



COLLECTIONS DEVELOPMENT POLICY

PUBLIC VERSION

03 NOVEMBER 2011

Version: 03.00

T +44 (0)1206 872572

E susan@data-archive.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

1. Background and Context.....	2
2. Scope of Collections	3
3. Criteria for Evaluating Datasets	3
4. Collections Development Strategies	5
5. Methods of Acquiring Data Collections	5
6. Procedures for Acquiring Data Collections	6

Scope

The Collections Development Policy provides an overview of the selection and appraisal criteria applied to the UK Data Archive holdings. It outlines the principles by which the Archive develops these data holdings to meet the needs of its stakeholder communities.

1. Background and Context

Since its establishment over 40 years ago the UK Data Archive has acquired, preserved and managed data and other electronic resources for the purposes of research and teaching. The Archive has also proactively promoted and disseminated these resources as widely and effectively as possible. It has acquired data from academic, commercial, and public sector sources within the United Kingdom and abroad. The collection encompasses a significant range of data relating to society, both historical and contemporary. The Archive has striven to ensure that the materials it acquires are suitable for informed use. The UK Data Archive takes as its primary user community social science data users within Higher and Further Education in the UK, though best efforts are made for all users.¹ It requires that all deposited materials are documented to a minimum standard and where possible to a far higher standard. The Archive acquires data for four central purposes:

- archival preservation;
- secondary use and analysis for research;
- teaching and learning use;
- replication and validation of research.

The UK Data Archive has an obligation to the communities that it serves to ensure that, as far as possible, these communities have access to the materials they require. Over the last two decades there has been a dramatic change in technology and an increase in the range and type of electronic materials that have been, and are being, created. This has impacted upon the work of the Archive in several ways. First, there is a greater diversity and quantity of material that places an increasing and exceptional burden upon the UK Data Archive resources. Second, the increasing use of the world wide web provides opportunities for organisations and individuals, including the Archive, to distribute their own materials. Furthermore, since 2005 the UK Data Archive has been designated a Place of Deposit which brings additional service provision and a widened Collections Development Policy. In response to these factors the Archive has developed the Collections Development Policy which is flexible, responsive and recognises the increasing role that the provision of information, as well as data, will play in the future. This policy, therefore, reflects not only the acquisitions policy of the Archive but also the policy for access to data.

The UK Data Archive Collections Development Policy is guided by the criteria outlined below. Data and other digital materials which fall outside these criteria may still be accessioned if they are of exceptional merit and can be handled within current resources. The Collections Development Policy of the Archive is under frequent review in the light of changing technologies, and the shifting requirements of its user community.

¹ The vast majority of the Archive's current users are staff and students at UK Higher and Further Education institutions, though some central and local government staff, commercial users and personal researchers are amongst its users.

Scope of Collections

The UK Data Archive collects data, information and other electronic resources of long-term interest and use across the range of social science and historical disciplines. They are acquired to support research and teaching activities in the UK and elsewhere. Materials of interest fall into four key areas:

- *data and electronic resources for research*: these are data that are suitable for informed use in a variety of research settings;
- *data and electronic resources for teaching and learning*: these are datasets that are accompanied by purpose-written teaching materials;
- *replication data and electronic resources*: these are collections of materials, (data, computer programs and instructions, and related outputs) necessary for the replication of published or unpublished research;
- *data and electronic resources which have a statutory obligation to be made available to the public*: these are data which fall under the remit of the relevant National Archives' Operational Selection Policy (OSP30).²

The UK Data Archive seeks to identify and acquire material within the following areas:

- *Discipline coverage*: at the broadest level, data and other electronic resources relating to society, in particular data about individuals or groups of individuals. This includes strategic social science and economic datasets, e.g. unemployment statistics and major household surveys.
- *Geographic coverage*: data across a broad geographic coverage focusing on the UK and cross-national datasets but including material from other countries where appropriate and in particular where these provide opportunities for comparative research e.g. European data.
- *Temporal coverage*: there are no restrictions on temporal coverage, although historical accessions are usually acquired through the History Data Service.
- *Time series and panel data*: data are sought which create or add to a time series and/or panel survey.
- *Thematic coverage*: data which assist in the creation of a coherent body of materials relating to a particular discipline or field of enquiry e.g. health, nutrition, etc.

