DOCUMENTATION INGEST PROCESSING PROCEDURES

PUBLIC

02 NOVEMBER 2015
Version: 08.00

T +44 (0)1206 872001
E sharonb@essex.ac.uk
www.data-archive.ac.uk

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

This work is licensed under the Creative Commons Attribution-ShareAlike 4.0 International Licence. To view a copy of this licence, visit http://creativecommons.org/licenses/by-sa/4.0/

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

Contents

Scope

What is in this guide?

This guide covers standards and procedures for the preparation and preservation of documentation for UK Data Archive studies.

What is not covered by this guide?

Data processing – see separate documents *Quantitative Data Ingest Processing Procedures* [CD081], *Data*

Ingest Processing Quick Reference [CD080] and *Qualitative Data Ingest Processing Procedures* [CD093]

- Administrative metadata (‘Read’ and ‘Note’ files) that provide extra information for users and describe the ingest processing history of each study.
- Conversion of study-related email correspondence to Adobe PDF format.

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. Introduction

Unlike data files, there are no ‘set’ ingest processing standards for documentation files. However, an ingest standard (A*, A, B or C) is allocated to denote the nature of the documentation materials deposited as part of a study/data collection (see separate document *Data Ingest Processing Standards* for further details*).* However, value is added to study documentation through the creation of user guides and other documentation, and the addition of bookmarks to aid navigation.

2. Documentation formats

The primary documentation dissemination format created at the Archive is Adobe Portable Document Format (PDF). As with data formats, successful documentation archiving requires a balance between effective archival preservation and the provision of documentation in popular and well-supported software formats to enable easy secondary use. While Adobe PDF is not an ideal archival format (though the PDF/A standard is developing – see section 2.1 below), Adobe Reader software is free and easily available on the web (see www.adobe.com), and most data users will be able to access it.

Adobe PDF format also has advantages in that documents are relatively difficult to edit, and so have some inherent protection against inadvertent change by the user.

Most documentation is currently deposited in MS Office formats, i.e. Word, Rich Text Format (RTF) or Excel. These formats are relatively easy to convert to PDF.

3. The PDF/A standard

Where possible, it is desirable to ensure that study folder materials are converted to PDF/A format. The Archive is moving towards using the PDF/A standard for all study documentation.

PDF/A is a file format for the archiving of electronic documents, based on Adobe’s PDF Reference Version 1.4. It is defined by ISO standard 19005-1:2005, published in 2005, and is implemented within Adobe Acrobat versions 5 onwards. The standard identifies a ‘profile’ for electronic documents that ensures they can be reproduced in exactly the same format in the future. Therefore, it requires PDF/A documents to be fully self-contained. All metadata and other information necessary to display the document in the same format each time is embedded in the file. This includes (but is not limited to) all content (text, raster images and vector graphics), fonts, and colour information. A PDF/A standard document cannot be reliant on externally-sourced information, e.g. hyperlinks. Further information on settings and the PDF/A standard is available at http://www.digitalpreservation.gov/formats/fdd/fdd000125.shtml.

The ISO standard defines two levels of PDF/A compliance for PDF files:

1. PDF/A-1a (level A compliance)
2. PDF/A-1b (level B compliance)

PDF/A-1b is the basic level of compliance, and aims at ensuring reliable reproduction of the document. Documents converted to PDF/A-1b are acceptable for archiving. PDF/A-1a is a higher level of compliance, which is harder to attain, and includes PDF/A-1b compliance plus document structure (‘tagging’), aiming to ensure that document content is also fully searchable. Microsoft Word versions 2010 onwards enable easier conversion to PDF/A-1a, which should be used where possible for Archive documentation. However, due to the diverse nature of documents deposited (which often includes PDF documents that are not PDF/A compatible, the standard is not currently rigidly enforced, but used wherever possible.

4. PDF/A and UK Data Archive data documentation

Conversion of born-digital UK Data Archive study documentation to PDF/A-1b is relatively straightforward using Microsoft Word 2010 and Adobe Acrobat version 9, but it is not infallible. It is also worth considering at this stage that the Archive’s holdings contain many older studies that include PDF documents scanned from paper, which are unlikely to become compliant to either PDF/A level.

5. Creating PDF/A compliant files in Word

Open the relevant document in Word (2010 and above). Convert it as follows:

- Select the ‘File’ menu (top left-hand tab), then the ‘Save and Send’ option from the left-hand side section of the screen.
- Select ‘Create XPS/PDF document’ from the middle section.
- Click on the ‘Create XPS/PDF’ button in the right hand section. The ‘Publish’ window will appear – check the filename and location is correct for the PDF file and then click on the ‘Options’ button on the bottom right.

