



Developing the National Data Strategy

Developments in Data Access

As part of the research infrastructure required to deliver elements of the National Strategy of Data Resources for Research in the Social Sciences it is proposed that two new services should be developed. These are detailed in the attached document.

Members are reminded that these documents are at the draft stage and may well be modified in discussion with key agencies. They are provided for information only and should not be circulated more widely at this stage.



A Secure Data Service: submission of outline proposals

Executive Summary

To support the development of the *National Strategy for Data Resources for Research in the Social Sciences*, the ESRC is seeking to develop a **Secure Data Service**. This service will provide controlled access to sensitive and/or disclosive personal or organisational information which cannot be released for research purposes under End User Licence or Special Licence conditions.

Access to data placed within the Secure Data Service (SDS) will be managed and operated by the service provider, according to rules and conditions established by the individuals or organisations with responsibility for the guardianship of the data. These may vary from source to source, but will provide for the following:

- a secure environment, within which sensitive microdata will be held;
- an information function, via which the research community will be made aware of the resources held within the SDS and the procedures required to gain access;
- an application, authorisation and authentication process through which researchers will make application to access data held within the Secure Data Service;
- a remote access facility, via which authorised and authenticated researchers will gain access to specific data and software held in the secure environment;
- screening procedures to ensure that research outputs requested by researchers are checked to ensure that the conditions specified by the data guardian(s) regarding release of research outputs are satisfied.

Background

The ESRC currently funds (or cofunds with others) the collection of information from individuals and/or organisations and makes such data available to the research community via the Economic and Social Data Service (ESDS). Basic identifying information (names of people/organisations, addresses, detailed spatial identifiers) is always removed from such data prior to their deposit with the ESDS, in accordance with guarantees of confidentiality given to respondents.

Many such datasets, particularly those which are longitudinal in nature, have now developed to the point where individuals or organisations could be identified by virtue of the detail available within the data that are collected for research purposes. In order to maintain guarantees of anonymity given to respondents, one approach to this problem of potential disclosure is to remove or aggregate variables in the dataset before making them

available for research access via bodies such as the Economic and Social Data Service. Another approach is to make datasets available with more restrictive conditions governing their use for research purposes. These cover the security of the environment in which they are held, backed up with severe penalties for breach of these conditions. This access regime, known as the *Special Data Licence*, has recently been introduced by the UK Data Archive, specifically for access to potentially disclosive datasets made available by the Office for National Statistics through the UK Data Archive.

There is a need for a third approach which, while more restrictive than the Special Data Licence, will permit researchers to carry out detailed work to link data and/or to create new analytical variables or to undertake analytical procedures which require access to detailed identifying information. Other examples include the development of linked data where the linked variables are disclosive, and the provision of administrative data from government departments/agencies where the provider stipulates that such data cannot be held outside a secure environment.

The service to be provided, known as the Secure Data Service, will operate in a manner analogous with that developed by the Office for National Statistics Virtual Microdata Laboratory (see www.statistics.gov.uk/about/bdl/). Data placed within the service will be held on a secure server in a secure environment, together with a range of analytical tools and software suitable for data management, statistical analysis and data visualisation. Access to this environment will be managed via access and authentication procedures agreed between the service provider and the body responsible for guardianship of data to be placed within the secure environment. Remote access will be provided to researchers under terms and conditions to be agreed between data guardians and researchers, to be managed by the service provider. The secure environment must have no provision for unauthorised extraction of any data, including research outputs, by the researcher who has been granted access. The service provider will operate methods for the release of research outputs to the researcher as agreed between the service provider and the data guardian.

