

Explanatory Notes for Dataset Transfer Form (Data Owner)

Introduction

The information that you provide on this form will supplement that provided by your Departmental Record Officer (DRO). He/she will have received a form (at the same time that you received this) focussing on the 'legal' aspects of the transfer including information on any relevant legislation, access/closure requirements and copyright. The DRO may contact you for information to assist him/her in completing that transfer form. There is one particular area which is covered by both forms; this is because it is necessary to ask the DRO to state whether there are parts of the dataset which need to be closed (for example by redaction/closure or aggregation), but it is unlikely that he/she will, in all circumstances, be able to provide a list of the relevant fields; we have therefore asked for that information on this form (see question 2-9).

There may be parts of the form which have been pre-completed from information provided by the DRO/TNA. Please amend or add to the information as appropriate.

Dates Wherever possible provide relevant dates in responding to questions. If you know the exact date (ie day, month, year) please provide it, but if not specify dates at the level of detail of which you are aware (eg month/year, year, or perhaps Season/year).

An explanation of some of the terms used in the transfer documents is provided in the Glossary - **Help & Advice**: Section (ii).

1 Dataset Description & Access

- 1-1 Dataset title** The name/title of the dataset as provided by your Departmental Record Officer (DRO) liaising with your Department's contact at The National Archives (the Client Manager). If the title is not that used within your Department for this dataset and/or you feel that it does not adequately describe the dataset, please amend the title.
- 1-2 Data to be transferred** Please provide any information additional to the Dataset Title, which is recorded on page 1 of the form, to describe the specific data to be transferred. In the case of a defunct system where *all* the data is to be transferred, little if any additional information will be required here. In other cases, what is being transferred may be a snapshot, a particular instance of a series, or a subset of a larger system (such as information relating only to England and Wales from a system which covers the whole of the UK). So, for instance, in the case of an annual survey please state the year; in the case of a snapshot state the information that identifies the snapshot, which could be the date on which the snapshot was taken. Please amend or add to the text in this box so that it adequately describes the data that makes up this particular transfer. Please include if possible the dates of the dataset *contents*. For example, the data may come from a survey which was carried out in January 1998, but the data may relate to the financial year 1 April 1996 - 31 March 1997.
- 1-3 Dataset Description** This may be pre-completed from information provided by the DRO/TNA, but please add additional details; including, for example, whether the dataset is based on a survey, in which case provide, if possible, the size and the response rate. Include the purpose of the dataset and a summary of how the data was used within your division of the Government Department/Agency; and if possible how it

is used elsewhere within or without the Department. Examples might be that it was used for in-house workload monitoring, or that it was used to answer enquiries from the public and local authorities on land use. If the data has not yet been used, please state this (ie do not leave the box blank). If you think that another unit has some claim to ownership of any of the data in the system, please provide details, including a contact name.

1-4 Description and dates of data input State which methods were used to enter the data into the computer system, for example using punched cards, mark-sense readers, optical character recognition (OCR), transfer from another computer system, or on-screen forms (ie 'standard' data entry by staff keying the data in via a terminal or PC). Wherever possible provide dates when the methods were used, particularly if data was transferred from another computer system. If the data is from a defunct system, please indicate the date when data was last added to the system and/or when data on the system was last modified.

1-5 Description of how the data was validated State which methods were used to validate the data, for example internal validation (eg by repeating questions, consistency checking by comparing with responses to related questions), dual-entry verification, validation programs (on-line or batch), comparison with output from related systems (specify which), comparison with output from independent sources (specify which), or feedback from inspection of later analyses.

1-7: Data Protection Acts Indicate whether the dataset is subject to the Data Protection Acts. If the dataset is subject to the Data Protection Acts indicate whether subject access (in the sense of the Act) is permitted and note the identifying information required eg National Insurance Number.