From the top down, set the following options (they only need to be set once, and should remain selected for future documents):

- Under ‘Page Range’ check that the page selection etc. is correct for the file to be created.
- Under ‘Publish what’, select ‘Document’ (not ‘Document including markup’)
- Under ‘Include non-printing information’, select ‘Create bookmarks’ with the ‘Using Headings’ option. (Note that if there are no headings in the document you’re using this option will not be available – try again with a multi-page document with headings) and ensure that the ‘Document structure tags for accessibility’ is ticked.
- Under ‘PDF options’, select ‘ISO 19005-1 compliant PDF/A’.

When ready, click ‘OK’ and then the ‘Publish’ button in the Publish window.

5.1. Reading the files in Acrobat and checking PDF/A compliance

When the document opens in Adobe Acrobat 9 and above, a header may appear stating ‘You are viewing this document in PDF/A mode’. If so, this setting needs to be changed, otherwise no edits or amendments can be made to the file. To change the settings (they only need to be set once, and should remain selected for future documents):

- Go to the Edit tab on the top menu
- Choose ‘Preferences’ (bottom of the menu list).
- Highlight ‘Documents’ in the left-hand section and under ‘PDF/A View Mode’ select ‘Never’ from the drop-down list instead of the ‘Only for PDF/A documents’ that is set by default. Click OK to save.

To check PDF/A compliance:

- Go to the Advanced tab on the top menu Preflight.
- In the Preflight window, select ‘Verify compliance with PDF/A1-a’ from the list.
- Click on the ‘Analyze’ button. Usually, it should return ‘No problems found’, meaning that the file is PDF/A1-a compliant. If problems are listed, try going back and selecting the ‘Analyze and Fix’ button instead.

Files created from electronic documents via Word 2010 will usually be PDF/A1-a compliant, but there are exceptions. If PDF/A1-a compliance fails, try checking for PDF/A1-b compliance instead, using the same method but choosing ‘Verify compliance with PDF/A1-b’ instead, though this may fail too. In case of query, or if attempts to produce PDF/A-compliant documents consistently fail, please consult the Data Curation Manager.

6. Excel documentation

Documentation may also be deposited in MS Excel format. The most common examples include variable lists and codebooks and occasionally questionnaires.

Excel documentation should always be checked carefully to ensure that it contains only material that can be classed as documentation and made available via the Discover catalogue record webpage. All spreadsheets and worksheets within the file should be checked accordingly, including linked data used to create graphs, etc. If data are found in the file, the depositor should be contacted and clarification sought as to whether the data element should be removed from the file or it should be treated as data and made available subject to authentication along with the rest of the data files.

If the Excel file includes text or set formatting, or makes use of a 'freeze pane' scrolling facility, it may not be easy to create successful PDF conversions. In this case, the document may be left in Excel format, and a suitable note added to the Note file. If the file to remain in Excel format is part of the documentation for secondary users, a copy should be archived under an SN/excel/ directory. The original document should also be archived under /noissue/.

6.1. Creating PDF documents from MS Excel

Microsoft Excel 2010 and earlier files can also be converted to PDF files in a similar fashion to Word 2010 documents

Open the relevant document in Excel 2010. Convert it as follows:

- Select the 'File' menu (top left-hand tab), then the 'Save and Send' option from the left-hand side section of the screen.
- Select 'Create XPS/PDF document' from the middle section.
- Click on the 'Create XPS/PDF' button in the right hand section. The 'Publish' window will appear – check the filename and location is correct for the PDF file and then click on the 'Options' button on the bottom right.

From the top down, set the following options (they should remain selected for future documents):

- Under 'Page Range' check that the page selection etc. is correct for the file to be created (usually 'All').
- Under 'Publish what', 'Active sheet' will be selected. (This can be changed to 'Entire workbook', but this may not be appropriate for the file to be converted – conversion by sheet may be a better option - this can be decided on a case-by-case basis)
- Under 'Include non-printing information', ensure 'Document structure tags for accessibility' is ticked.
- Under 'PDF options', select 'ISO 19005-1 compliant PDF/A'.

When ready, click 'OK' and then the 'Publish' button in the Publish window.

However, there are some points to remember (and some pitfalls to be avoided):

- It is very important to make sure all columns are set wide enough in the Excel file to display all the text within them prior to PDF conversion. If this is not done, the PDF file will display the cells with truncated text.
- Large Excel spreadsheets may also cause problems in Acrobat due to the limitations of the printable area, so such conversions should be checked very carefully. Reduction to a percentage of page size is possible in Excel via the Print dialog box in order to make the sheet display on one page, but for very large Excel sheets this may render the resulting PDF to such a small size it may need a great deal of magnification to be legible.

