Stewardship of digital research data: a framework of principles and guidelines

Responsibilities of research institutions and funders, data managers, learned societies and publishers

January 2008

research
information
network

www.rin.ac.uk

The principles

In order to produce high-quality research, researchers must have access to as wide a range as possible of the data and information produced by other researchers, as well as relevant information produced by other agencies in the UK and overseas. Similarly, successful dissemination and exploitation of research depends on effective flows of information between researchers and other individuals and organisations that have an interest in its results. A successful research and innovation system thus depends on the open exchange of ideas, information and knowledge.

Developments in information and communications technologies are transforming the nature and scale of research, enhancing both quality and productivity. They are facilitating new kinds of research, new organisational models, and collaboration across disciplinary, institutional and national boundaries. But they also demand new ways of thinking about how we manage data and information outputs, so that we can maximise their value, and ensure that precious resources are not lost. In pursuance of those goals, the fundamental policy objective is to ensure that

Ideas and knowledge derived from publicly-funded research should be made available and accessible for public use, interrogation, and scrutiny, as widely, rapidly and effectively as practicable.

To help achieve that objective, the following five principles provide a broad framework for developing good practice for universities, research institutions, libraries and other information providers, publishers, research funders as well as researchers themselves. The principles are pitched at a high level. The guidance set out in the main document provides pointers as to how policy and practice may need to be changed to ensure that they are to comply with the principles

I   **The roles and responsibilities of researchers, research institutions and funders should be defined as clearly as possible, and they should collaboratively establish a framework of codes of practice to ensure that creators and users of research data are aware of and fulfil their responsibilities in accordance with these principles.**

II   **Digital research data should be created and collected in accordance with applicable international standards, and the processes for selecting those to be made available to others should include proper quality assurance.**

III   **Digital research data should be easy to find, and access should be provided in an environment which maximises ease of use; provides credit for and protects the rights of those who have gathered or created data; and protects the rights of those who have legitimate interests in how data are made accessible and used.**

IV   **The models and mechanisms for managing and providing access to digital research data must be both efficient and cost-effective in the use of public and other funds.**

V   **Digital research data of long term value arising from current and future research should be preserved and remain accessible for current and future generations.**

Background and context

Introduction

1. This statement has been drafted by the RIN following discussions with key stakeholders and a public consultation conducted during the middle of 2007; a list of the organisations that responded to this can be found at **Annex A**. The document addresses some of the key issues that arise in managing the unprecedented quantities and varieties of digital data now being created and collected by researchers. It sets out a policy framework of five principles, with associated guidelines, to help ensure that such data are properly looked after.

Why do we need a policy framework?

2. Over recent years, much attention has focused on providing access to research **publications**. That is the focus of the RCUK Position Statement on Access to Research Outputs[1] published in June 2006. Less attention has been paid to issues surrounding the stewardship of the digital **data** created and used by researchers. But data are growing in importance as a product of research and an essential part of the evidence necessary to evaluate research results, and to reconstruct the events and processes leading to those results. We shall realise the value of data only if research policies, practices and support systems develop appropriately and we adopt common standards and good practice – in the UK and, where appropriate, internationally – in managing and providing access to them. Many of the issues to be addressed relate to the handling of all kinds of research data, in analogue as well as digital form: the essential principles are those of good records management in research. But stewardship has become a particularly urgent matter as increasing volumes and varieties of digital data are being produced, in forms which demand that stewardship and curation issues are addressed early in the research process.

3. No single agency can tackle all the issues. Rather, we need a cohesive framework of policies and procedures for key agents and stakeholders in order to maximise the potential benefits of digital research data. Some institutions, funders, or research groups have standards and protocols already in place; but there is a strong case for a co-ordinated approach, not least to promote coherence and avoid inconsistency or duplication – whilst at the same time recognising the distinctiveness of different communities. In the broader context of the development of codes of research governance it is thus important to establish transparent and consistent systems of information governance.

What kind of policy framework do we need?

