Data Sharing Review: Consultation on the use and sharing of personal information in the public and private sectors

Response from the Research Information Network

1. The Research Information Network is a strategic body established by the UK Higher Education Funding Bodies, the Research Councils and the three National Libraries. Its core role is to conduct research to improve our understanding of current and future developments in researchers' use of information resources and services; and to promote innovation by contributing to the development of effective policies and services that enhance the performance of the UK research base.

2. Access to and sharing of personal information is important in many areas of research. It plays a vital part in ensuring that the UK sustains a dynamic and internationally-competitive research sector that makes a major contribution to the expansion and dissemination of knowledge, to national well-being, and to the status and reputation of the UK.

3. The issues raised in the current review are thus of critical importance to many areas of research. We are therefore surprised that the consultation has not been more widely publicised, and that it is being restricted to such a short timetable, one that does not conform to the Government's Code of Practice that prescribes a minimum period of 12 weeks for responses. Nevertheless, we welcome the opportunity to contribute to the review.

Section 1: Background

4. The UK research base plays a vital part in sustaining the continued growth of the economy and improvements in our public services. In several areas of research – especially, but not only, in the social and medical sciences – access to and sharing of personal information is an essential part of the research process. It is therefore of critical importance that appropriate arrangements should be in place to ensure that accredited researchers can gain access to and share with others information relevant to the research questions they are seeking to answer; and that those arrangements should secure the confidence of the public at large.

5. The RIN has recently issued guidelines[1] on the stewardship of research data that address some of the issues raised in the consultation. We emphasise the need for a clear allocation of roles and responsibilities; the need for codes of practice; the need for clear processes for determining who may be given access to what data, and on what terms; and the need to conform with legal and ethical requirements.

6. It should also be noted that organisations that fund research gather personal information about researchers who apply for grants, and about other researchers who will be employed on research projects. Much of this information is shared with experts who review and assess applications, and this is a vital part of the system to ensure that research is of the highest quality. But there remains a need to consider carefully the circumstances and arrangements under which funding organisations will share personal information with other researchers and third parties.

---

[1] *Stewardship of Digital Research Data: A Framework of Principles and Guidelines* , available at http://www.rin.ac.uk/files/Research%20Data%20Principles%20and%20Guidelines%20full%20version%20-%20final.pdf.

Section 2: Scope, Benefits, Barriers and Risks

7. It is important to note that the personal information to which researchers need access is in different forms and from a number of different sources. In the social sciences in particular, much of the necessary information is gathered initially by the Government and other public sector bodies. Both survey data and administrative data are important sources for research, and there are concerns about the lack of consistency between Government Departments in their approaches to the disclosure of data of both kinds. We believe that there may be scope for extending more broadly and to other Departments the arrangements now established with the Office of National Statistics to secure access to specified sets of data for approved researchers.

8. Researchers themselves also generate sets of data in the course of their research. Most institutions in the higher education sector have Ethics Committees which oversee arrangements to ensure that research is conducted in accordance with appropriate ethical standards, including the arrangements for gathering, analysing and managing personal data where that is required. All research involving patients is must be passed by such committees.

9. Informed consent is of critical importance, and some of the major research funders provide advice to researchers and to institutions on the form and drafting of consent arrangements. As we note in paragraphs 12 and 17 below, however, some aspects of the current arrangements do not work well, and there is a need for greater awareness among researchers of the need to consider at the beginning of their research projects the potential for re-use of the data they gather, and how to ensure that barriers are not put in the way of realising that potential.

10. The benefits that arise from the sharing of personal information in clinical research are obvious. More broadly, in many areas of biomedical and social scientific research, access to and sharing of personal information is vital in leading to improvements in public policy, and in healthcare and other services. Data sharing is therefore of critical importance in

    - Enabling researchers to create new aggregations of data, merging or linking different datasets to facilitate fuller analysis, rather than analysing separate parts of the picture;
    - Promoting new research by allowing for the testing of new hypotheses and methodologies; and
    - Reducing the burden on respondents and avoiding duplicate data collection efforts.

11. The key barriers to researchers' gaining access to personal information gathered by Government and other public sector bodies are nervousness on the part of those bodies about the potential risks; misunderstandings as to why the information is needed and how it will be managed and used; and lack of consistency both within and between different bodies. There is an urgent need for clearer guidance to public bodies on how and in what circumstances they can provide accredited researchers with access to personal information.