7. Creating PDF documents from text files (RTF and plain text)

Text files (both plain text or RTF) can easily be converted to PDF via MS Word 2010. Any format editing of the file should be carried out in MS Word prior to PDF conversion. It should be possible to ensure PDF/A compliance via MS Word (see section 2.3.2 above).

8. Creating PDF documents from other word-processing software

Other proprietary formats may occasionally be deposited at the UK Data Archive (MS Works, WordPerfect, Claris Works, Open Office, etc.). Such files may be imported directly into MS Word and then converted to PDF (some editing may be required), or first exported from their proprietary format as RTF and then imported into Word for further conversion.

9. Paper (hard copy) documentation

The deposit of hard copy documentation at the Archive is becoming increasingly rare, but sometimes still occurs, meaning that the paper copy must be scanned to create electronic documentation for preservation and dissemination.

9.1. Creating Tagged Image File Format (TIFF) files from hard copy (paper) documentation

Before scanning, check with the depositor (if not already done) whether the document is available in electronic format.

If the hard copy documentation is double-sided, it should for ease be photocopied to single-sided format before scanning. Material that is primarily text should normally be scanned at 300dpi resolution, though higher resolution may be used where required. All material should be scanned into TIFF format, which is a flexible and adaptable format for handling images and data within a single file. Older studies in the Archive’s collection will have one TIFF file for each documentation page; this is preferable for archival standards in case of future file corruption, and should be the norm when processing hard copy documentation, but more recent studies may have many pages in one TIFF file. The TIFF files will be further converted to PDF, but the original TIFF files must also be archived on the Archive preservation system. Instructions on how to use the current Archive document scanner are available in the Appendix to this document.

Note: All original TIFF files for each study processed must be preserved on the Archive’s preservation system. To do so, all TIFF files produced from scanning hard copy documentation should be placed directly into a directory named after the four-digit Archive study number.

10. Creating PDF documents from TIFFS and other image files

Scanned images from paper documentation (TIFFs) and other image files supplied by the depositor (in any common ingest format, such as .jpg) are easily imported into PDF, using the ‘File’ drop-down menu and selecting ‘Create PDF’ (a multiple files version is available). If any image editing needs to be carried out to enhance legibility, it is generally easier to edit the TIFF images in a graphics program such as Paint Shop Pro or Adobe Photoshop, before conversion to PDF. If so, the editing work should be carried out on a copy of the TIFF file to avoid problems.

10.1. Optical Character Recognition (OCR)

All scanned hard copy material should be subjected to OCR, except where the text content is minimal (i.e. not scans of pictures or photographs, etc.). Such TIFFs can simply be opened into Acrobat, saved as a PDF file and then inserted into any PDF document (see below for details of merging files and moving pages). OCR may be carried out in Adobe Acrobat (depending on the version used). If the OCR option is available, select ‘Recognize Text using OCR’ and ‘Start’ from the ‘Document’ menu in Acrobat, and specify the following preferences before running the OCR:

- Primary OCR language: English (UK)
- PDF Output Style: Searchable image (exact)
- Downsample: Low (300dpi)

OCR should also be carried out on born-digital documentation for studies destined for Nesstar (see section 13 below).

11. Using the Adobe Acrobat ‘TouchUp Text’ Tool

The ‘TouchUp Text’ tool in Adobe Acrobat may be useful for limited editing of text once OCR software has been run on scanned hard copies. After selecting the icon, place the cursor over the text to be edited. A box will appear within which the text can be amended. This tool also allows very limited editing to be carried out, such as moving text along on the same line, which may be useful when pagination has gone astray. Extra lines of text cannot be added. More extensive editing will need to be done on the original tiff file using Photoshop or PaintShop Pro, as noted above.

12.Production standards for PDF study documentation

12.1. Filenames

The Archive has some standards for the naming of PDF and other documentation files, though a degree of flexibility is required and practised according to the requirements of the study. The filenames of deposited files are usually retained, though this may be decided on a case-by-case basis. The study number is normally added as a prefix, though some depositors (such as the Centre for Longitudinal Studies) prefer this is not done. All filename changes are recorded in the study Note file, and if the information is also needed by users, in the study Read file.

12.2. Document grouping

The work of the (now completed) Survey Question Bank project based at the Archive, -highlighted some useful groupings for particular types of study documentation, especially for larger studies. All studies are individual, and the groupings chosen will depend on the nature of the study documentation, but it may be clearer for users to group documents according to type rather than putting them all together in one ‘user guide’. For example, questionnaires may be grouped together, and fieldwork documents such as interviewer instructions may be grouped together..