4. Given the wide range of contexts in which research is conducted, the need is for a broad and shared framework to provide a basis on which key agents – universities, research institutions, libraries, publishers and other information providers, research funders and of course researchers – can develop approaches adapted to their needs. The first stage in developing such a framework is to establish a set of guiding principles that provide a foundation for a series of questions that need to be addressed in developing policy and practice at national and

---

[1] Available at http://www.rcuk.ac.uk/access/2006statement.pdf. Related statements from each individual Research Council are available on their respective websites. A consultation draft of the statement issued in June 2005 (http://www.rcuk.ac.uk/access/2005statement.pdf) made clear that it covered only journal articles and conference proceedings, and was not intended to cover other forms of outputs.

institutional levels. The guidance set out in this document is based around such principles, and some key questions which follow from them. But we do not provide prescriptive answers; different bodies will need to address the questions, and to develop their own policy and good practice, in the light of their own particular circumstances. As the framework is developed, however, there is a continued need for collaboration and co-ordination, to ensure that it meets a number of requirements.

Roles and responsibilities

5. The first requirement is thus for collaboration, to help to make explicit the roles and responsibilities of the key players in the research and research communications processes: researchers themselves, funders, the institutions in which researchers work, those who access and use data, and organisations such as libraries and archives that take responsibility for access and long-term preservation.

Sensitivity to the needs of researchers and the contexts in which research is conducted

6. A second requirement is for sensitivity to the different institutional, disciplinary and funding contexts in which researchers gather and use data. In some areas of research, where, for example, data are derived from human subjects, there is already a strong regulatory framework, and we must not add in a burdensome way to existing obligations. Elsewhere, commercial confidentiality is also a factor. One set of arrangements will thus not be appropriate for all circumstances. It is therefore essential that researchers themselves, in their various disciplinary and institutional settings, should be fully engaged in adapting and refining the framework, so that it takes full account of their needs and aspirations.

Sensitivity to the requirements of different kinds of research data

7. A third requirement is for sensitivity to the hugely varied kinds of digital research data – from texts and numbers to audio and video streams – and to how they were generated

   - for different purposes and through different processes

     □ scientific experiments, which may in principle be reproduced, although it may in practice prove difficult, or not cost-effective, to do so

     □ models or simulations, where it may be more important to preserve the model and associated metadata than the computational data arising from the model

     □ observations – from the astronomical to the zoological – of specific phenomena at a specific time or location, where the data will usually constitute a unique and irreplaceable record;

     □ derived data, resulting from processing or combining "raw" or other data (where care may be required to respect the rights of the owners of the raw data);

     □ canonical or reference data relating, for example, to gene sequences, chemical structures, or literary texts

   - by different groups of people and organisations

     □ from the research community itself in the course of research

          □   from a variety of bodies in the public, private and voluntary sectors for a wide range of purposes, where the data may nevertheless be of value for research

      ■   collected together for different reasons

          □   for the benefit of those engaged in a specific project, where some or all of the data may or may not retain a value beyond the life of the project

          □   for the benefit of a wider group within a discipline, or across disciplines, to provide reference information, or a basis for evidence-based policy-making.

8.   As is implied above, some of the data generated or collected in the course of research are of little value, and one of the key issues we address in these guidelines is the selection of those data which should be made accessible to others, and where appropriate preserved for the long term. It should be noted also that the scholarly communications process itself, of course, may lead to the modification of digital research data, or to the generation of new data in the course of selecting, processing, disseminating and preserving original data created by research.

International requirements and developments

9.   In the increasingly international context for research and information, policy and practice in the UK must take full account of international requirements and developments; and the UK should actively participate in relevant international policy forums. There have been significant developments in countries including the US, with the publication of reports by the National Science Board and the National Science Foundation[2];and in Australia, with the establishment of the Australian Research Information Infrastructure Committee (ARIIC) with a remit that focuses on improving researchers' access to information[3], and the development of proposals for a national data service[4].

10.   At international level, the OECD has also taken an interest, and following a Ministerial Declaration on Access to Research Data from Public Funding in 2004[5], has produced guidelines for governments and research organisations to facilitate data sharing among researchers, institutions, and national agencies. More recently, the European Union has underlined the importance of access to research data for analysis and use beyond what the originator had envisaged, and called for greater co-ordination of policy and practice between member states.[6]

How does the framework relate to other policy developments and frameworks?

