

MEMORANDUM FROM THE ECONOMIC AND SOCIAL RESEARCH COUNCIL IN RESPONSE TO THE MINISTRY OF JUSTICE CONSULTATION ON THE USE AND SHARING OF PERSONAL INFORMATION IN THE PUBLIC AND PRIVATE SECTORS

1 Introduction

- 1.1 The Economic and Social Research Council (ESRC) is the UK's largest funding agency for research and postgraduate training relating to social and economic issues. It provides independent, high quality, relevant research to business, the public sector and Government. It is the largest Research Council¹ in terms of the community it serves, accounting for over 25% of those staff returned in the last Research Assessment Exercise, across a wide range of disciplines.
- 1.2 For the UK science base to contribute fully to economic development, quality of life and public services, social science research is crucial. This is because human progress depends not only on natural science, engineering and technology but on the social and economic conditions that shape them and their subsequent implementation. Work supported by the ESRC contributes in many ways to the social, political, economic and intellectual life of the UK and beyond it. It directly informs public policy and professional practice across all sectors of the economy. It is therefore essential in meeting the Government's economic objective of *'building a strong economy and a fair society, with opportunity and security for all'*.
- 1.3 Increasingly, the information which social scientists need to draw upon to provide evidence for policy and practice is from large scale national resources. These include national census and survey data, or data which are generated via administrative records, such as hospital visits, social benefits received or schools attended. Census, survey and administrative data, when suitably organised, stored, documented and made accessible for research purposes, form part of the *'national infrastructure for social research'*. Appropriate sharing of data is therefore critical for high quality, relevant social science research.
- 1.4 The ESRC response to this consultation is given from the researcher perspective and as such does not strictly follow the questions set out in the consultation document. ESRC's organisational perspective in terms of the types of personal information the Council holds and shares is set out in the response provided on behalf of Research Councils UK (RCUK) and is not repeated in this document. However please note that some parts of this response are replicated in the RCUK paper.

2 Interest in Data Sharing

- 2.1 The ESRC has a strong interest in data sharing to promote new and exciting data resources to the social science research community. Particularly in the case of data sharing across government departments, such resources have the potential to provide a strong evidence base for policy-relevant research, and to reduce surveying costs by utilising existing administrative data. Such potential has been

¹ There are seven research councils who collectively form the body Research Councils UK (RCUK)

recognised by departments such as the Cabinet Office² and Office for National Statistics³, who also raised important issues relating to the need to preserve confidentiality and to access such data within a clear legislative framework.

- 2.2 In terms of promoting responsible data sharing, the Economic and Social Data Service (ESDS), an ESRC-funded resource service, provides an integrated service offering enhanced support for the secondary use of data across the research, learning and teaching communities. ESRC grant-holders are obliged to offer for deposit with the ESDS a copy of any data resulting from ESRC-funded research. If accepted by ESDS, this data is usually made available via an End User License (EUL), are fully anonymised and certain variables may be suppressed to minimise risk of disclosure.
- 2.3 The EUL makes several references to the users obligations towards maintaining the confidentiality of the data they have downloaded. In particular, users agree to *'preserve at all times the confidentiality of information pertaining to individuals and/or households in the data collections where the information is not in the public domain.'* Any breach of the terms and conditions set out in the EUL would result in immediate termination of the users access to all services provided by ESDS and the UK Data Archive⁴.
- 2.4 Access to the large majority of micro data only requires registration with ESDS, agreement to an EUL and the registration of usage details. These registration details are then authenticated (currently using the Athens system) to verify the identity of registered users. However, it is recognised that some researchers need access to more detailed data. To facilitate this and to increase the range of data available for statistical research, the Office for National Statistics (ONS), in collaboration with the UK Data Archive (UKDA) and ESDS Government, developed a strategy to provide access to social survey datasets that are detailed, yet anonymised. Since these data pose a higher risk of disclosure, they have additional special conditions attached to them that take the form of a Special Licence (SL). The SL requires the signature(s) of the researcher(s) and the institution with responsibility for the researcher. It also needs the explicit permission of the data owner to release the data to the researcher(s).
- 2.5 The SL contains:
- the conditions for access for statistical research purposes;
 - the obligation of the researcher(s) and the measures for protecting and respecting the confidentiality of statistical data;
 - statistical purpose and how the data are to be used;
 - justification why access is needed to the more detailed version of the data;
 - standards and methods for disclosure control for any outputs;
 - requirement on the researcher to supply the bibliographic details of any published work, based wholly or in part on the data collection(s) accessed;
 - clauses relating to data and site security and destruction of the data on completion of the project; and
 - sanctions to be applied to breaches of confidentiality.

