

ECONOMIC AND SOCIAL RESEARCH
COUNCIL

DATA POLICY

April 2000

Further information can be obtained from:

Research Division
Polaris House
North Star Avenue
Swindon
Wiltshire
SN2 1UJ
Telephone: 0793 413023
Fax: 0793 413001



ESRC DATASETS POLICY

Council policy on the acquisition, maintenance and support of datasets is set out in the following sections:

- (i) Principles
- (ii) Current Activities
- (iii) Dataset definition.
- (iv) Data acquisition
- (v) Continuing Commitment.
- (vi) Access to data and user support
- (vii) Dissemination and stimulation.
- (viii) Collaboration with other agencies.
- (ix) Charging policy
- (x) Technical Developments
- (xi) Licensing Arrangements
- (xii) Copyright and Confidentiality
- (xiii) Ethics and Professional Code of Conduct
- (xiv) Intellectual Property

(i) Principles

- 1 ESRC has a responsibility for the development of new and improved data resources and the preservation and support of all kinds of social science datasets.
- 2 A dataset is defined, for the purposes of the policy, as any collection of data generated by, or of particular interest to, the social science community which may be considered for archiving. This may include computer-readable data, audio and visual recordings, hand-written documents such as diaries and fieldwork notes, photographs and artefacts.
- 3 ESRC support emphasises the creation of effective data services, rather than the purchase of ready-made datasets.
- 4 All applicants for ESRC research funding are required to demonstrate in their application that they have completed a "data review" to ensure funds are not being requested for data already available to the academic community. If applicants are aware of the existence of a dataset, but are unclear whether it is available through the Data Archive or via Qualidata, then the Archive or Qualidata should be contacted. If the data is unavailable then the Archive or Qualidata may be able to acquire the data on the applicants behalf.
- 5 All award holders are required to offer their computer-readable data for deposit, prepared to a standard which may be used by a third party, within three months of the end of an award. In the case of very large and complicated datasets, the ESRC is prepared to consider alternative deposit arrangements or a waiver, to be negotiated in advance with the Data Archive or Qualidata and on a case by case basis. Time and funds within the award are available for preparation of data for archiving. The final payment of an award will be withheld until data has been deposited in accordance with the requirements. The appropriate Board of Council can agree to alternative deposit arrangements or a waiver of deposit if a convincing case can be made, for example on the grounds of confidentiality.
- 6 All major dataset acquisitions by the ESRC will be subject to a full life-cycle cost justification.
- 7 Any organisation acting as an ESRC dataset depository will produce guidance documents for prospective depositors.
8. Access to data from any ESRC facility will be governed where necessary by user licence agreements.
- 9 Individual Boards of Council will identify datasets of highest priority and will seek to promote their use by researchers.
10. The citation of datasets in published materials is encouraged by the ESRC.

11. The ESRC will collaborate with other national and international agencies in funding and promoting access to datasets.
12. Charges will not normally be levied for academic use of ESRC - funded datasets.
13. The ESRC will seek the best use of different technologies in the methods of distribution of data to meet the varying needs of the academic community.
14. The ESRC expects researchers to meet the copyright requirements set out in the 1988 Copyright, Designs and Patents Act. A guide to the requirements of the Act are included at Annex I of this document.
15. The ESRC's position on exploitation of intellectual property rights is set out at Annex II.
16. The ESRC will consider the ethical issues involved in the use of any dataset at the time of application or acquisition. Applicants are expected to fully anticipate and resolve any such ethical issues.

(ii) Current Activities

The Economic and Social Research Council (ESRC) has a responsibility - unique among social science funders - for the development of new and improved data resources to enable increasingly more sophisticated social science research to be undertaken. Along with other funders, the ESRC also has a responsibility for the long-term preservation and support of social science datasets of all kinds.

The ESRC already has major commitments to dataset creation, acquisition and support, including: 1991 Census of Population; the National Child Development Survey; the British Household Panel Study; the British Election Surveys, the Workplace Industrial Relations Surveys and a network of Resource Centres. For full details see the ESRC Website.

Council's continued investment in its Resource Centres demonstrates a commitment to facilitating secondary analysis of datasets, providing access to and support for both primary data gathered by researchers and others, and data generated as a by-product of administrative or commercial processes.

All ESRC award holders are required to offer copies of their machine-readable data to the Data Archive and in addition, the Archive is required by the ESRC to negotiate on behalf of the social science community access to a wider range of data from a number of governmental and non-governmental sources. These include major government surveys such as the General Household and Labour Force Surveys and public opinion polls.