12. With regard to the data that researchers themselves gather in the course of their research, the major barrier to sharing and re-use by other researchers is the lack of training and awareness of good practice in the stewardship of research data, including such critical matters as the requirement for high-quality documentation and metadata. While the guidance provided by bodies such as the Medical Research Council and the Wellcome Trust is very detailed, it is often the case that, since the potential for data sharing is often not given priority attention when research projects are initiated, issues

relating to consent for sharing and of ownership are frequently considered too late, at the end of the project. Difficulties also arise as a result of uncertainties about the ownership of research data, and whether researchers or their institutions actually have the right to share the data with others. We suggest that funders and Ethics Committees need to provide guidelines to ensure that these issues are considered before data collection begins, or to take steps to encourage compliance with the existing guidelines provided by other bodies.

13. The greatest risks in data sharing are that data subjects' privacy will be lost, through loss or misuse of data. Such risks are particularly high when different datasets are aggregated in a single location. And the biggest risk for the future of UK research is that high-profile examples of loss of privacy will lead to irrevocable damage to public trust in the arrangements for data sharing.

14. In order to mitigate these risks, we support the arrangements that have been put in place to provide access to some datasets containing personal information only to approved researchers who specify the use(s) to which the data will be put; and commit themselves and their institutions to approved standards and methods for disclosure of any outputs, and to security measures to protect the data. We believe that these arrangements could be extended more widely.

Section 3: The legal framework

15. Research funders and institutions take very seriously their obligations under the Data Protection Act; and in the light of recent incidents, many of them are reviewing their policies with regard to the transportation of data outside their premises.

16. There are three issues that we believe require further consideration. First, the relationship between the DPA and Freedom of Information legislation is of critical importance. While the exemption of personal information under Section 40 of the Freedom of Information Act 2000 is reasonably clear, it is important that any revision to the DPA should consider the relationship between the two pieces of legislation. Second, technological change is proceeding very fast, and there is a need to ensure that the legislation is as near as possible future-proof in the light of such changes. Third, we believe there is a need for greater clarity as to how researchers among others may share personal information that comes in the form of audio or visual data.

Section 4: Consent and Transparency

17. As we have noted above in Section 2, the requirement for informed consent in gathering data in the course of biomedical and social scientific research is made absolutely clear to all researchers. Guarantees as to the adoption of trusted arrangements for the anonymisation or pseudonymisation of data are of critical importance in securing consent. Difficulties arise, however, when data that is gathered for a particular purpose in one research project may be of value for different purposes to researchers working subsequently on another project. The precise terms of consent gained from data subjects may frequently preclude such use, or even the preservation of the data. Hence it is important that consideration should be given to forms of consent that allow for the sharing of anonymised data for further research where that is appropriate. These issues are not well-covered in current guidelines on consent, and that gap needs to be remedied.

18. If researchers are to gain access to administrative and survey data gathered by Government Departments and other public bodies, it is important that informed consent is gained in gathering the data, in similar vein to the procedures adopted in the research community. There are particular problems here in relation to administrative data, where consent for sharing for the purposes of research is unlikely to have been

sought. Given the potential value of administrative data for research – particularly in policy-related areas – it is important that clear guidelines should be developed as to the circumstances under which such data can be shared.

Section 5: Technology

19. It is now possible to share data for the purposes of research over the web, and researchers are making increasing use of this facility. This gives rise to significant security concerns, however, and while technological advances mean that advanced security measures can be used, there is a need to ensure that appropriate measures are in fact adopted in all cases. Since technological change advances so fast, however, it may be difficult to legislate in this area. More generally, we believe there is a need for greater availability and use of secure environments in which authorised researchers can be given access to personal information but where other users are firmly excluded.

20. Data mining and text mining technologies promise to be important research tools for the future, enabling researchers to extract implicit, previously unknown, and potentially useful information from very large aggregations of data. Such developments are already being used by business and financial analysts; and they will become an increasingly important part of the research landscape in the next few years. In relation to personal data, they raise significant issues relating to privacy, since the result of data mining is to generate information that would not be available otherwise. It is impossible therefore for individuals to ascertain the nature and extent of the information that is being generated about them; and there are legitimate concerns about the quality of data sources that are being analysed, as well as the interpretations that are being put on the inferential information that is being generated. These issues need to be subject to urgent review by experts in the field, and codes of practice developed to ensure that large aggregations of data are not mined inappropriately, to fish for information which the data are not designed to provide.

Section 6: International Comparisons

21. The US approach to the protection of privacy is different from that applied in the EU in general, and the UK in particular. The US Safe Harbor framework provides a mechanism to enable research organisations in the UK to share personal data with related organisations in the US. In practice, however, sharing is limited because only a limited number of organisations in the US have joined the programme. This has a damaging effect on collaboration between researchers in the UK and the US, and there is a need to find ways to address this problem.