12.3. General PDF editing operations

This section covers some of the most common procedures used in the Adobe Acrobat software.
Note: these instructions are based on Adobe Acrobat 9 Professional, and may differ from other versions of the software. See the ‘Help’ guide and documentation in other versions of Adobe Acrobat software for alternative specifications and instructions.

12.4. Amalgamating files

From the ‘File’ menu in Adobe Acrobat, select ‘Create PDF’ then ‘From multiple files’. This will show an interface which can be used to browse, select, and set the order and combination of, multiple PDF files (held together in one directory).

Alternatively, open the current PDF file at the page at which another file is to be inserted, ‘drag’ the desired file onto the open PDF file and ‘drop’ it in the main text area (not the bookmark margin). Selecting the ‘Document’ drop-down menu, ‘Insert Pages’, then choosing the file to insert, will also perform the same action.

12.5. Cropping

The cropping tool utility is useful for editing documentation pages scanned from paper hard copies. Unwanted marks from document pages, e.g. dark margins or staple marks can be removed. To use the cropping tool, select the cropping icon from the toolbar, or ‘Crop Pages’ from the ‘Document’ menu. Draw a box around the area to be kept; anything outside the box will be deleted.

12.6. Deleting/rotating pages

Choose ‘Delete Pages’ from the ‘Document’ drop-down menu. A dialogue box will appear as a final prompt before the decision is made to delete a page. To rotate misaligned pages, select ‘Rotate Pages’ from the ‘Document’ drop-down menu.

12.7. Moving pages

To move pages around within a PDF document, click on the ‘Pages’ tab at the left-hand side. The images will appear in the left-hand section. Images can then be ‘dragged’ as necessary to a different location in the document.

12.8. Adding notes to PDF files

Notes may be added to PDF documents by selecting ‘Comments’ and then ‘Add Sticky Note’ from the ‘Tools’ menu. A note can be added on any page by clicking the mouse and typing in the box that appears. Notes are sometimes used to draw users’ attention to anomalies in the documentation, for example filenames and formats referenced that differ from those available from the Archive. Ingest processing staff should be aware the ‘author’ of the note will appear as the licensed owner of the Adobe Acrobat software, which will often be the login name of the staff member. This should be amended to ‘UK Data Archive’ by changing the ‘Author’ box entry under the ‘General’ tab in ‘Properties’, which may be accessed by clicking on the ‘Options’ tab within the open note.

13. Enabling easy Nesstar text entry

Where the study in question is destined for Nesstar, certain operations performed on the Adobe PDF file can make question text entry from PDF easier in the Nesstar interface. These include saving the file in the latest version of Adobe Acrobat (9.0 or above) and running Optical Character Recognition (OCR) on the file, even when it is born digital rather than scanned from hard copy paper format. To OCR a PDF file, go to the Document menu and select ‘OCR Text Recognition’ > ‘Recognise text using OCR’ and follow the instructions onscreen.

14.Bookmarking

All PDF documentation files should be ‘bookmarked’ to aid user navigation, whether they are in the format of a single user guide, or multiple volumes. The optimum density of bookmarking is obviously content-specific and depends on the length and section division of the document. Do not ‘over-bookmark’, which can be distracting for the reader. Any necessary amalgamation of files, or deletion/movement of pages within the PDF file should be carried out before bookmarks are added.

To create a bookmark in Adobe Acrobat, select the ‘Bookmark’ tab, along the left-hand side of the visible document screen. This will open a section at that side, in which the bookmarks will be created. Choose the ‘Select’ tool icon. The word(s) required on the document page (e.g. a heading) can be highlighted. Once this is done, press ‘Ctrl+B’, and the selected text will appear as a bookmark in the Bookmark pane. Alternatively, a new bookmark may also be created using the ‘Edit’ drop-down menu and selecting ‘Add Bookmark’. The resulting bookmark text may be edited as necessary. However, if a blank bookmark is required for text to be typed in, press ‘Ctrl+B’ and an ‘Untitled’ bookmark will appear, which may be edited as necessary. Insert the bookmark text in ‘sentence case’ (upper case initial letters for all nouns, lower case for conjunctions, etc.). If numerous bookmarks have been inserted, remember to save the file regularly. Many files are now deposited in PDF format, often with bookmarks already added by the depositor. In many cases, these bookmarks have been created by the Adobe 'Distiller' plug-in, and may need - editing to conform to Archive standards. If the editing required is significant, recreating the bookmarks from scratch may be the easiest option.