The ESRC's Qualitative Data Archival Resource Centre (Qualidata) is responsible for collecting, cataloguing and making available qualitative data through a distributed

network of Archives, which includes the Mass Observation Archive and the Modern Records Centre.

The ESRC's commitment to bibliographic data is demonstrated by its funding of an on-line service for the International Bibliography for the Social Sciences. This provides access to bibliographic records for 2,500 national and international journals largely in the disciplines of economics, sociology, politics and anthropology. The Council also supports the Social Science Information Gateway, a meta-data resource, which promotes improved access to data resources available over the internet of interest to social science researchers.

(iii) Dataset definition

A dataset is defined as any collection of data generated by, or of particular interest to, the social science community. Datasets of potential interest to the social science community, and therefore to the ESRC, cover a wide range from the computer-readable datasets of major surveys to small scale data collected as part of individual grants. They include datasets used for research and analysis, the (usually) numeric machine-readable results of surveys, censuses and administrative activity, bibliographic data (but not notes from reading published texts), textual material from the transcription of historical sources or collected through qualitative research (such as diaries and fieldwork notes), and other qualitative data including audio and visual recordings of interviews and events, and photographs and artefacts. The obligation of award holders is to make available for archiving those data generated by their research projects, including appropriate documentation. However, archiving will be selective, both between and within projects.

(iv) Data acquisition

ESRC support will emphasise the creation of effective information services, stimulated and monitored by the Research Resources Board, rather than the purchase of ready-made datasets. The ESRC's Resource Centres will continue to acquire research data from deposits by award holders and by negotiation with governmental, commercial and academic bodies as well as via its links with international data organisations. The Data Archive and Qualidata are also able to examine any dataset referred to them, in order to advise on the feasibility of archiving.

The ESRC currently requires all its award holders to offer their data for deposit within three months of the end of an award. This relates not only to datasets arising from preliminary data collection, but also to derived datasets resulting from ESRC funded work.

The ESRC will withhold the final payment of an award if the dataset has not been offered for deposit to the required standard at the Data Archive or Qualidata within three months of the end of the award, except where a modification or waiver of deposit has been made in advance (see below)

The ESRC recognises that the time and funds necessary for preparation of data can be substantial, particularly in the case of qualitative data which must be anonymised. For

this reason, the Council is prepared to give support within an award for this task. Applicants who wish to be advised on the likely costs involved in preparation of data for deposit are asked to contact either the Data Archive or Qualidata prior to making their application.

Applicants will also be asked to justify why there is a need for new data collection, indicating why currently available datasets are inadequate for the proposed research (ie a "datasets review" to complement the normal "literature review"). If applicants are aware of the existence of a dataset but are unaware of whether it is available through the data Archive or via access arrangements established by Qualidata, then they should, once again, contact the Data Archive or Qualidata. If the data is not available, then either the Data Archive or Qualidata may be able to acquire it on their behalf.

Should any problem relating to the deposit of the data be foreseen, then they should be discussed with either the Director of the Data Archive or Qualidata, or the ESRC at the earliest opportunity. Problems may include confidentiality; data ownership; copyright; subject anonymity etc. It is possible that a request for a modification or waiver of the deposit requirements may be granted by the ESRC, on the advice of the Director of the Data Archive or Qualidata, whichever is relevant, should there prove to be a strong case for doing so. These requests will be dealt with on a case by case basis. If no such request has been made to the Archive, the standard deposit requirements and sanctions as outlined above will be assumed to be in force.

Outside of awards, dataset acquisition will only occur if it is consistent with policies and priorities developed by an ESRC Board. Any acquisition will need to be supported by evidence of:

- (a) Likely usage.
- (b) Reasonable cost effective mechanisms for future maintenance and support.
- (c) Key funding or leverage role through ESRC support.

In line with Council's general policy of obtaining value for money, a full life-cycle cost justification for any acquisition will be required. This will take into account not only initial purchase costs, but the costs of maintenance and user support at an appropriate level to ensure adequate use of the data. The expertise and advice of the Data Archive and/or Qualidata will be sought in all cases.

(v) Continuing Commitment

The ESRC has a long term commitment to fund the continuing maintenance and exploitation of data. The costs involved should be quantified at the time of acquisition. Some datasets increase in value with the elapse of time, others decline in value if not updated and renewed. All require maintenance. This is particularly the case with those in active use, but also applies to archived data, if only to ensure continued accessibility across succeeding generations of hardware and operating systems. In any case, raw information is of almost no value without at least some continuing support for user exploitation.

