fundamental to data processing, and these checks are common to all categories of dataset (A*, A, B and C).

In general, no information that breaches the confidentiality of the respondent or any other person or entity must be present in any issueable format of the dataset. However, in some cases, respondent permission may have been gained for re-use without anonymity, for example for research with elites or life story material. Furthermore, in cases where data cannot be completely anonymised, data may only be released with restricted access and user undertakings, and typically only with permission from the depositor. Finally, in signing a legally binding access agreement to re-use data, users of the UKDA promise to respect guarantees of anonymity, consistent with the original investigator's undertaking.

It is important to arrive at an appropriate level of anonymisation. In some cases it can be difficult to disguise the identity of participants without introducing an unacceptable distortion into the data. Indeed, this lessens the potential for re-use. The level of anonymisation to be adopted for any one dataset depends on the nature of the study, and each case presents its own unique set of issues.

Acquisition procedures now in place encourage depositors to ensure that data being deposited conform to ethical and legal guidelines with respect to the preservation of anonymity (where requested). In practice this means that data collections should have identifiers removed in advance of deposit. A major part of processing work is to check, where present, the extent and system of anonymisation used by the depositor to ensure it has been done correctly and to the right standard. For datasets that have not been anonymised in advance, but are accepted to A

or A* standard, anonymisation work will be conducted by Qualidata processing staff. In these cases, the procedures used are always guided by and negotiated with the depositor.

The techniques of anonymisation used are:

- To remove major identifying details, i.e. real names, place and company names, street names etc. and replacement with pseudonyms where appropriate. Automated search and replace techniques are used. Additional proofreading should always be carried out in case of variations in spelling of the replaced words or the presence of unanticipated words or identifiers that may require alteration.
- To use the same pseudonyms and place names used in any prior publication by the investigator.
- To use a cross-referencing system for the pseudonyms to the original names, that will not be made available to users.
-
- For mixed methods studies, check links from qualitative data to survey data.

Where any major problems are encountered, in particular citation of third parties, e.g. slanderous or libellous comments, sections should be deleted, closed or the item withheld to make the collection issueable. In these circumstances, the depositor's permission must always be sought before this is carried out. For further information on legal issues surrounding confidentiality and copyright, see the Qualidata Web site under Depositing section.

The original data, including any 'problem' transcripts that may breach confidentiality, are always retained and preserved, but are not issueable. A note should always be made in the *read* and *note* file if information was deleted or altered for confidentiality reasons. Where complete anonymisation is not possible (but where researchers are happy to share their data for extended reanalysis), restricted or conditional access to data is given. In these cases registered users will usually need to need to gain permission from the depositor and will undertake not to quote identifying information.

Finally, as it is so time consuming to anonymise audio recordings, this procedure is not carried out. Rather, anonymised excerpts of audio 'sound bytes' are prepared.

2.5 Digitisation

Whilst Qualidata will not be undertaking any in-house preservation of paper materials, paper collections that are selected to be worthy of digitisation for all or parts of the collection, will require that the paper is prepared.

The nature and form of the raw materials leads us to consider a number of issues for digitisation:

- the suitability of the material for digitisation (paper colour and quality; type face; content);
- the proportion of the collection to be digitised;
- how the collection should be prepared with respect to physical, organisational or intellectual considerations;
- to what extent text should be made machine-readable (i.e. level of Optical Character Recognition (OCR));
- what level of cleaning up of images should be performed.

Perhaps the most critical consideration is whether images should be created and stored just as images (i.e. a graphical picture of the paper) or whether they should be converted to fully searchable text (i.e. OCRed). Due to the complex nature of qualitative data collections, which can include printed paper questionnaires and schedules with typed and hand-written comments, some materials may not be suited to OCR.

For materials that contain poor typeface, handwriting, tables or drawings, the paper is scanned and saved as Tagged Image File Format (TIFF) format. For each 'record', for example a case or transcript, all the constituent TIFFS are converted to Adobe Acrobat Portable Document Format (PDF). Using PDF, the look of the original paper can be preserved as they are similar in concept to a book displayed on screen. The PDF is then bookmarked to provide a contents page with headings giving clickable links to pages, and annotations. Security can also be applied to the files where necessary. Full details of cleaning up images, PDF conversion and bookmarking are set out in detail the UKDA *Documentation processing* Process Guide. All materials that are not destined for OCR should be photocopied prior to digitisation, and the copy anonymised with a black permanent marker pen.

For straight forward clean text, the images are or OCRed, edited and saved as RTF documents. Even with new advanced OCR software it is still time consuming to produce a perfect text

document. The OCR process is discussed more fully in the UKDA *Documentation Processing Process Guide* and in the *Edwardians On-line Project Guide*

Audio collections are only ever digitised on a case-by-case basis rather than full conversion. Excerpts of sound 'sound bytes' are prepared to add an additional dimension to the text transcript. The procedures for this process will be detailed in the next version of this guide.