The specific service requirements

Those bidding to provide the Secure Data Service must provide the following:

1. A secure environment for microdata resources (datafiles, textfile, visual data) which are deemed sensitive by the originators of such information (or those that have responsibility for the guardianships of such data) and which are not being made available for access by usual methods such as portable file copies on digital media or by file transport protocol or web download, either under End User Licence or Special Licence conditions.
2. A web-based information environment, through which the service can make potential users aware of the data available, the scope for detailed analysis that the service can be provided, the access protocol arrangements and the training that is required in order to make use of the service.
3. Access protocol procedures which specify how access arrangements are conducted, the necessary agreements that must be achieved between researchers and the originators (or guardians) of data resources.
4. Training for researchers who request and are granted access to data resources held in the secure environment. Training will focus specifically upon the

recognition of disclosive research results and the use appropriate research methods designed to prevent disclosure of sensitive information held within the secure environment.

5. Accurate and up-to-date management information on access applications, access sessions, the nature of research applications and research outputs.

In addition, the following optional services might be provided:

1. Technical assistance to researchers wishing to engage in various forms of linkage.
2. Assistance to researchers with the selection of appropriate statistical methods, and referral to appropriate training courses provided by organisations such as the National Centre for Research Methods.

Outline bids

At this stage outline bids only are requested. Outline bids will be no more than 2,000 words in length. These should state clearly how the service will be provided; demonstrate the track record that the bidder has in terms of maintaining a secure environment for sensitive microdata and should give indicative costs (full economic costs, distinguishing between direct staff costs and other costs) for providing this service over the first three years. They should give a provisional indication of the level of staffing they anticipate to provide the service in its first year and for the full service as it roles out in years two and three.

Organisations bidding to provide this service may wish to operate as consortia. In such cases, the lead organisation must be clearly identified and the rationale for the arrangement justified, either on scientific grounds or in terms of efficiency.

If short listed, bidders will be required to provide further details about how they might construct and deliver the Secure Data Service. In particular, they will be asked to state how the following guarantees of service can be implemented and upheld.

1. The user identity guarantee
What procedures (administrative, physical and technical) will be employed to ensure that user access to the secure Data Service facility is restricted without exception to those who have been granted access to specific data?
2. The exclusive use guarantee
What procedures (administrative, physical and technical) will be employed to ensure that users who are granted access are restricted without exception to those data sources within the Secure Data Service facility to which they have been granted access?
3. The secure outputs guarantee
What procedures (administrative physical and technical) will be employed to ensure that any research outputs released from the secure environment meet the required standard for release of non-disclosive information?
4. The secure data management guarantee

What procedures (administrative, physical and technical) will be adopted to ensure that the service provider maintains the security and the integrity of the service at all times whilst effectively managing the service (*e.g.* whilst installing new data sources, data updates, software and hardware upgrades)?

5. The secure data recovery guarantee

What procedures will be employed to ensure that conditions 1-4 can be maintained, whilst providing for recovery from disaster (*e.g.* power failure, fire, flood, vandalism).

An important aspect of the Secure Data Service is that it will, from time to time, receive updated versions of datasets. Researchers may wish to work on more than one version of a dataset. Additionally, replication of research may not be possible without access to earlier versions. To avoid confusion, whilst facilitating replication of research findings, a version management control system must be in place.

Timetable for commissioning the Secure Data Service

March 9 th 2007	Research Resources Board approval for commissioning process
May 1 st 2007	Invitation to submit outline bids is published
July 1 st 2007	Deadline for submission of outline proposals
September 1 st 2007	Commissioning Panel invites short-listed bidders to submit full bids
October 31 st 2007	Deadline for receipt of full bids
September 30 th 2007	Commissioning Panel announces winning bid
January 1 st 2008	Secure Data Service commences operation



An Administrative Data Service: submission of outline proposals

Executive Summary

To support the development of the *National Strategy for Data Resources for Research in the Social Sciences*¹ the UK Data Forum and the Economic and Social Research Council (ESRC) commissioned a selective audit of the availability and use of administrative data sources. This audit focussed upon a number of areas of potential interest for social scientific and related research interests and examining within these areas both the types of administrative data available and their potential as research resources.