Any organisation with which the ESRC makes arrangements as a dataset depository will produce guidance documents for prospective depositors. The Data Archive, along with Qualidata, will act as a source of advice on arrangements for continuing data dissemination and related publication practices, together with appropriate conditions of use, for data arising from ESRC funded research.

(vi) Access to data and user support

The archiving of data can be distinguished from user access to that data. User access can also be distinguished from value-added information services. For example, whilst the archiving and supply of data in various formats, along with basic user support, may be appropriate for the Data Archive, specialised support services may be supplied via a designated distribution centre. For example, the Council funds four specialist small units to support and provide access to Census datasets. Similarly, the IBSS Resource Centre provides specialist bibliographic support services.

The funding of all Resource Centres will normally be for a rolling period of five years and will be based upon an agreed set of service definitions, reviewed every three years. Each centre will then be reviewed against specific performance criteria. These criteria will include the levels of use, and satisfaction with, the services provided to the academic community. The likely future value of the service will also be assessed in relation to the cost of provision and alternative sources. In monitoring performance particular attention is given to datasets whose value decreases with time.

Access to datasets can only be achieved if the user community knows what is available, where they are and how to gain access to them. This becomes even more important when the data are distributed in a number of centres. The maintenance of a central catalogue and subject index such as that provided by the Data Archive is crucial to access.

Access to data from any ESRC data facility will be governed where necessary by user licence agreements which will be standardised insofar as they are not affected by the specific requirements of a particular dataset (for example, legal or ethical requirements above the standards normally required, or specific conditions of confidentiality such as commercial sensitivity associated with a particular dataset).

(vii) Dissemination and stimulation

Whenever collection of data likely to be of use as a general resource is an aim of research funding, consideration will be given at the outset to earmarking monies for later support and exploitation. Value for money can only be achieved through use of a resource. The ESRC therefore places a high priority on mechanisms for stimulating use of datasets. Individual Boards will identify datasets of highest priority and will seek to promote their use by researchers, with collaboration between Boards when appropriate. In promoting researchers' exploitation of data services, the Research Resources Board will consider applications for: the support of relevant training; the development of standards for cataloguing; documenting and citing datasets; and the promotion of techniques for linking datasets.

The monitoring, maintenance and development of information support services will be an explicit responsibility of the Research Resources Board.

All Boards will consider whether dissemination of particular datasets beyond the researcher community should be encouraged.

The ESRC encourages citation of datasets in published materials (in order to acknowledge the scientific work involved in the creation of a dataset as well as to allow identification of data sources) as another way of stimulating awareness.

(viii) Collaboration with other agencies

The ESRC will monitor the scope for the purchase of social science information services from agencies, commercial or non-commercial, outside the UK social science community.

The ESRC will collaborate with other national and international agencies - especially other Research Councils and the Joint Information Systems Committee of the Higher Education Funding Councils - in funding and promoting access to datasets. To this end, ESRC will take part in inter-agency fora to discuss priorities in data provision.

(ix) Charging policy

Charges will, for the moment, not normally be levied for academic use of ESRC-funded datasets but this policy will be kept under review. Charges will be appropriate for the following:

- (a) use of the service for non-academic purposes, for which charges will normally be levied on a commercial basis.
- (b) assistance to users, academic or commercial, which extends beyond basic support.

Where the exchange of money is inappropriate (for instance, where the necessary procedure would be uneconomic in operation), ESRC will favour allocation procedures which provide an assessment of changing patterns of demand for use of the data over time, and management and evaluation of performance using indicators and other systems.

(x) Technical Developments

The capabilities of modern networking make it possible within Higher Education Institutions to:

- (a) distribute data from multiple sites;
- (b) separate physical holdings from human support activities; and
- (c) expand international collaboration.

However, although large scale concentration of data holdings and support is now feasible, there are advantages to be obtained in developing experience at more than one site and in promoting distinctive expertise in appropriate locations. Indeed, the factors mentioned here promote specialisation in some of these functions. The ESRC considers flexibility to be essential in the methods of distribution to meet the varying needs of the academic community, and will therefore seek the best use of different technologies.

(xi) Licensing Arrangements

It is important that the ESRC maximises its investment in data acquisition or collection. Data should be used as widely as possible provided this is consistent with the undertakings made to data providers, whether data has been acquired from a third party, or from respondents in social science surveys. It is essential therefore to consider at the time of acquisition or collection how data use might be promoted and to avoid unnecessary restrictions on the use of data.

In some instances data may have significant economic value and attention needs to be paid to achieving a balance between maximising use and ensuring that financial returns are obtained particularly if the data may be used for non academic purposes.

