

External Document

Data Processing Standards

UK Data Archive

Creator:	Data Services
Maintained by:	Data Services
Contributor:	
Version:	01.00
Controlled document No:	79
Last Amended:	2 March 2009
Review due date:	2 March 2010

Data Services
UK Data Archive
University of Essex
Wivenhoe Park
Colchester
Essex
CO4 3SQ

email: help@esds.ac.uk
Tel: +44 (0) 1206 872001
Fax: +44 (0) 1206 872003
data-archive.ac.uk

Table of Contents

Table of Contents	2
Document Control	2
Scope	2
1. Allocation of the processing standard	3
2. Quantitative processing standards	3
2.1. Processing standard A*	3
2.2. Processing standard A	4
2.3. Processing standard B	4
2.4. Processing standard C	5
3. Qualitative processing standards	5
3.1. Processing standard A*	5
3.2. Processing standard A	5
3.3. Processing standard B	5
3.4. Processing standard C	5

Document Control

Version	Notes	Last Amended
01.00	First external version	2009-01-08

Replaces or supersedes:

UKDA-DataProcessingTechniques-Internal

Review terms:

Annual

Is related to:

- UKDA-DSS-Data Processing Standards*
- UKDA-DSS-Processing Quick Reference*
- UKDA-DSS Qualitative Processing Procedures* (in development)
- UKDA-DSS-Pre-processing Checks* (in development)
- UKDA-DSS-Confidentiality Checks* (in development)

Scope

What is included in this document?

This document defines dataset processing standards currently in use at the UKDA, and the criteria used to allocate that standard. External readers should note that some of the documents referenced above may not be available for external use.

What is not included in this document?

This document does not include details of quantitative or qualitative processing procedures. Further details of procedures may be found in the documents *UKDA-DSS Quantitative Processing Procedures* and *UKDA-DSS Qualitative Processing Procedures* (in development).

1. Allocation of the processing standard

All datasets added to the UKDA collection are processed to one of four set standards (A*, A, B or C), which determine the level of work to be undertaken to prepare them for secondary use.

For both quantitative and qualitative datasets allocation of the ingest processing standard depends largely on:

- condition of the data and documentation;
- anticipated level of secondary use.

The processing standard is therefore usually allocated based on the results of pre-processing checks carried out on the dataset component files (see separate documents *UKDA-DSS-Pre-processing Checks* and *UKDA-DSS-Confidentiality Checks* (both in development)).

However, some set rules for processing standard allocation do exist:

- quantitative datasets destined for the UKDA's Nesstar browsing tool (whether part of an ESDS service series or not) will be processed to A* standard;
- if not routinely destined for Nesstar, quantitative datasets covered by the ESDS Government service are usually processed to A standard;
- quantitative datasets deposited by the Office for National Statistics (ONS) or another government department, that are not covered by ESDS Government are generally processed to A standard, with a few exceptions;
- quantitative datasets deposited by a major research centre such as the National Centre for Social Research (Natcen) are generally processed to A standard, with a few exceptions;
- if the quantitative dataset is part of a series (regular or otherwise), the processing standard allocated should be that previously used for the series - it is desirable to use the same standard in order to keep the series consistent, but if the condition of the dataset components differ from the previous deposit, the standard may be raised or lowered;
- most quantitative studies originating from academic sources (added to the main UKDA collection rather than UKDA-store) will be processed to B standard or higher;
- quantitative processing standard C is rarely used, although it may be allocated where, for example, no improvement to the materials is possible due to circumstances beyond UKDA control, or where the materials are in a software-dependent format with no alternatives available;
- qualitative datasets destined for ESDS Qualidata Online will be processed to qualitative processing standard A*;
- all qualitative datasets archived at the UKDA are covered by the ESDS Qualidata service, and are usually processed to qualitative processing standard A (see section 3.2 below);
- the vast majority of mixed methodology (qualitative + quantitative) datasets are covered by the ESDS Qualidata service, and the respective components are processed to qualitative processing standard A and quantitative processing standard A.

2. Quantitative processing standards

The following sections describe the quantitative ingest processing work undertaken for each processing standard (A* to C).