Building upon the recommendations in this audit, outline proposals are being sought for the development of an **Administrative Data Service** for the research community – a service which will work in close collaboration with a number of government departments and agencies, together with other ESRC-funded resource centres. This service will have a number of functions. First and foremost, it will become a focal centre for knowledge about the availability of appropriate administrative data, their suitability for specific research purposes and the procedures required to gain access to and use such data. Second, it will work in tandem with government departments and agencies, seeking to develop and improve the use of administrative data resources for research purposes. The service will not hold or store such data, but will act as an intermediary between researchers and the organisation(s) providing access to administrative data for research purposes.

Background

Administrative data describe information which arises via the operation of a transaction, registration or as a record of service delivery. They relate specifically to the administration of a system or process and are not primarily generated as research resources.

While such data are not necessarily the preserve of government, all government departments and agencies keep records of the variety of services they deliver and the processes they register, often storing this information as electronic records that relate to individuals and/or organisations. These records have the potential to inform social scientific research - either directly through analysis of such data at the micro level or, via data linkage techniques, to enhance existing research resources.

Administrative records cover a wide variety of fields in both the public sector and private sectors, including demographics, consumer behaviour, education, social care and community support, crime, transport, health, taxation, social security, housing and migration. Via personal or organisational identifiers, data from different sources have

¹ See

the potential to generate rich resources for research purposes. This potential has been recognised in a series of recent reports which have also raised a number of important issues relating to the need to preserve their confidentiality, concerns about the privacy of individuals and the legality of sharing such information between government departments and agencies². However, access for research purposes and the scope for data linkage may be restricted by ethical considerations surrounding the sensitive nature of the linked data, by legal requirements and by the resource costs of undertaking linkage.

A selected audit of administrative data sources

An audit was recently conducted³ to establish the scope, quality and potential that public sector administrative data can offer as research resources. Given the wide variety of administrative data that exist, this audit focussed specifically upon resources held by the Office for National Statistics, the National Health Service, the Department of Work and Pensions, and the Department for Education and Skills.

In summarising the findings of this audit a number of key observations were made:

- the scale of administrative data resources available within government departments is extensive. While this selective audit was focussed within a few departments, a large number of different resources was revealed. A considerable amount of research has already been conducted using these resources, yet they remain relatively undiscovered within the wider social science research community.
- By their very nature, existing administrative data resources are growing continually and their complexity is increasing via linkage between them. Furthermore, new developments are underway (*e.g.* ‘Connecting for Health’) which will have a major impact upon the extent of future resources.
- Access to these resources for research purposes varies from department to department. Depending upon the perceived sensitivity of the data and/or the legal framework governing data sharing arrangements some departmental ‘gatekeepers’ regulate access conditions tightly.
- In some instances the ESRC has established good links with organisations which provide access to controlled microdata or has funded preliminary work to explore these resources further (*e.g.* the Pupil Level Annual Schools Census User Group). In other areas, no such links or preliminary work have been established.
- At the time the audit was conducted, there was uncertainty about the nature of the powers that an independent ONS will have to centralise access to

² See, for example:
Better use of personal information: opportunities and risks (Council for Science and Technology, November 2005)

(<http://www.cst.gov.uk/cst/reports/files/personal-information/report.pdf>)

Privacy and data-sharing: The way forward for public services (Performance and Innovation Unit, Cabinet Office, April 2002)

(<http://www.strategy.gov.uk/downloads/su/privacy/downloads/piu-data.pdf>)

Data Sharing for Statistical Purposes: A Practitioners’ Guide to the Legal Framework (Office for National Statistics, Sep. 2005)

(http://www.statistics.gov.uk/downloads/theme_other/NSDataSharing.pdf)

³ Jones, P. and P. Elias (2006)

administrative data resources. While legislation may give the Office for National Statistics the authority to undertake data linking across departments, this may only happen on a case by case basis and through additional legislative ‘gateways’.

- Knowledge about the suitability of different administrative data resources for various research purposes is limited and fragmented. There is a need to harness and coordinate this knowledge in a systematic manner. This includes:
 - knowledge about their existence, scope and potential to inform research issues or enhance other resources;
 - knowledge of procedures, gatekeeping arrangements required to gain access;
 - knowledge about the technical skills required to handle large scale datasets and/or to establish appropriate linkage mechanisms between different sources;
 - knowledge about the ‘quality’ of the administrative data (*i.e.* whether or not a specific source of administrative data will provide the additional information required to address a specific research purpose).

