National Digital Archive of Datasets (NDAD)

NDAD Transfer Procedures: Help & Advice

(iv) Advice for IT Staff

1 Introduction

The Public Record Office (PRO) preserves the records of central government departments and their agencies, acting as the repository of the national archives for England, Wales and the United Kingdom. What it has done for paper records for many decades, it now must also do for computer-based records. NDAD is fulfilling this role for the PRO.

As a member of IT staff involved with government IT systems, there are certain tasks which you will need to perform for any dataset that is selected (jointly by the PRO and your Departmental Record Officer (DRO)) for transfer to NDAD. This document explains what you need to know, what you need to do, and who to contact.

2 Objectives

- to preserve the data - by transforming it into a standard, non-proprietary format
- where possible to preserve the structure (as well as the content) of the dataset as used by the Department
- to specify/describe the files and fields which form the dataset - in order to enable NDAD software to provide access to the data
- to provide sufficient metadata and contextual and background information to allow researchers to use the data effectively; this includes how the data was collected and validated
- to provide, in the catalogues, historical information on the original IT system and its use within the Department; including perhaps how the capabilities of the system limited the ways in which the data was used

In order to meet these objectives, the Department:

- supplies the data in a format which NDAD can process and on approved media
- provides a data dictionary and, where appropriate, a data model
- provides explanations of the meanings of any encoded data
- provides copies of appropriate system and/or user documentation
- provides background and contextual information/documentation on the data, the organisation, the IT systems etc.

3 Overview of Transfer Procedures

As a member of staff in the IT department (or an external IT contractor), you may be required to assist the DRO and/or Data Owner (the person who has a say as to who is allowed access to/use of the data) with some of the following steps in the process of transferring data to NDAD:

Step 1 Completion of the transfer forms

The forms are designed to facilitate the transfer. The DRO or Data Owner may require your assistance in completing the more 'technical' information on the forms.

Step 2 Further liaison prior to transfer

As well as assisting with the completion of the transfer forms, you may be required to help provide additional information or documentation needed for the transfer process.

Step 3 Transfer of the data to NDAD

The Department prepares the data for transfer in the format, and on the media, specified by your Department on the transfer forms. It is likely that you will be involved at this stage.

Step 4 Provision of further information to NDAD

Following transfer, you may be required to provide further facts, background and contextual information to NDAD to assist in processing of the dataset, and for inclusion in the catalogues.

The bulk of the remainder of this document explains these steps in more detail. At the end of the document, you will find a section entitled Frequently Asked Questions, which covers some specific points in more detail.

Step 1 Completion of the Transfer Forms

NDAD sends out two separate forms for the Department to complete and return. The first form, which goes to the DRO, is concerned mainly with legal aspects of the transfer and with the Department's access/closure requirements for the dataset; whereas the second form, for the Data Owner, deals with some of the more technical considerations. It is likely that, as IT representative, you will be asked to assist with the technical questions on the Data Owner form. Notes are issued with both forms and you should refer to these when completing the forms.

It should be borne in mind that the NDAD transfer forms have to cover a wide range of datasets, from the very complicated to the straightforward, from large statistical systems to small PC-based systems, from bespoke systems set up in the 1960's to modern systems entirely based on packaged software. In addition, a form may cover a number of years of data and/or a number of 'generations' of the system. Perhaps a 'tall order' for any form! You may therefore find that not all questions are relevant to your dataset or there may be a number of answers that you need to provide to a particular question. A separate sheet can be used to clarify responses to specific questions and/or to provide additional information.

The information on the transfer forms is required in order to ensure that:

- the data is transferred in a format which can be processed by NDAD (see section 3.1 below)
- sufficient metadata is available, and is provided to NDAD, to enable the data to be processed
- background and contextual information are provided for inclusion in the catalogues - for the benefit of future researchers; this will include the meaning of any encoded data
- relevant documentation is supplied with the dataset

Step 2 Further liaison prior to transfer

As well as assisting with the completion of Transfer Forms, you may be required to provide more detailed information or documentation. Please refer to the *NDAD Transfer Procedures: Help and Advice (i) Documentation*, which lists in more detail than is included here the technical information/ documentation required:

- A data model – a description or diagram explaining the system structure, including all inter-relationships between files and identifying any key fields
- Name of each file, description and content of file, description of key fields
- Name of each field, description, data-type, size, validation information, whether mandatory field, missing/unknown values, audit trail, cross-field validation

Please note that it is the *information* that we need and the above items do not necessarily need to form a separate document; such information may be embedded in the system.