2.1. Processing standard A*

The text below describes the quantitative ingest processing work undertaken for processing

standard A*. This text (or a variant thereof) appears in the 'Read' file associated with each quantitative dataset processed to A* standard:

"The data were processed to the UK Data Archive's 'A*' standard. This is the UKDA's highest standard, and means that an extremely rigorous and comprehensive series of checks was carried out to ensure the quality of the data and documentation. Briefly, the most important procedures were as follows. Firstly, checks were made that the number of cases and variables matched the depositor's records. Secondly, checks were made that all variables had comprehensible variable labels and all nominal (categorical) variables had comprehensible value labels. Where possible, either with reference to the documentation and/or in communication with the depositor, labels were accordingly edited or created. Thirdly, logical checks were performed to ensure that nominal (categorical) variables had values within the range defined (either by value labels or in the depositor's documentation). Lastly, any data or documentation that breached confidentiality rules were altered or suppressed to preserve anonymity.

All notable and/or outstanding problems discovered are detailed under the 'Data and documentation problems' heading below."

2.2. Processing standard A

The text below describes the quantitative ingest processing work undertaken for processing standard A. This text (or a variant thereof) appears in the 'Read' file associated with each quantitative dataset processed to A* standard:

"The data were processed to the UK Data Archive's 'A' standard. A rigorous and comprehensive series of checks was carried out to ensure the quality of the data and documentation. The most important procedures were as follows. Firstly, checks were made that the number of cases and variables matched the depositor's records. Secondly, checks were made that all variables had variable labels and all nominal (categorical) variables had value labels. Where possible, either with reference to the documentation and/or in communication with the depositor, absent labels were created. Thirdly, logical checks were performed to ensure that nominal (categorical) variables had values within the range defined (either by value labels or in the depositor's documentation). Lastly, any data or documentation that breached confidentiality rules were altered or suppressed to preserve anonymity.

All notable and/or outstanding problems discovered are detailed under the 'Data and documentation problems' heading below."

2.3. Processing standard B

The text below describes the quantitative ingest processing work undertaken for processing standard B. This text (or a variant thereof) appears in the 'Read' file associated with each quantitative dataset processed to B standard:

"The data were processed to the UK Data Archive's 'B' standard. A substantial series of checks was carried out to ensure the quality of the data and documentation. Firstly, checks were made that the number of cases and variables matched the depositor's records. Secondly, logical checks were performed on a sample of 30 + 10% of the remaining nominal (categorical) variables to ensure they had values within the range defined (either by value labels or in the depositor's documentation). Thirdly, any data or documentation that breached confidentiality rules were altered or suppressed to preserve anonymity.

All notable and/or outstanding problems discovered are detailed under the 'Data and documentation problems' heading below."

Note: the '30 + 10%' sample is a minimum, rather than a rigidly observed rule; checks performed at B standard are dependent on the nature of the dataset. The Data Services team member responsible for the dataset will use their expertise to decide on the exact checks needed.

2.4. Processing standard C

The text below describes the quantitative ingest processing work undertaken for processing standard C. This text (or a variant thereof) appears in the 'Read' file associated with each quantitative dataset processed to C standard:

"The data were processed to the UK Data Archive's 'C' standard, which is the Archive's minimal processing level. A series of checks was carried out to ensure the quality of the data. Checks were made that the number of cases and variables matched the depositor's records. Any data or documentation that breached confidentiality rules were altered or suppressed to preserve anonymity.

All notable and/or outstanding problems discovered are detailed under the 'Data and documentation problems' heading below."

3. Qualitative processing standards

The following sections illustrate the ingest processing work undertaken on qualitative datasets. Further details of qualitative ingest processing procedures may be found in the document *UKDA-DSS Qualitative Processing Procedures* (in development).

3.1. Processing standard A*

- Data are fully digitised and anonymised.
- Data are accessible through Qualidata Online and UKDA.
- Some or all of the data will never have been available in digital format before.
- Metadata are fully digitised and anonymised.
- Metadata are accessible through Qualidata Online and UKDA.
- Additional work has been done to create new metadata that adds to understanding of the collection.

3.2. Processing standard A

- Data are fully digitised and anonymised.
- Data are accessible through UKDA.
- Metadata are fully digitised and anonymised.
- Metadata are accessible through UKDA.

3.3. Processing standard B

- Data are digitised at least to the level of scanned TIFF images and anonymised.
- Only major problems with data are resolved.
- Data are accessible through UKDA.
- Metadata are digitised at least to the level of scanned TIFF images and anonymised.
- Only major problems with metadata are resolved.
- Metadata are accessible through UKDA.

3.4. Processing standard C

- No checks are made.
- Data remains in the format in which it was received.
- A basic catalogue record only is created.
- Non digital collections are not anonymised or digitised and transferred to another repository.

Note: in practice most qualitative datasets are processed to 'A' standard with a select few being nominated for 'A*' (see also section 1 above). The remaining qualitative processing standards, B and C, are seldom used.