These issues should be discussed with the data provider where data are acquired from a third party in order to clarify conditions of access to the data. When data are being collected directly these decisions will be within the responsibilities of the grant holder. In either case, help and advice may be sought from the Data Archive, Qualidata or another organisation which will be distributing the data.

It may be helpful to draw up a license with the data providers and with the data distributors specifying the conditions of access to the data, payments due and any conditions or responsibilities, including any liability for deficiencies in the data. Access for academic purposes should, in principle, be free. It is also important to consider whether undertakings are required with individuals who will subsequently use the data for secondary analysis. If there are restrictions on the use of the data, undertakings will be necessary. Issues which might be covered by such undertakings are:

- the conditions of access;
- passing data on to third parties;
- use of data by students;
- the requirement to give feedback on errors found in the dataset and to redeposit derived data;
- confidentiality of the material
- restrictions due to copyright
- the requirement to acknowledge and cite the data source and data providers

The Overseas Use of data

Charges will be levied for services supplied to overseas academics, except where reciprocal arrangements are in place or could be agreed. Under such circumstances charges may be waived.

For example, where the ESRC has distribution rights to datasets from the UK Census of Population (eg the Sample of Anonymised Records), these can be made available for academic use overseas on a comparable basis to academic use in the UK. However, any services required by overseas academics should normally be subject to charge.

Where the distribution rights to UK Census datasets resides with the census offices or others (eg Small Area Statistics/Local Base Statistics), ESRC supports access for overseas academic purposes, subject to the agreement of all concerned.

(xii) Copyright and Confidentiality

The ESRC expects researchers to meet the copyright requirements set down in the 1988 Copyright, Designs and Patents Act. Issues of confidentiality tend to derive from case law rather than specific Acts of Parliament. To assist researchers in these complex areas the ESRC has produced a set of Guidelines that are attached at Annex I to this policy document. These Guidelines are also available in stand alone format either in hardcopy or on the ESRC Website. It must be stressed that responsibility for ensuring compliance with all laws and other legal instruments rests with researchers and/or their institutions. The ESRC cannot accept liability for any complaint or legal action taken against a researcher for infringements of copyright, defamation and any other data protection requirement.

The Guidelines cover:

- Copyright
- Ownership
- Infringement and Exceptions
- Employers and Copyright
- Transfer of Copyright
- Confidentiality
- Defamation and Libel
- Data Protection
- Further Advice

(xiii) Ethics and professional Code of Conduct

The ESRC would expect researchers to be guided by the professional code of conduct governing their field of work. The ESRC will consider the ethical issues involved in the collection and use of any dataset at the time of application or acquisition. Council will enforce conditions of confidentiality of collection or use of data on a case by case basis, and this will often be governed by law, as with the requirements of the Data Protection Act, or by the specific terms of supply of data to which the ESRC has agreed with a data supplier, as with the 1991 Census data. The ESRC expects applicants for research funds to fully anticipate and resolve any ethical issues involved in the proposed use or, in the case of primary data collection, acquisition of that data. ESRC's

general guidelines on research ethics are contained within the Research Funding, booklet.

(xiv) Intellectual Property and Commercial Exploitation

It is the Council's policy that potentially valuable results or products arising from ESRC research should, where practicable, be exploited for the benefit of the UK economy and the quality of life of the nation, the award holder and his/her institution and the social science community. To this end, the ESRC requires that the joint Research Councils 'Guidelines for Arrangements on Exploitation' are followed at all times.(Annex II)

The ownership of the intellectual property arising from a research project should be clear from the outset. ESRC will assume that this will rest with the university or institution receiving the ESRC award, unless otherwise stated to the contrary.

The distribution of royalty/income sharing also needs to be clear from the outset. Whoever holds the intellectual property, it should be clear what the distribution of any income might be. For income up to £10,000 ESRC will assume that this will normally rest with the principal investigators and the research team, unless otherwise stated to the contrary. For income between £10,000 and £50,000 ESRC will assume that the major part will accrue to the university and institution, and ESRC reserves the right to reclaim up to one-third of the total, up to the value of its original award. For income over £50,000 ESRC requires that the major part of any income should accrue to the university or institution, and again reserves the right to reclaim up to one third of the total up to the value of the original award.

Whilst computer software, firmware and related documentation are governed by the general conditions outlined above, the ESRC requires award holders to take particular care to protect the value of any computer software, firmware and related documentation resulting from its awards by:

- i) confining public statements about the software or firmware merely to what it does without revealing details of the software or firmware itself;
- ii) ensuring that copies are not supplied to third parties except to individuals who have undertaken to keep them confidential and not to use them for commercial purposes.