An Administrative Data Service

The ESRC, in collaboration with a number of key government departments (*e.g.* ONS, DWP, DfES, NHS, Home Office) wishes to address these problems directly. It is proposed that this should be achieved by establishing an **Administrative Data Service**.

The proposed service will aim to provide knowledge about, technical assistance for and organisational help with access to some of the major administrative data resources available in the UK. To achieve these aims, the service will:

1. help researchers to gain access to and link with administrative data resources.
2. collect, develop and help disseminate information about the variety of administrative data resources potentially available for research in the social sciences and related disciplines.
3. work closely with the departments/agencies responsible for the guardianship of administrative data resources, to explore the potential that such data have to inform research in the social sciences and related disciplines
4. explore the scope for linkage between various administrative data sources and to other personal records (*e.g.* survey data, census information), thereby enhancing and extending existing resources.

The role of the service will include:

- agreeing procedures for research access where such procedures do not currently exist;
- harmonising these procedures where possible;
- promoting the use of administrative resources for research purposes via conference, workshops, websites etc;

- assisting researchers who wish to make use of administrative data resources for research purposes;
- undertaking quality reviews of administrative data resources to establish their strengths and weaknesses as research resources;
- assisting in the preservation of linked data sets via archiving and data curation, subject to conditions laid down by data owners;
- liaising with the Economic and Social Data Service to determine whether or not a wider range of historical and anonymised administrative datasets can be lodged therein.
- liaising with the National Archive to ensure that, with the agreement of those bodies responsible for the guardianship, relevant data and documentation are preserved in accordance with established standards;
- where permissible, studying the feasibility of data linkage between different data sources and testing possible linkage procedures.

It is important to note that the proposed **Administrative Data Service** would not seek to become a repository for administrative data, which would remain within the control and protection of the agency responsible for their safe-keeping. It will work across a variety of government departments and agencies to promote access to and appropriate use of administrative data sources. It will not seek to replicate work being undertaken elsewhere (*e.g.* the Secondary Uses Service being developed by the National Health Service), working instead to assist the academic community to make more and/or better use of these rich resources.

The timetable for development of the service

It is intended that the new service should start on January 1st 2008. It will be necessary for the ESRC to conclude agreements with various government departments and the Office for National Statistics before the Administrative Data Service becomes fully established, and this may not take place until after the Statistics and Registration Bill is enacted later this year. With this in mind, it is likely that the service will be rolled out gradually over the first two years of its development.

Submission of outline proposals

Suitably qualified individuals/organisations are invited to submit an outline proposal to provide this service, illustrating the way they see this service developing and their ideas about the manner in which it should be delivered. Outline proposals should consist of no more than 2,000 words and should specify:

- the experience and qualifications that bidders will bring to the services they propose to provide;
- the nature of the service to be provided;
- the manner in which the service will be provided;
- how they envisage the service will develop within the first two years of its operation, and
- the timetable for the roll out of the service within the first year of funding.

They should give a provisional indication of the level of staffing they anticipate to provide the service in its first year and for the full service as it rolls out in year two. Provisional costs of providing the service must be given (full economic costs), distinguishing between direct staff costs and other costs.

The proposed service could be provided by a single body, or by a consortium of interested parties. However, if bid to provide the service comes from a consortium, there must be strong justification for working in this manner, and sound managerial arrangements in place to ensure that the service provided is well integrated.

Timetable for commissioning the Administrative Data Service

March 9 th 2007	Research Resources Board approval for commissioning process
May 1 st 2007	Invitation to submit outline bids is published
July 1 st 2007	Deadline for submission of outline proposals
September 1 st 2007	Commissioning Panel invites short-listed bidders to submit full bids
October 31 st 2007	Deadline for receipt of full bids
September 30 th 2007	Commissioning Panel announces winning bid
January 1 st 2008	Administrative Data Service commences operation