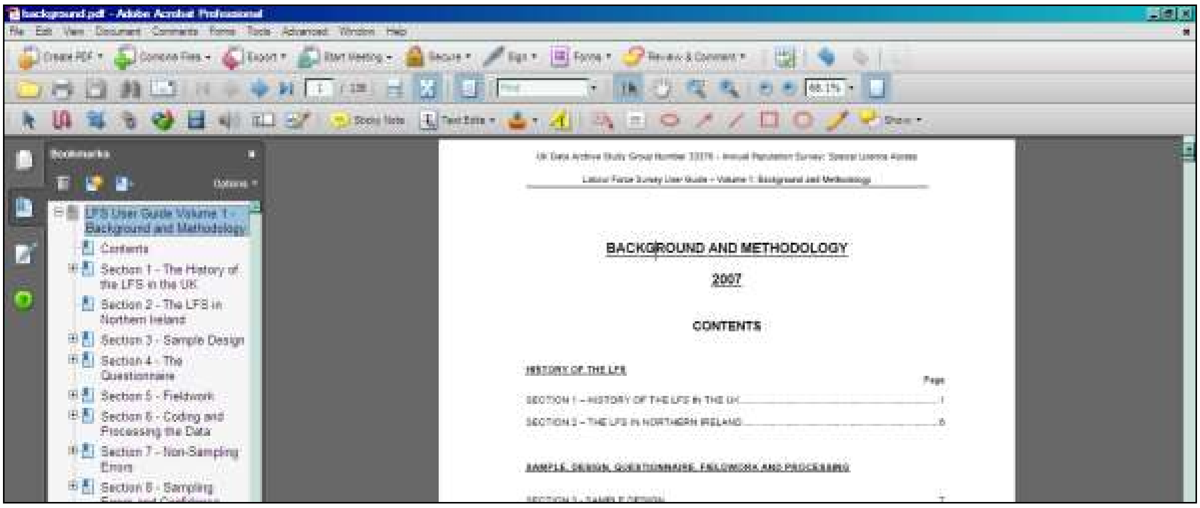
14.1. Setting bookmarks to 'Fit Page' magnification

All bookmarks should be added with the document page set at 'Fit Page' magnification, which is the Archive standard. Note that if a bookmark is added whilst the page is magnified, that is how the bookmark will be set. Bookmarks that open with pages in an assortment of magnifications make for a very unprofessional appearance, and should be avoided. Files may arrive from the depositor with bookmarks set like this, in which case all bookmarks must be reset to 'Fit Page'. Remember to proof-read bookmarks thoroughly: there is no facility within Adobe Acrobat to spell-check them.

14.2. Hierarchical bookmark structures

For documentation associated with quantitative studies, bookmarks should be 'nested' within a hierarchical structure. As a general rule, if the document contains a comprehensive contents page, the titles of the bookmarks and their hierarchical structure should reflect that. When all bookmarks have been created, the hierarchy should be closed leaving just the top bookmark visible (normally 'User Guide'). However, there are some special cases where the depositor's preference is that the bookmark hierarchy is left extended with major section bookmarks visible.

For documentation associated with qualitative data collections, it is preferred that the bookmarks hierarchy is also left extended with major section bookmarks visible. See section 8 below for details.



LFS User Guide Volume 1 - Background and Methodology
- Contents
- Section 1 - The History of the LFS in the UK
- Section 2 - The LFS in Northern Ireland
- Section 3 - Sample Design
- Section 4 - The Questionnaire
- Section 5 - Fieldwork
- Section 6 - Coding and Processing the Data
- Section 7 - Non-Sampling Errors
- Section 8 - Sampling

BACKGROUND AND METHODOLOGY

2007

CONTENTS

HISTORY OF THE LFS

Page

SECTION 1 – HISTORY OF THE LFS IN THE UK 1

SECTION 2 – THE LFS IN NORTHERN IRELAND 6

SAMPLE, DESIGN, QUESTIONNAIRE, FIELDWORK AND PROCESSING

Example of a hierarchical bookmark structure

14.3. Setting the final PDF document properties

Document properties should be set so that the PDF document opens with the bookmarks panel visible, and so that the correct document metadata (now routinely read by internet search engines such as Google) settings are present. In addition, if a Document Title is added and the file saved in Acrobat version 9 or later, the latest zipdiss12 labelling program will automatically pick up the title to add to the 'makelbl' program file list.