If subject access is not permitted indicate under which of the areas defined in Part IV of the Act exemption is claimed. The exemption areas are:

- National security
- Crime & Taxation
- Health, Education & Social Work
- Regulatory Activity
- Journalism, literature & art
- Research, history & statistics
- Information available to the public by or under enactment
- Disclosures required by law or made in connection with legal proceedings
- Any other exemption (please specify)

2 Technical Details

2-1 Software Please give details of *all* the software used. There could be a number of software packages and / or bespoke software involved. For example Microsoft Access version 2.0 could be used for entry of the data onto an Oracle database (version 7.1), with the SPSS statistical package (version 7.5) being used for data analysis. Please provide the version numbers, where available.

If the software was developed (in-house or contracted out) specifically for this dataset, please list the programming languages used, in addition to any packages.

2-2 Computer/Operating System State the type of computer/operating system for example IBM or compatible PC with MS-DOS version 4, IBM mainframe with MVS version 4.3, VAX with VMS version 5.3, ICL 2900 with VME SV292+ , SUN Unix with Solaris 2.3, AIX Unix version 3.1.5.

- 2-3 Previous system** Please provide this information if any or all of the data was converted from another system. If the previous system 'belonged to' another part of your Department or to a unit outside your Department, please provide appropriate details including the name (both then and now if possible) of the unit owning the system.
- 2-4 Quantity of data** Specify your answer in kilobytes, megabytes or gigabytes. An approximate figure is acceptable. If at this stage you have no idea of the size, specify the number of records and if possible give an indication of the average size of record (for instance average number of characters). It would also be useful if you could provide figures for the total numbers of files and the approximate number of records in each.
- 2-5 Physical media** Include the numbers of items eg 5 CDs. The agreement between TNA and the UL specifies the types of media which NDAD is obliged to accept for the transfer of data, and these are listed below. If you anticipate needing to use any media not in the list below, you should make your requirements known **immediately** - either to your Client Manager or to the NDAD contact (at ULCC) for this dataset - the names are listed on the front of this form.

Please provide as much information as possible, eg for 9-track tape state the density (6250 or 1600). The 'approved' media are:

- 9-track reel tape (6250 BPI GCR encoded and 1600 BPI phase-encoded);
- CD (High Sierra format or ISO 9660 format);
- 90mm (also known as 3.5") floppy disk and 5.25" floppy disk, in MS-DOS, MacOS, VMS or multi-volume TAR format (as produced by GNU tar);
- Exabyte 8200 and 8500 format 8mm tape cartridge. (These are the 2.2 Gbyte and 5 Gbyte version of the Exabyte. No other densities are supported.);
- DDS2/3 DAT (4mm) tape cartridge. (These are the 4 Gbyte and 12 Gbyte DAT tape types. DDS2 specifies an international standard for the compression mechanism used to get 4 Gbyte on a tape. Some older DAT drives use proprietary schemes to get more than 2 Gbyte on a tape – these are NOT accepted by NDAD. You can usually force a tape drive to write a tape without using its proprietary compression scheme.);
- 3480, 3490 and 3490E 0.5" cartridge tapes. (Also known as 'square tapes.' These hold 200 Mbyte, 400 Mbyte and 800 Mbyte respectively);

- 2-6 Overall file structure of the media** Please provide sufficient information to enable us to read the data off the media. For files supplied on floppy disk, CD etc, it is sufficient to state the type of system from which the files were transferred, for instance PC, Mac, Unix. If the files are compressed it would also be useful to mention the compression technique used eg Winzip. More detail is required for tapes, so this could mean that you need to specify the block size as well as any archiving/compression format (eg tar, VMS backup). For example, the ideal level of detail provided would be 'each tape is unlabelled and contains a single tar-format archive per year's survey, using a blocking factor of 20 (10 Kbyte blocks)'.
- 2-7 Data format** Because of the variety of software packages used by Government Departments over the years, it is not possible to specify a single format in which the data should be transferred. A general rule, particularly for older packages, might be that

data should be transferred to NDAD in a non-proprietary format, and this will usually be the format intended for export to another system. (Almost every piece of software includes a mechanism to do this). However, provision of data in the package's 'native' format will usually mean that more information on the data is accessible (for example access to the digital data dictionary); and therefore if the data is in a current or recent version of a widely-available package (such as MS Access, Foxpro, SPSS), it may in fact be preferable for the data to be sent in the package's own/native format. Please state the formats in which you could send the data, indicating if appropriate your preferred format; we will contact you to discuss the format in which the data will be transferred.