11.   The framework draws on the principles set out in the RCUK Position Statement on Access to Research Outputs and in the OECD Declaration, as well as on the work of the Digital

---

[2] *Long-lived Data Collections: Enabling Research and Education in the 21st Century*, National Science Board, Washington, October 2005, available at http://www.nsf.gov/pubs/2005/nsb0540/nsb0540.pdf; *New Cyberinfrastructure for 21st Century Discovery*, National Science Foundation, January 2006, available at http://www.nsf.gov/od/oci/ci_v5.pdf; and *Cyberinfrastructure Vision for 21st Century Discovery*, National Science Foundation, March 2007, at http://www.nsf.gov/pubs/2007/nsf0728/index.jsp.

[3] See
http://www.dest.gov.au/sectors/research_sector/policies_issues_reviews/key_issues/australian_research_information_infrastructure_committee/default.htm

[4] *Towards the Australian Data Commons*, October 2007, available at
http://www.pfc.org.au/twiki/pub/Main/Data/TowardstheAustralianDataCommons.pdf

[5] The Declaration is based on ten principles: openness, transparency, legal conformity, protection of intellectual property, formal responsibility, professionalism, interoperability, quality and security, efficiency, accountability. The guidelines are available at http://www.oecd.org/dataoecd/9/61/38500813.pdf

[6] See *Council Conclusions on scientific information in the digital age: access, dissemination and preservation*, November 2007, available at http://www.consilium.europa.eu/ueDocs/cms_Data/docs/pressData/en/intm/97236.pdf.

Preservation Coalition and the Digital Curation Centre, and on discussions with those Research Councils that are developing policy in this area. The RIN will continue to monitor developments of this kind, to gather evidence on changes in policy and practice, and to broker relationships with the aim of promoting greater co-ordination in the development of policies and services.

Won't a framework like this be very costly to implement?

12. There will of course be significant costs associated with implementing the framework, and both funders and research institutions will need to assess priorities in providing the necessary resources. But the bigger cost, although a hidden one, lies in losing valuable data or in duplicating research that has already been done because the data are not available. Cost-effectiveness in one of the key principles on which the framework is founded. One of our key aims is to maximise the benefits from investments already being made in the stewardship of data; but more significantly, the benefits and cost-effectiveness of the very considerable investment in research itself.

13. Neither costs nor benefits have been clearly identified at present, and so it is difficult to construct a clear business case for investment in the infrastructure of services necessary to ensure that research data are properly managed. It is important that assessments of costs and of the evidence of value are built into pilot and early implementation projects.

Isn't it premature to develop such a framework?

14. Unless we address the issues now, there is the real risk that valuable data will be damaged irreparably or lost. Tackling all the issues across the whole of the research sector is, of course, impracticable. But an the development of an overall policy framework should provide a helpful context for pilot projects and early implementation in specific areas. Such work should also help in moving towards greater clarity as to the roles and responsibilities of different stakeholders and agents.

The five principles

15. The five principles are set out below in an order that reflects the lifecycle through which digital research data are created, used and made accessible over the long term. We believe that the principles should be adopted by universities and other research institutions, libraries and data services, publishers, research funders, as well as researchers themselves. Each principle is accompanied by a list of key issues and questions to be addressed, and an indication of where primary responsibility lies in addressing them.

Principle 1: roles and responsibilities

16. Without clarity as to the roles and responsibilities of different agents, there is the risk of misunderstandings, of wasted efforts, and growing difficulties in managing and providing access to digital research data. Increases in the volume and complexity of data mean that traditional informal arrangements alone are no longer adequate to ensure the effective stewardship of data, and may need to be complemented by more formal codes.

17. In order to facilitate the development of a co-ordinated set of policies and procedures covering the roles and responsibilities of the key players and stakeholders in relation to all the foregoing principles, it is essential that

   The roles and responsibilities of researchers, research institutions and funders should be defined as clearly as possible, and they should collaboratively establish a framework of codes of practice to ensure that creators and users of research data are aware of and fulfil their responsibilities in accordance with the principles set out in this document.

18. Roles and responsibilities in ensuring that data are properly looked after are shared across all the key stakeholders, including researchers themselves and the support services in libraries, archives and data centres. Precise arrangements will necessarily vary to reflect the range of practices in different disciplines, funding contexts and institutional settings. Moreover, both needs and roles may change, or new ones emerge. But research funders and research institutions must play a critical role in setting the policy framework and determining the processes through which the policies are implemented. In most cases, the main role of funders will be to set the broad policy framework, rather than detailed procedures, where the default responsibility rests with research institutions, or with their research groups or departments; but funders may have a more active role if they are responsible for running sustainable data centres.