The types of technical documents (either paper or digital) which are likely to be vital for transfer with the dataset include:

- System documentation
- User manuals
- Supplier's documentation

Please note that documentation can include digital documents as well as paper. (Wherever documents are available in digital format, please supply them as such - in addition to avoiding unnecessary scanning, this results in the provision of a better quality digital image and/or text to the public). In the case of paper documents, these can be loaned to NDAD if the Department needs to retain them rather than deposit them in the Archive with the data. Alternatively, if it is impossible to transfer an original document, even temporarily, a photocopy would be acceptable.

Step 3 Transfer of data to NDAD

Once NDAD has received both sets of completed Transfer Forms, the preparation and transfer of the data can take place. This step usually involves the IT representative from the Department and/or an IT contractor.

It would be very useful if, in the course of preparing for the transfer to take place, you could make NDAD aware of the following:

- Convenient or inconvenient times for the transfer to take place
- Any potential technical problems you can foresee, for instance in preparing and/or transferring the data and/or essential documentation
- Contact details of any other personnel who might be able to assist, e.g. by providing technical advice about the current/original implementation or use of the system

3.1 Preparation of the data

At this stage please ensure that the data and documentation are as described on the Transfer Forms (i.e. the same number of files, on the same media, and in the same format as stated on the forms). If there are differences and these have not been discussed/agreed with NDAD, please ensure that NDAD is made aware of these differences, and the reasons for them, prior to transfer of the data.

If there are problems or delays in preparing the data, or if you have any questions whilst transferring the data files to the chosen media, please contact the appropriate data specialist at NDAD, either directly or via the Help Desk on 020 7692 1212, or e-mail support@ndad.ulcc.ac.uk

Please see the *Explanatory Notes for Dataset Transfer Form (Data Owner)* - question 2-5 - for a list of the **media** on which data can be transferred to NDAD.

The *Explanatory Notes for Dataset Transfer Form (Data Owner)* - question 2-7 - provide guidance on the **formats** that can be used for transfer of the data but see also the **FAQ** below.

Bear in mind that our objective is to retain as far as possible the structure of the data as held in the system used by the Department. So if the database is held in a number of separate tables, these should be transferred as separate files. If a particular field contains a code (and the meaning of the code is picked up from a lookup file), it should be transferred as a code (along with the lookup table) rather than being translated before transfer. In general, data within a file should be transferred in the order in which it is held (where that concept has any meaning), for instance this may be the order in which the data was entered rather than transferring it sorted into the order in which it is included in reports.

3.2 Sending the data to NDAD

The Department can now begin packing the material to send to NDAD. Detailed information on packing and transport is provided in *Help & Advice: Packaging & Transport*, but in particular please remember:

- Label all media, using only the correct labels which have been supplied with the media
- Always activate the write-protect tabs, or remove the write-ring on open reel tapes, before packing
- Complete a Transfer List, listing all the items contained within a parcel/box, and include this form within the parcel/box
- Always protect media in transit by packing in purpose-designed containers. Only use rigid containers (i.e. not padded envelopes), and ensure media are enclosed in suitable packing material, leaving at least 80mm at all points between the media and the outside of the container

If required, data can be sent to NDAD via e-mail. However, please note that this would not be suitable for large datasets and that a limit is imposed on the size of a file that can be sent via e-mail. For ULCC, this is *currently* 10 MB (i.e. you cannot e-mail a data file of greater than 10MB to NDAD without splitting the file, and this should be avoided except in exceptional circumstances). Your organisation is likely also to restrict the size of file that can be sent - you will need to check this limit before undertaking to transfer a dataset via e-mail. These limits do not apply to data sent via FTP and this method may be suitable for small to medium-sized datasets (datasets up to a few hundred MB).

*Please note that confidential data, in particular any files containing data that will be closed for a period following receipt by NDAD, should **not** be sent via e-mail.*

Step 4 Provision of further information to NDAD

There may be a need for NDAD staff to contact you after transfer has taken place; either to clarify responses to questions on the transfer forms or to fill in gaps in the information provided. Of course, the more comprehensive the information provided at the earlier steps, the less we will have to bother you at this stage! If there are particular times or periods when you would find it inconvenient for us to contact you for further information, please let us know.

4 FAQ (Frequently Asked Questions)

(i) Can we/should we transfer the data in xxx format? Or should data always be exported to one of the standard recommended formats (eg CSV, Fixed Length)?