To do this, go to the 'File' menu and select 'Document Properties' and add the following settings:

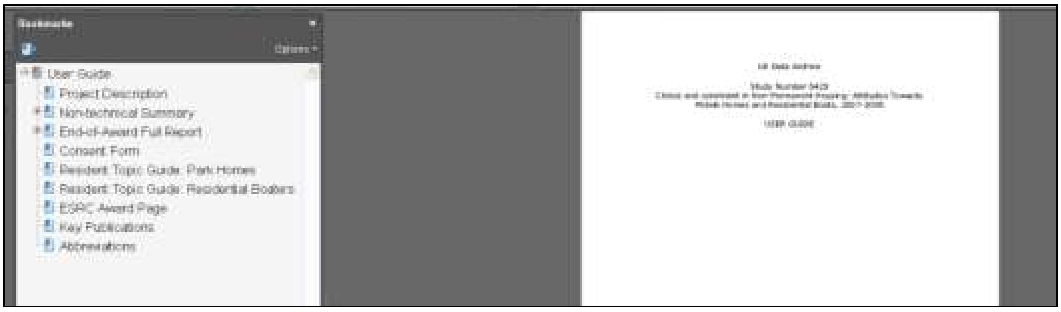
Under the 'Description' tab, add a suitable document title in the 'Title' box, such as 'General Household Survey Questionnaire 2005'. Under 'Author', add 'UK Data Archive', or if the PDF file has been created by the depositor prior to deposit, the original organisation name should be used. Current practice is to leave the subject field blank, though this may change as metadata becomes increasingly important in the future.

Under the ‘Initial View’ tab, to ensure that the document opens with the correct magnification and that bookmarks are displayed, the following items should be checked:

- Show: ‘Bookmarks and Page’
- Page Layout: ‘Single page’
- Magnification: ‘Fit Page’
- Window options: ‘Centre Window on Screen’.

For qualitative data collections where a user guide has been created from a combination of files, the bookmarks hierarchy should be left extended with major section bookmarks visible. This helps users to see at a glance the contents of the user guide, as given in the example from study 6429 below. The header page information described section 14.2 below may also be seen in situ here.

Bookmark hierarchy example (from study 6429)



- User Guide
  - Project Description
  - Non-technical Summary
  - End-of-Award Full Report
  - Consent Form
  - Resident Topic Guide: Park Homes
  - Resident Topic Guide: Residential Boaters
  - ESRC Award Page
  - Key Publications
  - Abbreviations

15.Adding headers to documentation (branding)

15.1. Background

Google and similar web searches now return results that include searchable PDF documents. If a search results in a ‘hit’ on a document in the Discover catalogue on the UKDS web site, it may not be obvious to the searcher that the document is part of a study held at the UK Data Archive.

Therefore, where possible, an informative header should be included to ‘brand’ the first page of each PDF file as part of the UKDS study documentation. (The principle is similar to the addition of a header to qualitative interview transcripts.) This policy applies whether the documentation consists of one file or multiples. Note that the presence of an Archive header does not imply any claim on copyright, which remains with the original document copyright holder. It is merely a branding device used to aid web searchers and identify a component of the Archive collection.

15.2. Exceptions

There are some occasions where branding with suitable headers may not be possible or desirable.

- **Where the depositor has specified particular requirements for documentation,** including bookmark formats and the retention of file names (see also section 9.2 above)
- **‘Locked’ (password-protected) PDF files where no editing is possible.** Some PDF files may have been created with a high level of security, which limits user options. Unless the password to ‘unlock’ them is available, neither headers nor bookmarks can be added. This may occur when, for example, the PDF file has been created on behalf of the depositor by a secondary survey contractor. All reasonable efforts should be made to obtain either a Word version of the PDF file, an ‘unlocked’ copy of the PDF file, or the password. If this proves fruitless, it should be recorded in the study Note and Read files as appropriate that the file was unable to be edited and so it has not been processed to the usual Archive standard.
- **The depositor objects to the presence of an Archive header in their documents.** Most depositors will not mind a discreet header, but if a request to remove it is received, this should of

course be done. The files should be edited accordingly and replattered. The depositor's objection should be recorded in the Note file for future reference, and the Data Curation Manager informed.

- **Software used for documentation file creation limits or precludes header creation** (e.g. Windows Help (.hlp) files).
- **For qualitative data collections, a separate header page may be used** rather than the addition of a header to the first document page. However, this depends on the nature of the collection and is only likely to be of use where several small disparate documents are to be combined into a user guide. An A4-size header page template is available for ingest processing staff to use. (See also section 13.2 above on bookmark hierarchies.)

15.3. Adding headers to Word documents

Where documentation is received in Word or RTF, the addition of a header may be done relatively easily. First, make a copy of the original to work on (the depositor's original should remain as received without header). The copy file will need to be saved once the header is added before it can be converted to PDF, which is another reason for working on a copy rather than the original. If under any circumstances the original is used, ensure that the header is removed from it once the resulting PDF file has been checked.