In addition, the UK Data Archive will seek to acquire material:

- at the specific request or recommendation of a user or group of users;
- when the data collection has been fully or partially funded by the ESRC or when data have been obtained from another source using ESRC funds;
- when the data collection at least partially fulfils the subject criteria above and the costs of ingest, preservation and dissemination are paid for in advance;
- for replication purposes;
- accompanied by appropriate teaching materials.

2. Criteria for Evaluating Datasets

To ensure that the UK Data Archive continues to build a collection of value, criteria are used to evaluate potential datasets acquired by or submitted to the Archive.

- Ensure they fall within the 'Scope of Collections' (see Section 2 above).

² www.nationalarchives.gov.uk/documents/osp30.pdf

- Assess their content, long-term value and the level of potential interest in their re-use. Factors influencing this evaluation include:
 - the geographic and/or temporal scope is significant;
 - the subject coverage of the data is broad and may be of interest across the social science and humanities disciplines;
 - the data are not generally available in any other form e.g. paper;
 - accession into the Archive makes the resource more accessible;
 - the dataset adds to or is made more valuable by existing holdings, in particular where it fits into an existing series;
 - the dataset fills a gap in the existing holdings;
 - there is research and/or teaching activity in the subject area covered by the data;
 - the data are such that their continued access would otherwise be threatened.

- Determine if they may be viably managed, preserved and distributed to potential secondary users. Factors influencing this evaluation include:
 - the data are of a type with which the Archive has expertise or may easily obtain expertise or expert advice;
 - the data format can be converted to suitable dissemination and preservation formats;
 - the level and quality of documentation reach an appropriate standard to enable a secondary analyst to make informed use of the data. We welcome datasets that are documented to Archive standards as outlined in the Research management and documentation guidelines.

- Determine if there is another archive, data centre or institutional repository more appropriate for dissemination, curation or preservation of these data.

There are also criteria for rejection of data. These are for guidance only and the UK Data Archive may be prepared to accept data that fall into one or more of these categories after discussion and agreement.

The criteria for rejection are as follows:

- there are problems with the sample size, the data are very localised or are of peripheral interest for social scientists;
- the documentation is insufficient to enable understanding and further analysis of the data;
- the data would be more effectively dealt with by another organisation or institution;
- there are insurmountable legal obstacles e.g. rights management issues have not been, or are unlikely to be satisfactorily resolved before ingest;
- full use of the data would not be possible without infringing legislation, e.g. Data Protection Act;
- ethical issues remain unresolved before ingest;
- the depositor wishes to impose unnecessarily stringent access conditions or wishes to place an indefinite embargo on use. Unless this is unavoidable and the data are judged to be of long-term value, the UK Data Archive may be willing to accept material for preservation-only subject to payment of the appropriate charges.

3. Collections Development Strategies

The UK Data Archive undertakes a varied and diverse range of activities to identify data collections.

- An awareness-raising programme encourages and promotes the deposit of collections of interest through a variety of activities. These may include workshops, conferences, and training events.
- Individuals, groups and institutions creating data resources are identified and encouraged to deposit their materials with the UK Data Archive. The Archive ensures that they are contacted at an appropriate point in the data creation process and provided with sufficient explanatory material to facilitate deposit of well-documented, high quality data and documentation materials.
- The Archive works to facilitate and influence, where appropriate, funding bodies, researchers, and other groups involved in the data creation process to encourage funding and creation of materials suitable for secondary analysis and to make provision for access and long-term preservation.
- An agreement is held with the ESRC whereby its grant holders are required to offer their electronic materials for deposit with the UK Data Archive. The Archive, with the agreement of the ESRC, contacts grant-holders to discuss deposit of digital materials arising from their funded research project.
- Agreements are held with many of the other major funding bodies for the social science disciplines. The Archive, with their agreement, contacts grant-holders to discuss deposit of digital materials arising from their funded research project.
- Extensive links with government departments and with the Office for National Statistics (ONS) have been established resulting in the acquisition of many datasets sponsored or collected by the ONS. The Archive works to extend and enhance these links to ensure continued access to these data and to identify and acquire new sources of data and information.
- Links with the National Archives are in place to allow identification and acquisition of datasets which fall within the remit of OSP30.
- Good working relations have been founded with a number of non-profit-making public and charitable bodies and commercial companies specialising in data creation and collection. The Archive works to extend and enhance these associations to ensure continued access to the data collected and to identify and acquire new sources of data and information.
- A number of memberships and 'data exchange agreements' are held with other digital archives. Materials are made available to the UK higher and further education communities through these agreements. The Archive seeks to identify other archives and data centres holding materials of interest and to negotiate for similar data exchange agreements.
- Materials of interest are actively sought through publications, conferences and through regular searching of web sites.
- Teaching materials are actively sought, working with the teaching community to identify and acquire suitable teaching materials.
- The Archive seeks to develop collaborative partnerships and projects which may result in data creation by its partners.
- Contact is made with relevant journals and users of data in order to identify, acquire and facilitate access to collections of materials necessary to the replication of published or unpublished research.
- Material available through the world wide web is identified and the Archive explores options for providing gateways to these distributed resources.