NDAD can receive data in the ‘native’ format of some software packages, and in these cases it is usually preferable for the data to be transferred in this format because it will result in more information on the data being accessible to NDAD (for example the digital data dictionary). Currently data in the following packages can be transferred in their native format (but please ensure that the Transfer form states the *version* of the package in which the data is held by the Department, in order that NDAD can verify *before transfer* that there should be no technical problems):

- Microsoft Access
- Foxpro
- SPSS
- Dataease (some versions)

If ‘your’ software is not listed here but it is a current or recent version of a widely-available package, then please consult NDAD (either the data specialist allocated to your dataset or the Help Desk on 020 7692 1212, or e-mail support@ndad.ulcc.ac.uk).

However, the general rule, particularly for older packages, is 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.

There are particular challenges in the case of systems consisting of, or including, a GIS (Geographic Information System); the format in which they are transferred will always need to be discussed in advance with NDAD. NTF (National Transfer Format) is the preferred format.

(ii) Why do you need to know about data entry and validation?

Information on how the data was collected, what methods were used to enter the data onto the system and how it was validated (at data collection, at data entry and subsequently), is in some cases essential for researchers when using or evaluating the data. Please note that if, over the years that the system was in use, more than one method was used, each method should be listed with an indication of when it was used.

(iii) Why do you need to know such detail of the system’s hardware and software if the data is to be transferred in, for instance, comma-separated format, and there is no attempt to preserve the original software?

Details of the hardware and software (including the operating system) used in the current system are needed in order to judge the best format and method of data transfer. In addition, information regarding previous hardware and software will be of interest to researchers considering how the data was originally used within the Department. If more than one type of hardware and/or software was used (either consecutively or concurrently), please provide details of each.

(iv) I am concerned about the confidential data within the system; this should not be released to anyone outside the Department and, besides, could be of no possible interest to researchers/is not in my view a ‘public record’.

All data used by a Government Department in the course of its administrative or policy work is a ‘public record’. Although in the spirit of open government the PRO wishes to make government data available, it can be kept closed for up to 100 years from the date of creation. In general, once a dataset has been selected for preservation by your Departmental Record Officer and the PRO there is no scope for debate, but we do suggest that you make your concerns known to your DRO as it may affect the frequency of any future transfers. More importantly, you should make your DRO aware immediately if you believe that there is a statute-bar on transfer of the data outside your organisation.

(v) How are the Department's obligations under the Data Protection Act affected by the transfer of the dataset to NDAD?

If the Department retains the data, it has to continue to register the dataset under the DPA. Following transfer, the PRO will also be registered as an owner, although the purpose registered will be different. If the Department is no longer retaining the transferred information, it can opt for the PRO to take responsibility for subject enquiries. Any other considerations, for example subject exemptions, are not affected by the transfer to NDAD - but it is essential that NDAD is informed of exemptions and other pertinent factors.

The DPA seeks to protect individuals' privacy by controlling the way in which their personal information is processed - whether by government or by the private sector. The new act does allow personal data to be preserved and made available by archives, but it regulates the way in which this can be done. Although there is no reason why a dataset subject to the DPA should not be transferred to NDAD it is essential that NDAD is made aware of the data protection status of any dataset to be transferred.

(vi) We only have a single paper copy of the system (or user) documentation and need this within the Department; what are the options?

Options would include one or more of the following :

- Loan the documents to NDAD so that they can be scanned
- Provide photocopies
- If electronic copies of some of the documents are available, transfer these as digital documents

(vii) The system and user documentation are available in electronic format - should we print them out for transfer to NDAD?

No. Whatever documents are available in digital format should be transferred in that format (this avoids unnecessary scanning and results in the provision of better quality digital images and/or text files to the public). The documents transferred to NDAD should be the versions used by the Department; if the paper and digital versions differ and both are or were used, it may be appropriate for both versions to be transferred.

(viii) The information about the original system (or on ….) will take some time to get together; should we retain the transfer form until all questions can be answered?

If the information to answer a particular question will not be available for some time, note this on the transfer form, indicate who will provide the information when the Department has obtained it, and return the form to NDAD.

5 Access to NDAD

In addition to preserving the datasets, NDAD provides a range of services to the public so that, subject to any closure requirements, the datasets may be searched and viewed by users, or copies of the data and/or accompanying documents provided on request.

See the NDAD web-site in order to try out the facilities for yourself and/or see how we use the information that you provide in the NDAD catalogues: **http://NDAD.ulcc.ac.uk/**