   ■ Are the roles and responsibilities of researchers, institutions, funders and other agents clearly set out and understood? *(funders; research institutions)*

19. Codes of practice for handling research data should be should be seen as an essential part of the codes now being developed to cover research governance in general; but they must be light in touch and based on mutual trust between researchers, their institutions, and their funders. A framework of codes of this kind would help to minimise complexity and regulatory burdens in arrangements for the stewardship of data. In addition to the other matters covered in this paper, the codes should address such matters as ethical and legal requirements relating to data protection and freedom of information, confidentiality (both corporate and personal), privacy, intellectual property rights, and national security.

   ■ Do the arrangements for research governance set by funders and/or applied in your institution include agreed codes of practice for those who create, collect or use digital research data? *(funders; research institutions)*

   ■ Do the codes of practice cover the issues arising from the principles set out in this document, and other relevant ethical and legal issues? *(funders; research institutions)*

Principle 2: standards and quality assurance

20.　If we are to optimise the value and potential use of digital research data, it is essential that

> ***Digital research data should be created and collected in accordance with applicable international standards, and the processes for selecting those to be made available to others, should include proper quality assurance.***

21.　Effective management of digital research data begins early in the research process, to ensure that data are created and collected to high standards, and that thought is given to whether, how and when they should be made available to others. Funders themselves need to be closely associated with the development and setting of appropriate technical and procedural standards, ensuring that common data formats are used wherever possible, but also that they are sufficiently adaptable to avoid the risk that innovation in data creation and collection is unduly constrained. Precisely how the data are managed depends among other things on the extent to which researchers themselves own or control the data in question, and the policies of the relevant funding bodies.

22.　The adoption of common standards and formats wherever possible is of critical importance in ensuring usability. Especially where data are conceived as a primary output of research, and intended for re-use, user-testing is essential.

- Are policies and procedures in place to ensure that appropriate standards are specified for each of the main categories of data that researchers create and collect; and for effective adoption of those standards? *(funders; research institutions)*

- Are policies and procedures in place to ensure that data standards and quality assurance are explicitly addressed in research proposals, and that data are created and collected in formats that facilitate interoperability with other datasets? *(funders; research institutions)*

- Where appropriate, are procedures for user-testing of data formats in place? *(funders; research institutions)*

23.　Where data are to be made widely available, researchers and other potential users need to know that they are authentic, reliable, and of high quality. Hence consideration should be given to whether and how data quality can most effectively be assured and assessed, through peer review or other procedures; and how that assurance can most effectively be conveyed to users, especially through the recording of provenance and lineage.

- Have clear policies and procedures been defined and implemented to ensure that data made available to others are authentic, with clear documentation of origin, including the methods, instruments and techniques used in their creation or collection? *(funders; research institutions)*

- Have arrangements been agreed and implemented where appropriate for peer review of the content and formats of data to be made available to others, and for providing a clear indication to users that the data have been subject to such review? *(funders; research institutions)*

Principle 3: access, usage and credit

24. In order to achieve the fundamental policy objective of availability and accessibility, it is essential that

    Digital research data should be easy to find, and access should be provided in an environment which maximises ease of use; provides credit for and protects the rights of those who have gathered or created data; and protects the rights of those who have legitimate interests in how data are made accessible and used.

25. The policy objective requires that the research community as a whole, and any others who have an interest in the data should have timely, user-friendly access to relevant data, and at the lowest possible cost. Free and open access, without restriction as to use, should be the default option wherever possible. But the access objective does not imply that all data should be made available immediately to all those who may have an interest in it, and some restrictions on access may be legitimate or necessary. Indeed, there is a need to balance conflicting interests and rights, where some rights – of research institutions or funders, data owners or subjects, or researchers themselves – may trump others. Hence, the requirement may frequently be for access to be provided in a managed environment; and in order to achieve that, a number of issues must be addressed, in addition to those set out under Principle 2.

    ### Data Creators and Collectors

26. Especially where it is decided that data should be made widely available, problems may well arise unless it is clear who owns the data, what rights are associated with that ownership, and who else may have rights associated with the data.