Comma separated files (also referred to as CSV: Comma Separated Values) are often used to facilitate transfer of data between applications. The data are in the form of a table, with each field separated by a comma; text may be enclosed in double quotes.

Fixed length files are those where each field has a specific length (for example the 'County' name field is 30 characters long, whether or not all 30 are used); this means that each record of a particular type (eg staff names and addresses) has the same length.

Binary files are those where data, for example numeric data, are held in a compressed'/ binary format.

2-9 Legislation/Access Government Departments should advise NDAD of any concerns regarding access to sensitive data. Under the Freedom of Information Act the default thirty-year access rule for public records will no longer apply. The dataset will be considered **open to the public**, unless an FOI exemption is identified. The DRO will be expected to provide the detail of such an exemption, using the DRO Transfer Form. If such an exemption is applicable, then various options are available for closing data:

1. Closure of an entire dataset.
2. Closure of one or more tables within a dataset.
3. Closure of selected fields within a dataset.
4. Aggregation of data to a higher level.
5. Any combination of the above.

In option (3) for example, information which could identify individuals will be removed from the version of the dataset which is available to the public. In option (4), the data can be averaged or aggregated in some way (by date or by geographic area, for instance) to provide a useful research resource which nonetheless preserves individual or commercial confidentiality.

If the transferring Department is of the view that some form of closure or redaction of the data is necessary, it should be noted that this work will be carried out by NDAD after the data has been transferred. Data should therefore be provided in its "raw" form so that any data that is to be closed or aggregated can be permanently preserved. A version of the data containing aggregated fields or excluding the closed fields will be made available for research. If you are not certain at this stage whether confidential data, needs to be protected by closure or aggregation, either list the fields that contain confidential data or give some indication of the type of data involved eg names, addresses, telephone numbers. Following transfer of the data, NDAD will report back to you with confirmation of the details and field names involved.

3 Documentation and Further Descriptive Information

- 3-1 Documentation** You are asked to indicate here the types of documents which are available and which could provide information, background and context to the dataset - these would include documents transferred with the dataset to form part of the Archive, documents held by the Department but which cannot be deposited with NDAD (possibly because they are still in use) and documents held elsewhere (for instance published reports, or documents held by Departmental or specialist libraries).

Further guidelines on the types of documents which could usefully be transferred to NDAD or made available (either via loan or photocopies) are provided in the **Help & Advice** (i) Documentation.

- 3-4 Data Dictionary** It is essential that a list of the fields that form the dataset is supplied. The information provided should normally include all of: field name, description/explanation, data type (eg character, integer, date), size (eg number of characters), validation. Further explanation of the technical information that should be provided is included in the **Help & Advice** (i) Documentation.

The data dictionary may be part of the system or it may just be a paper list of fields/their attributes from the system and/or user documentation.

- 3-5 Coded fields** In many datasets, there are fields where the data is encoded, for example a number is held instead of a county name, or a list of diseases is represented by alphanumeric codes. The meanings of the codes may be held in another table within the dataset or they may only be available on paper. To make use of the data, researchers need to know the meaning of any encoded data. If the information in the coded data fields is available in a standard publication, provide details of this publication (title, publication date etc.) on a separate sheet of paper if necessary. If an explanation of all of the codes is not available at the time of completing this form, please indicate this by annotating the form.

- 3-6: Digital media** The media on which the data was transferred to NDAD can be returned to your Department (in which case the data will be erased prior to despatch). If you do not want the media returned, it will be disposed of securely (according to the specification agreed with TNA).
- 3-7**