2.6 Creating Data Listings

Data Listings are prepared to help users to identify particular types of interviews or transcripts (such as women of a particular age in the sample). A Data Listing lists the key characteristics of interviewees for example, such as date of birth, gender, and perhaps geographical region. Depositors are asked to collate and supply this information themselves, which, many have already done in the course of conducting fieldwork or analysing data. Any key characteristics that initially defined the sample should be listed.

For every collection, a full listing, whether of interviews, field notes or other items should be drawn up. The list should indicate where parts or sections of the data are missing, e.g. partial or summary transcripts only. Each item, e.g. transcript, should carry a *unique identifier*, and sometimes, for individuals, a pseudonym. These should be consistent to enable links to be made between interview transcripts, tapes, field notes, etc. For each interview/observation, a summary sheet or distinct row in a table should be provided indicating the biographical details of informants and information about the interview itself. These must be referenced to the data by a unique case identifier. Table 7 sets out the mandatory and optional fields recommended by Qualidata.

Table 7. Mandatory and optional fields recommended for Data listing

Interview ID	Mandatory
Date of Birth	Mandatory
Gender	Mandatory
Occupation	Mandatory
Organisation	Optional, where appropriate
Position	Optional, where appropriate
Marital Status	Optional, where appropriate
Family Code	Optional, where appropriate
Relationship	Optional, where appropriate
Number of Children	Optional, where appropriate
Ethnicity	Optional, where appropriate
Interview Summary	Optional, where appropriate
Audiotape No.	Optional, where appropriate
Place of interview	Mandatory
Date of Interview	Mandatory
Number of Pages	Mandatory
Text File Name	Mandatory
Audio File Name	Mandatory, where appropriate

Completed examples of listings are shown in Tables 8, 9 and 10 below, which highlight the minimal information requirements.

Table 8. Example of a Data Listing for a set of individual interviews

SN QXXXX							
Title: Schoolchildren Project							
Depositor: Dr. L. Smith							
List of Interviewees (Individual Interviews)							
Interview ID	Interviewee's Name(s)	DOB	Gender	Date(s) of Interview	Number of pages	Text File Name	Audio File Name
1	James	14	M	24.2.99	11	I01t.rtf	I01a.wav
2	Gary	15	M	27.2.99	19	I02t.rtf	I02a.wav

Table 9. Example of a Data Listing for a set of focus groups

SN QXXXX							
Title: Schoolchildren Project							
Depositor: Dr. L. Smith							
List of Focus Group participants							
Interview ID	Interviewee's Name(s)	DOB	Gender	Date(s) of Interview	Number of pages	Text File Name	Audio File Name
10F2	Mary, Florence and Diana	Year 10	F	3.3.99	24	FG10F2.rtf	FG10F2.wav
10M3	John, Paul, George and Ringo	Year 11	M	5.3.99	17	FG10M3.rtf	FG10M3.wav

Interviews from follow-up studies, or groups where relationships are an important part of the research design, should be logically named and cross-referenced to enable link-up between waves or family members. For example I1F would be identified as being assigned to Family group 1 and Wife of I1M, whilst IC1 would be linked to IF1 and IM1 as child of IF1 and IM1, such as illustrated in Table 10. Text and audio file names should use the same naming convention.

Table 10. Example of a Data Listing for a set interviews with members of families

SN QXXXX				
Title: Family Study				
Depositor: Dr. L. Smith				
List of interviewees in family groups				
Interview ID	Family Code	Relationship	Gender	etc ...
1F1	1	Wife of 1M1	F
1M1	1	Husband of 1F1	M	
1CF1	1	First child of 1F1 and 1M1	F	
1CM2	1	Second child of 1F1 and 1M1	M	
2F1	2	Wife of 2M1	F	

The naming scheme shown in these tables is not prescriptive as researchers often use their own systems for identifying materials. The important issue is that the naming system is logical and

adheres to the UKDA guidelines minimal information requirements and file naming and labelling standards.

3. Guidelines for the physical preparation of paper and audio/video-cassettes

Qualidata's strategy from 1994 - 2001 has been to prepare all formats of qualitative data for preservation in host archives across the UK. Accordingly, guidelines for preparing paper and audio-video collections have been established. Qualidata recommends the adoption of these steps in the preparation of such materials:

For all materials:

1. A Data Listing prepared according to Section 2.6, with added information about the physical location of items in folders and boxes e.g. Box 1, Folder 1.
2. Naming and file labelling according to UKDA name and file labelling conventions.
3. The production of electronic files for metadata, e.g. User Guides in Adobe PDF format.

For paper:

1. The use of archival quality acid-free folders and linen ties for paper files.
2. The use of archival quality photocopying paper, wherever possible.

For Audio/Video:

1. The cassette box and the tape itself is marked with an identifying reference number, using a permanent pen.
2. The cassette box is labelled with the project title, investigators names, and ID of the interviewee/recording.
 1. The security lugs are removed from the tape itself, to prevent accidental over-taping
 2. Care is taken to ensure that elements that may infringe depositors' access conditions are removed (e.g. real name labels on cassettes or cassette boxes where anonymisation is necessary), but as explained earlier, anonymisation of the recording itself is not normally practical.