    - Are procedures in place to determine ownership and for identifying and managing data rights? *(funders; research institutions)*

    - Is clear guidance available on copyright and other instruments of intellectual property (including database right, which applies in the EU but not in the US)? *(funders; research*

27. The ability of users to find and re-use data depends on making available to them a full record of information relating to the content, structure, context and source of data that may be relevant to their needs. Such information – data about data – is commonly referred to as metadata. Wherever possible, metadata should be created and made available in accordance with recognised international standards. But the essential requirement is as full a record as possible of the context in which the data were created or collected. Without such information, users cannot be aware of the precise nature and limitations of the data (including any modifications made since their original creation), and will not be able to interpret them properly, or use them effectively.

    - Have procedures been determined and implemented to provide for the benefit of users appropriate information about the nature and context of the data? *(funders; research institutions)*

    - Are metadata machine-understandable, and in compliance with recognised standards? *(funders; research institutions)*

28. Data are of no use without the facilities, software applications and other tools required to access and use them. In some cases it may be reasonable to assume that potential users have ready access to the relevant applications and tools. But frequently neither access nor use of data will be possible without special steps to preserve and/or make the necessary applications and tools

available; and in some cases, the sophisticated tools required for access and use many need to be future-proofed beyond the likely life of the group that created the data.

■ Have procedures been put in place to ensure that the facilities, software applications, user interfaces, and other tools necessary to access and make use of data generally available to users, including where appropriate in the very long term; or are special measures required to provide them? *(research institutions; funders; data centres)*

29. The access objective implies also a clear process for determining who should be given access to what data, and on what terms. Otherwise there is the danger of either unwarranted access or unnecessary restrictions. Access therefore requires research funders, researchers, and their employing institutions to decide how the benefits of access to data can be optimised within an agreed and clear access management regime, while providing safeguards against mis-use, whether intentional or unintentional. Authentication and authorisation of users needs to be easy and clear, and well-integrated with licensing arrangements.

■ Are there arrangements to determine what is made available, in conformity with legal, ethical and other considerations? *(funders; research institutions)*

■ Are there arrangements to set levels of access – including read-only, manipulation and annotation rights – and the authentication and authorisation processes necessary to secure that access? *(funders; research institutions)*

■ Is it clear who is responsible for the necessary decisions, and on what criteria? *(funders; research institutions)*

■ Are safeguards in place to protect against mis-use of data? *(funders; research institutions)*

Data Services, Publishers and Users

30. Researchers and others who wish to use data need services through which they can in user-friendly fashion authenticate themselves and secure the authorised level of access.

■ Have service providers established procedures and audit trails to establish who owns each dataset, and who may have and who has had access to it? *(institutions; repositories and data centres)*

■ Are authentication and authorisation arrangements clear, and are users required to confirm their acceptance of the terms and conditions of access? *(repositories and data centres)*

■ What procedures are there to monitor use and to ensure that users act in accordance with those terms? *(repositories and data centres)*

31. Increasingly, users wish to gain access to digital research data alongside publications, and to be able to link the two, in order to assess the evidence on which reported results are based. Such linking also facilitates the granting of credit to data creators.

■ Are arrangements in place to ensure that the data underpinning research findings reported in scholarly publications are made accessible to readers, and that references to data are included in relevant papers, along with appropriate links? *(funders; publishers; research institution; repositories and data centres)*

Credit, Citation and Evaluation

32. As the use and re-use of data become more common, it is critically important that researchers should receive appropriate credit for the work that has gone into creating or collecting data, and

for its use by others. Citation of digital research data (alongside citation of publications) is becoming common in some areas, and is likely to become more so. Indeed, citation and use of data, and the allocation of credit, will play an increasingly important part in debates about the evaluation of research quality.

- Have clear protocols been established for the citation of datasets of different kinds, including such matters as formats for referencing, identification of different versions, and linkages between published outputs and underlying datasets? *(funders; learned societies and professional bodies; repositories and data centres; publishers)*

- Are there clearly-understood arrangements for allocating credit to data creators in recognition of the value and the use of their data? Are such arrangements integrated with research assessment processes? *(funders; learned societies and professional bodies; research institutions)*

- Have appropriate bodies established mechanisms and criteria for evaluating the importance and significance of different datasets and their use? *(funders; learned societies and professional bodies; research institutions)*

- Have the roles and responsibilities of the various bodies concerned in the development of protocols, criteria and other arrangements been clearly delineated and agreed? *(funders; learned societies and professional bodies; research institutions)*

Principle 4: benefits and cost effectiveness

33.   In order to maximise the benefits and minimise the costs of managing and providing access to digital research data, it is essential that

The models and mechanisms for managing and providing access to digital research data must be both efficient and cost-effective in the use of public and other funds.