² Cabinet Office (2002) *Privacy and data sharing: The way forward for public services.*

³ ONS (2005) *Data sharing for statistical purposes: A practitioners guide to the legal framework.*

⁴ ESDS is hosted and managed by the UK Data Archive.

- 2.6 Data are only provided to those researchers who can justify their need for the data for statistical research purposes and agree to meet the terms of the SL⁵.

3 Data Linking

- 3.1 ESRC is not only interested in the potential for data sharing; the Council also recognises the potential that data linking poses for research across disciplinary boundaries, where such linkage would be made possible by data sharing in the first place. Data linking involves taking information about the same data subject from two or more datasets and merging these data to form a new dataset containing matched records. The practice is sometimes complex but the rewards are great.

- 3.2 The main obvious benefit of data linking *'lies in being able to bring together existing resources to create new datasets which in turn would allow researchers to address complex research questions relating to health, socio-economic status, education and so on'* (Elias and Jones, 2006). In addition to this, linking administrative sources to survey datasets such as the UK Decennial Census has the potential to fill any gaps that may result from non-sampling error.

⁵ Paragraphs 2.4, 2.5 and 2.6 have been replication from <http://www.esds.ac.uk/orderingdata/specialLicence.asp> with kind permission of the ESDS.

4 Key Benefits of Data Sharing

- 4.1 From a research perspective, ESRC appreciate in the broadest sense the many potential benefits of data sharing to individuals and society as a whole. Data sharing:
- Facilitates high-quality, policy-relevant research by sharing and then combining datasets from different departments and agencies to form a full picture rather than analysing separate pieces of a jigsaw.
 - Reinforces open scientific inquiry thereby improving methods of data collection and measurements through the scrutiny of others.
 - Promotes new research and allows for the testing of new or alternative methods.
 - Reduces costs by avoiding duplicate data collection efforts.
 - Allows the creation of new datasets through the merging or linkage of two or more existing sources of information.
 - Provides an important resource for training in research by enabling new researchers to utilise existing data.
 - Can reduce the burden on respondents caused by multiple data collection efforts.
 - Reduces the information security risks associated with maintaining duplicated datasets in more than one location.
- 4.2 A potential area of social science research made possible through data sharing relates to evidence-based policy research in education utilising the National Pupil Database (NPD). The NDP is a census of all pupils in England in Local Education Authority maintained schools. It includes attainment and other pupil-level information in the Pupil Level Annual Census (PLASC). The sheer size and completeness of this dataset means that is potentially a rich research resource that is unmatched in other areas of educational research.
- 4.3 There are many examples of the potential for data sharing for research purposes. For an overview of thee in the areas of education, the labour-market, health, business and demographics, please see a report commissioned by the ESRC entitled *Administrative data as a research resource: a selective audit*⁶(a copy is attached to this response). This audit focused upon a number of areas of potential interest for social scientific and related research interests and examined within these areas both the types of administrative data available and their potential as research resources.
- 4.4 Building upon the recommendations in this audit, the ESRC will shortly publish an Invitation to Tender (ITT) to develop an Administrative Data Liaison Service (ADLS) for the research community – a service which will work in close collaboration with a number of government departments and agencies, together

⁶ Elias, P. and Jones, P. (2006) *Administrative data as a research resource: a selected audit*. A report commissioned by the ESRC Research Resources Board.

with other ESRC-funded resource centres. The full ITT is attached to this response for information.

- 4.5 This Service will have a number of functions with a key aim of assisting researchers who wish to make use of administrative data resources for research purposes to do so ethically and within a clear legal framework. 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. It will work in tandem with government departments and agencies, seeking to develop and improve the use of administrative data resources for research purposes. Please note that 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.