These instructions are for Word 2010 and current versions – other versions may differ:

- Open the document in Word
- Click the header and the 'Header and Footer' tools tab will open
- To add the header to the first page only, tick the 'Different First Page' box.
- In the header area of the document, add the study number and title in the following format:

UK Data Archive Study Number 5640 – General Household Survey, 2005

Style Specifications:

- Font: Verdana, normal (i.e. no bold or italics)
- Size: 8pt (large enough to be legible and small enough to be unobtrusive)
- Justification: centred, and at the top of the header box, so that it is suitably distinct from the text in the body of the document.
- Ensure that there is a spaced hyphen between the study number and the study title.

This style has been agreed as an Archive standard, and should be used. However, it is possible that this style may cause display problems depending on the document. For example, if the header text appears too near the first line of the document text, the page margins can be adjusted (see Page Setup box above) to move it further away. Also, if the document design means that there is a dark background at the top, or other headers are present, the information may be added as a 'Footer' instead.

When the header has been added, save the document copy, and make any other adjustments deemed necessary before creating the PDF file according to normal documentation procedures. The PDF copy must be checked to ensure that header transfer has been successful. If so, the Word file copy may be deleted.

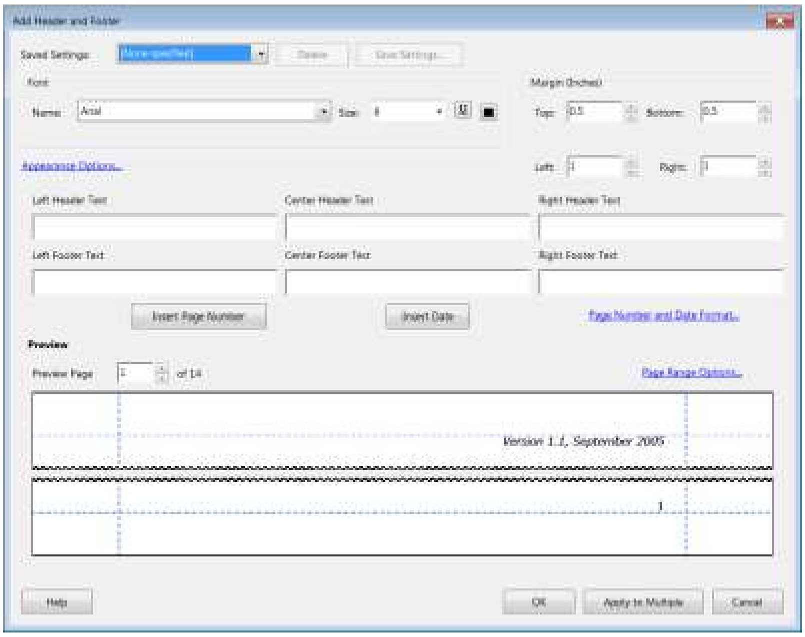
15.4. Adding headers to Adobe PDF documents

Where no Word or RTF copy is available, the header may be added within the PDF document.

Note: these instructions are based on Adobe Acrobat version 9 Professional (it is possible to add headers in versions 6.0 Standard onwards). The procedure is as follows:

- Open the PDF file
- Go to 'Document' on the top menu, then choose 'Headers & Footer', then (usually) 'Add' from the selection of 'Add', 'Update' or 'Remove'. If there is an existing header or footer, a further prompt will

then appear to ‘Add new’ or ‘Replace existing’. Usually, ‘Add new’ will be the option of choice. Once this is selected, the following box will appear:



Add Header and Footer

Saved Settings: [None specified]

Font

Name: Arial Size: 8

Margin (Inches)

Top: 0.5 Bottom: 0.5

Left: 1 Right: 1

Left Header Text

Center Header Text

Right Header Text

Left Footer Text

Center Footer Text

Right Footer Text

Preview

Preview Page 1 of 14

Version 1.1, September 2005

1

- Select the following font settings:
  - Font: Verdana, size 8
  - Margin (Top): 0.5 (may be adjusted to 0.2 or 0.1 if another header is present)

Add the required header text into the ‘Center Header Text’ box to ensure central justification. The text wrapping does not matter here, as it will appear in one line on the page.

Click on the ‘Page Range Options’ hyperlink, select ‘Pages from’, and adjust the range to show ‘1’ to ‘1’, so the header will appear only on the first page). **Note that ‘All Pages’ is the default option - if this is not reset, the header will appear on every page and will need to be removed and then the process started again to add the correct header to the first page.**

Click OK and then view the header as it appears on the first page. If alterations need to be made, repeat the process. The position on the page can be adjusted within the ‘Margin’ (Top) section as necessary.

Once the header is correct and added to the first page, check the display in the Adobe PDF window. Then, you can either click OK for one file, or use the ‘Apply to Multiple’ button (new in Adobe 9) to add the same header to the first pages of other PDF documentation files. Follow the onscreen instructions to select files and do this – ensure that the ‘Do not save files’ button is **not ticked** or the headers will be lost.

