



The work is licensed under the Creative Commons Attribution-Non-Commercial-Share Alike 2.0 UK: England and Wales Licence. To view a copy of this licence, visit creativecommons.org/licenses/by-nc-sa/2.0/uk/

Data Processing Standards

Version 2.0

PUBLIC VERSION

07 October 2010

T +44 (0)1206 872001

E help@esds.ac.uk

www.data-archive.ac.uk



UK DATA Archive
UNIVERSITY OF ESSEX
WIVENHOE PARK
COLCHESTER
ESSEX, CO4 3SQ

WE ARE SUPPORTED BY THE UNIVERSITY OF ESSEX, THE ECONOMIC AND SOCIAL RESEARCH COUNCIL, AND THE JOINT INFORMATION SYSTEMS COMMITTEE

Contents

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

Scope

What is included in this document

This document defines study processing standards currently in use at the Archive, and the criteria used to allocate that standard.

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 Data Collections Processing Procedures*.

It should be noted that some of the documents referenced within the text below are not publicly available, but external readers may of course contact the Archive in case of query.

1. Allocation of the processing standard

All studies added to the Archive 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 studies and qualitative data collections, allocation of an 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 study component files (see separate documents *UKDA-DSS-Pre-processing Checks* and *UKDA-DSS-Confidentiality Checks*).

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

- Quantitative studies destined for the Archive's Nesstar browsing tool (whether part of an ESDS service series or not) will be processed to A* standard (see list in Appendix).
- If not routinely destined for Nesstar, quantitative studies covered by the ESDS Government service are usually processed to A standard.
- Quantitative studies 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 if in

suitable condition, with a few exceptions (such as the *International Passenger Survey* (IPS)).

- Quantitative studies deposited by a major research centre such as the National Centre for Social Research (Natcen) are generally processed to A standard if in suitable condition.
- If the quantitative study 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 study components received have improved (or markedly deteriorated) since the last deposit, the standard may be raised (or lowered).
- Most quantitative studies originating from academic sources (added to the main Archive collection rather than UKDA-store) will be processed to B standard or higher. However, if the study components are in suitably good condition, or a high level of reuse is anticipated, a higher processing standard may be allocated.
- Quantitative processing standard C is only used for those studies where the materials are in very poor condition with little improvement possible, or are in a software-dependent format with no alternatives available (the depositor should be contacted at the pre-processing check stage to ascertain whether extra or better materials are available).
- Qualitative data collections destined for Qualidata Online will be processed to qualitative processing standard A*.
- All qualitative data collections archived 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) data collections 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 study processed to A* standard:

"The data were processed to the UK Data Archive's 'A*' standard. This is the Archive'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 study 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 study 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 study.

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 study 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 ingest processing work undertaken on qualitative data collections. Further details of qualitative ingest processing procedures may be found in the document *UKDA-DSS Qualitative Data Collections Processing Procedures*.

3.1. Processing standard A*

- Data are fully digitised and anonymised.
- Data are accessible through Qualidata Online and the UK Data Archive
- 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 UK Data Archive
- 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 UK Data Archive
- Metadata are fully digitised and anonymised
- Metadata are accessible through UK Data Archive

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 UK Data Archive
- 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 UK Data Archive

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 data collections 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.