5 Consent

- 5.1 The ESRC recognise that ‘consent’ is a complex concept. It is generally used to indicate that a party has agreed to enter into an arrangement or understanding with another. It can be implied or explicit. In terms of the research that ESRC funds, the Council requires that all is conducted ethically and with appropriate consent. Where consent in this instance must be informed, freely given and explicit.
- 5.2 However, despite advice to the contrary, researchers do not always obtain appropriate consent. For example, an ESRC-funded researcher conducting a study on a sensitive topic such as social influences on sexual behaviour may seek to reassure their study participants by stating in the consent form that the resulting data will only be used by the research team. This then causes a problem for the researcher as they are obliged by ESRC to offer their dataset for deposit at the Economic and Social Data Service. This is a particular problem with qualitative, rather than quantitative data collections.
- 5.3 What this example serves to illustrate is that explicit and informed consent should be obtained that not only means the participant understands their involvement and what their data will primarily be used for, but also highlights the fact that their anonymised data may be shared with others where appropriate.
- 5.4 An issue that requires clarity is whether consent should be sought for the sharing of administrative data. By its very definition administrative data has not been obtained through research and therefore is not designed for research purposes. As such, consent to share such data is unlikely to have been sought which, if enforced, may hinder data sharing initiatives across government. Given the enormous potential that administrative data and subsequent data linking has for policy research and thus the advantages that its appropriate use poses for society, it is critical that this ambiguity is clarified.

6 Technology

- 6.1 Technological advances pose both risk and reward for data sharing. We are living in a more mobile society with internet enabled mobile phones and PDA’s being common place. Faster and more secure internet and email have meant that more data is being sent electronically than ever before, and the ‘paper-less’ office has resulted in more data storage on PC’s and data distribution via CD’s or other

electronic formats. All of these, whilst being more convenient and faster to use than manual, paper methods, do pose some risk to data security from hackers, internet viruses and data loss. One way to counter this risk would be to better educate the public about using technologies safely in terms of protecting their personal data from fueling criminal activity such as identity theft

- 6.2 Having said that, in many ways technological advances have enhanced data security by helping develop more sophisticated ways to control for disclosure of personal information, creating firewalls and installing advanced password protection systems. It has also helped to facilitate data sharing. For example, researchers can access electronic copies of large survey datasets from data archives and conduct powerful secondary analysis on them; local doctor's surgeries can quickly share patient information with a hospital if that patient is taken seriously ill. There are many more examples of the societal benefits that technology has enabled.
- 6.3 In terms of facilitating research, advanced technology has enabled researchers to access data that previously would have been unavailable. For example, the ONS shares sensitive data with researchers via their Virtual Microdata Laboratory (VML) ⁷. The VML is a research access facility through which confidential, sensitive data can be accessed within a secure environment. Due to the nature of the data, access is tightly restricted and procedures and audits are in place to ensure that only non-disclosive analysis is carried out.
- 6.4 The ESRC are looking to develop a Secure Data Service (SDS) that will be modeled on the ONS' VML. The SDS will provide controlled access to sensitive and/or disclosive personal or organisational information which cannot be released for research purposes under End User License or Special License conditions. The aims of the SDS are to provide:
- 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;
 - training for potential users of the service, to ensure that they are aware of the varying requirements to preserve the confidentiality of personal or organisational records in any research outputs derived from such records.

⁷ The VML can be accessed from all four ONS sites at Titchfield, Newport, London and Southport.

6.5 The full ITT for this Service is attached as information to this response. The ESRC's long-term vision for this Service is to make secure environments available in every Higher Education Institution to enable and facilitate the analysis of sensitive data.

6.6 However, technology moves very fast and new developments provide new opportunities to link increasingly diverse data sources. To take advantage of these vast opportunities it is essential that any amendments to the DPA and related legislation recognise the fluid nature of technology and its subsequent impact on research.

7 Summary

7.1 The opportunities that improved data sharing and access has for research, particularly policy-relevant research, is enormous considering the vast number of different administrative datasets that are in existence from public bodies. The ability to share and subsequently link these datasets has the potential to form a rich evidence base from which research can be undertaken that would benefit the whole of society in areas such as health, crime and social care. However, the legislative framework for this practice is unclear and requires clarification to protect both individuals who provide personal information, and researchers who utilise such sensitive information. ESRC would therefore like to take this opportunity to stress that when considering the benefits and risks of sharing personal data the perspective of researchers (both academic and non-academic) is taken into account.