Note: it is possible to save settings in Adobe Acrobat based on a basic header creation, for addition to subsequent documents. Obviously, the study number and title will need to be edited according to the study, but settings such as header on the first page only, and font and font size can be saved so that they do not need to be reset on each occasion.

15.5. Adding headers to other documentation formats

Excel

Header information will only display in printed Excel files, not in the normal spreadsheet view, which rather defeats the object of adding a header for these purposes. As it is currently unlikely that a

Google web search will pick up an Excel file, adding a header should depend on the nature of the Excel file. If it possible to identify the file as part of Archive documentation in some form, it should be done. For an example of how to do this, see the Excel file 5640_changes_2004_to_2005.xls included in the GHS 2005 documentation.

Powerpoint

Documentation may occasionally include Powerpoint presentations, for example, the Family Resources Survey (FRS). It may be possible to create PDF files from Powerpoint, but the same principles should be applied to these as to the Excel files, i.e. header addition should depend on the nature and contents of the file.

HTML

It may not be possible to add header information to .html documentation pages, but again this depends on the nature of the file. Procedures for dealing with .html documentation are currently in development.

Software package files (e.g Windows .hlp)

Header addition is unlikely to be possible.

16. Special Licence Data Dictionary documentation

This section covers Special Licence Data Dictionary processing only. For full guidance on all aspects of Special Licence study processing, see document *Quantitative Ingest Processing Procedures* Note that this guidance is likely to change during late 2013/early 2014 as new ingest processing scripts are developed.For standard End-User Licence (EUL) SPSS-format studies, top-line variable frequency distributions for each data file are usually added to the catalogue record from the DDI-XML files created by the data processing script. However, according to Office for National Statistics (ONS) preference, this is not done for ONS Special Licence Access (SL) studies. For consistency, this practice has also been extended to SL studies sourced from depositors other than ONS. However, in response to user requests for more information on the contents of SL studies as opposed to their corresponding EUL versions, it has been agreed that the Archive Data Dictionary file(s) for SL studies should be converted to PDF so that they will be visible with the other study documentation files via the catalogue record. This will provide variable information, though not frequencies. Therefore, for SL studies, the Archive Data Dictionary file created in RTF format by the processing script is converted to PDF format and added alongside the other PDF documentation. Note that a study with more than one SPSS data file will have separate data dictionaries for each file.

The generation/conversion procedure is as follows. For details of processing script functions and outputs, and other data processing details, see the documents *October 2005 Processing Script Procedures (not currently a controlled document due to ongoing script development) and Quantitative Data Processing Procedures:*:

1. Generate the Archive Data Dictionary/ies via the data processing script as normal.
2. Convert the resulting RTF Archive Data Dictionary file(s) to Adobe PDF, and name it/them as follows: ####_<filename>_Archive_Data_Dictionary.pdf (where #### is the study number, and <filename> is the individual file name that will differ for each separate SPSS file).
3. If the study contains more than one file, the resulting PDF converted data dictionaries may be combined into one PDF document. However, if the combined file size will be >10mb, they should be left as separate files.
4. Add header and bookmarks as for other PDF documents, and set to ‘Fit Page’, etc. as normal (see section 6.3 above).
5. Add the document(s) to mrdoc/pdf/ alongside the other documentation files, and delete the mrdoc/allissue directory and contents, as the RTF data dictionary file(s) is/are no longer needed.

6. When generating the study .lbl file, use the label 'UK Data Archive Data Dictionary' for the PDF data dictionary. (if there is more than one file (see point 3 above), individual filenames may be added to the label for clarification).
7. For new editions, at the study plattering stage, ensure that any /mrdoc/allissue directory from the previous edition of the study is deleted.

17. Creating index files

Some older studies in the Archive collection with multiple documentation volumes may have had a hyperlinked ‘Index’ file to the documentation created. This is no longer necessary now that the Archive has moved to a more intuitive filenaming convention for documentation. It is also not desirable for the PDF/A documentation preservation standard, as hyperlinks to external files may not be permissible. However, a certain degree of flexibility is permitted so that an index file may be created for those studies with a lot of documentation in multiple formats (e.g. the *Family Resources Survey*). If it is felt necessary to create an index file for a new study or series, please consult a senior member of the Ingest Services team for advice.

18. Administrative metadata: Read and Note files

In addition to the documentation supplied by the depositor and processed as described above, additional documentation (metadata) is also created by the Archive. This consists of the html format ‘Read’ and ‘Note’ files, created via the Calm database to accompany each study processed at the Archive. Procedures for creating these files are covered by the document *Creating Read and Note files in Calm* (not currently controlled due to ongoing database development).