4. Methods of Acquiring Data Collections

The UK Data Archive strives to provide access to the increasing and varied range of electronic resources that fulfil the scope and criteria outlined above. Data are obtained, or made available, in several ways.

- *Direct deposit by individuals and organisations:* the UK Data Archive undertakes a programme to increase the holdings of the Archive according to the objectives outlined in the Scope of Collections (see Section 2 above). These data are accessioned into the UK Data Archive and made available through the usual Archive procedures. The Archive's self-archiving repository (UKDA-store) provides a means for data creators to make available datasets and related information at a lower standard of documentation for the short to medium term. The UKDA-store service provides a catalogue of material which can be selected for the permanent Data Collection.

- *Collaboration with other agencies:* the UK Data Archive seeks to identify, and to provide access to, distributed collections. The data in these collections are unlikely to be held physically at the Archive but will be made available through it. Exchange agreements with other data archives and centres fall into this category.
- *Information provision:* the Archive recognises that it has limited resources for the acquisition of datasets. In addition, there is a proliferation of data resources of interest to social scientists that are managed and made available elsewhere, particularly through the World Wide Web. To this end the Archive is exploring the option of introducing a programme of discovery of, and mechanisms for pointing to, other electronic data collections of interest.

5. Procedures for Acquiring Data Collections

The UK Data Archive has established procedures for acquiring data.

- *Identification:* identifying potential acquisitions and seeking information on the contents of datasets. The Archive identifies potential datasets through a variety of mechanisms:
 - links with funders;
 - publication trawls;
 - conference attendance;
 - user groups, workshops, seminars;
 - links with data creators;
 - contact by depositor;
 - availability through UKDA-store;
 - recommendation by the Advisory Committee.
- *Negotiation:* the UK Data Archive contacts data creators and undertakes a process of negotiation to offer a variety of services including:
 - evaluation of a dataset to ensure that it meets the criteria outlined above;
 - advice and information on the deposit process;
 - advice and discussion on contracts, licences, and access conditions;
 - advice on the appropriate formats for deposit and the method of deposit;
 - advice on the preparation of appropriate documentation and acceptable formats for discussion.
- *Guidance for depositors:* the UK Data Archive believes in providing the widest range of guidance and information on the benefits and procedures for depositing material, on best practice and on relevant standards for data and documentation creation. Publications currently available include:
 - notes and forms for depositors which provide explanatory information on depositing material with the Archive, including acceptable formats, preferred methods of data delivery, and the metadata requirements;
 - guidelines for documenting datasets, which detail the range and type of information which should be included with all digital materials.
- *Legal framework:* the UK Data Archive provides a legal framework within which materials may be accessioned into the Archive, including:
 - a deposit agreement and licence form which confirms the rights and obligations of both parties and offers an opportunity for depositors to specify the conditions under which access may be given to third parties;

- discussion of copyright and intellectual property rights to ensure that the data creator has cleared all necessary copyright permissions;
 - discussion of legal and ethical issues to ensure that the data creator has received sufficient clearance for the Archive to disseminate the resource;
 - where necessary, the UK Data Archive negotiates for licence agreements with third parties to enable the Archive to distribute the material for the purposes of research, learning and teaching.
- *Administration*: the UK Data Archive's Acquisitions Review Committee administers the acquisitions and access procedures by:
 - maintaining a database of potential acquisitions which records details of the contact name and address, brief details of the data materials and a full record of the contact made and follow-up procedures undertaken;
 - contacting potential depositors in the initial stages of a project which includes an element of data creation and providing them with information which informs them about the Archive and the procedures for deposit;
 - maintaining regular contact with potential depositors in order to ensure the smooth accession of their material into the Archive and to answer any queries they might have;
 - providing help and additional information where needed.