

Related documentation

NDAD staff will work with the Departmental Record Officers (DROs) and the National Archives (TNA) to establish guidance on what material should be selected for preservation or transfer in addition to the computer data itself. The agreement between the University of London Computer Centre (ULCC) and The National Archives makes provision for two types of accompanying material (either of which may be in paper or digital form, or a combination of both):

- a. Material (known as 'Dataset documentation') which will form part of the archive, and thus part of the accession, but which is not itself part of the dataset, for example: departmental publications produced using the data; internally-produced user manuals; user group minutes and other feedback from users of a dataset; and registered files relating to the planning and implementation of a dataset. (Explanations of some of the document types listed in Question 3-1 of the Transfer forms are included in **Help & Advice** (ii) Glossary of Terms Used).
- b. Non-archival material (known as 'Accompanying documentation') which will help NDAD staff deal with the process of accessioning and conversion, but which will not form part of the permanent archive itself and can be treated as a loan. This might include supplier's manuals, notes on the dataset format or system, or even just instructions for reading a tape. Although some of these items may be considered working papers, and would not be formally regarded, or catalogued, as public records, they can be invaluable for helping with the interpretation of data. If material in this category is to be returned to a department, the Department should make this clear at the time of transfer to NDAD.

It is appreciated that not all of this documentation may survive, particularly for older datasets, but anything which is available can be of use. It is possible for documentation to be closed for a fixed period if a department does not consider that it should be open immediately on transfer. Copies of documentation are acceptable if a department feels it necessary to retain the original e.g. it is the only copy of a manual.

Registered files which have not yet been appraised may contain important documentation on a dataset. It would be helpful if these or related documentation within them could be considered for transfer (either permanently or on loan) with the dataset. In some cases it may be appropriate to supply NDAD with photocopied extracts, e.g. where a file predominately relates to another subject but contains some material which is of use in interpreting, or providing context to, a dataset.

Documentation can be in a variety of physical formats: in addition to paper, they could be on microfiche or film or be digital documents such as word processed files. See **Help & Advice** (ii) Glossary of Terms Used for a definition of 'digital documents'.

Although any available documentation should be transferred with the dataset, it is likely that discussions following transfer will identify the need for further documentation or establish its existence.

Technical documentation required

Whilst it is not possible to be prescriptive (the type of information required will differ, at least to a certain extent, according to the type, size and complexity of system), the following can be used as guidelines on the minimum information usually required. Related areas (e.g. whether a file/field is closed or requires some form of anonymisation) are not covered here.

In addition to the general information on the system [including software e.g. package/version, hardware e.g. make/model of computer & operating system, format in which data will be transferred to us, all of which are covered by specific questions on the Dataset Transfer form (Data Owner)], detailed information is required on each file and all fields within each file. (A file is a collection of information treated as a unit by the computer. A file will usually contain a related collection of records and each record would consist of fields for individual data items).

The Department should aim to provide the following (but note that most, if not all, of this information will be present in the system documentation if the system was 'properly' documented and this documentation was kept up-to-date):

- i. Data model -description and/or diagram showing the structure of the system including the inter-relationships between files [Note that the fields acting as the links between the files (i.e. the 'foreign keys' where a field in one file is the key of another file) should be identified]. A dataset may consist of a single file in which case clearly the description of the structure of the dataset will be much shorter!
- ii. For each file:
 - Name of the file (i.e. the file name used by the Department)
 - Description, including the purpose/content of the file
 - Where appropriate (e.g. in the case of a relational database), the field(s) that form the unique key of the file
- iii. For each field:
 - Name of the field (i.e. where available, the field name used by the Department)
 - Description/explanation (if the name of field is not self-explanatory or if there is additional information available)
 - Data-type (e.g. character, date, integer, real number, logical)
 - Size (number of characters; if a real number held as characters, number of decimal places); although, for instance, a field may be held and/or transferred to us as a variable length or binary field, it is useful to know what was the maximum number of characters/digits that could be entered by the Department using their data input screen/program)
 - Validation e.g. the acceptable range for a numeric field; a list of the valid entries (e.g. must be one of Sunday, Monday, Tuesday, ...Saturday); whether the data entered must match one of the entries in a 'look-up' file
 - Whether the field is mandatory
 - Any particular value used to indicate that the field value for that record is missing or unknown
 - An audit trail [e.g. whether the field was used throughout the period covered by the dataset (it may have, for instance, been introduced half-way through the year and therefore the first 6 months records do not contain data in this field) or whether it is not used at present but will be in future]
 - It would be useful also for any cross-field validation imposed by the system to be identified (e.g. the date in the field 'driving test passed' must be later than or equal to the date in the '17th birthday' field; or there should be no data in certain field(s) if the answer to a previous question was 'no')