34.   Managing, preserving, and providing access to digital research data involves significant costs; and these are likely to increase as the volumes of data increase, and the complexity of the tasks become more apparent. This will present significant challenges to research institutions and funders. But the risks of not rising to these challenges are the loss not just of data, but of the potential benefits in enhancing the overall efficiency of research; avoiding unnecessary duplication of effort and needless costs in experimentation, observation, or the creation of new data; and improving knowledge transfer. As data management becomes increasingly integral to the research process itself, **all** those with responsibilities for data – researchers, research institutions, library and data services, and research funders – need to balance the costs and the benefits, and to ensure that their policies and practices operate cost-effectively.

35.   Effective and efficient management of data requires investment in specialist professional support services, to ensure that data are properly selected and stored, that they can readily be accessed, and that their integrity can be assured over time. These services need to be readily available, so that researchers can get professional advice at all stages of the research process. Without such support, there is the danger that data will be created in unusable forms, managed inappropriately, or stored ineffectively.

- Is there an appropriate and sustainable level of investment in professional support services, whether on a centralised or a distributed model? *(funders; research institutions)*

- Are both researchers and professional support staff appropriately trained, and are there effective arrangements for promoting good practice? *(funders; research institutions)*

- Are there appropriate incentives for researchers to make use of good professional services? *(funders; research institutions)*

36.   Unless data management and access regimes are sustained as key elements in the infrastructure for research, there is the risk that the overall efficiency of research will fall. But funders face growing challenges in determining the balance of their overall funding between support for research activity itself, and for the necessary infrastructure of data and related services. Hence it is crucial that data management and access regimes and services should be subject to regular and systematic review and evaluation, to ensure that the benefits to researchers and other users are realised, and that needless costs are avoided. Such reviews should assess how benefits to researchers and others might be enhanced as well as how costs and other burdens might be reduced.

- Have frameworks been established to ensure that there is a sound basis for assessing and making judgements about the benefits as well as the costs, and the performance and sustainability of all the processes and parties involved in creating, managing, providing access, and using research data? *(funders; research institutions)*

Principle 5: preservation and sustainability

37. As data plays an increasingly central part in the conduct of research, it will become increasingly important that

     Digital research data of long term value arising from current and future research should be preserved and remain accessible for current and future generations.

38. As we argue throughout this paper, research data should be considered as a valuable resource. But by no means all digital research data are of long-term value. Many datasets may have little value beyond the life of a specific project. It is essential in the interests of all those involved in supporting, promoting and undertaking research to ensure that there are effective procedures for determining which data are of continuing value, and ensuring that arrangements for their stewardship are sustainable – not least the training and supply of a cadre of specialist curation personnel. Otherwise there is the danger of loss or damage to valuable data.

     ■ Are there clear procedures and criteria for appraising the value of datasets, and determining which should be selected for preservation, by whom, in what repository (local, national or international), for what period, in what format, and for what purposes? *(funders; research institutions)*

     ■ Are there sustainable procedures to ensure that the data are protected through explicit security protocols from unauthorised modification, damage or destruction? *(funders; research institutions; repositories and data centres)*

     ■ Are appropriate policies, infrastructure, training and other arrangements in place to ensure that data are properly managed and maintained in formats that support continued storage, access, including, where appropriate, migration and emulation? *(funders; research institutions; repositories and data centres)*

39. Some datasets are of such strategic significance and value that they need to be preserved and made accessible over the very long-term. In order to focus effort and attention on such data, and to guard against loss, clear procedures and criteria are needed, developed in consultation with key stakeholders, to identify data of this kind, and to ensure their preservation in an appropriate repository.

     ■ What arrangements are there to ensure that data of potentially high value over the very long term are transferred to a recognised specialist repository, that they are complete, and free from corruption? *(funders; research institutions; repositories and data centres)*

40. As digital research data are increasingly updated, amended and annotated over their lifecycle both by data creators and by subsequent users, provenance protocols and audit trails are needed to indicate clearly who has annotated or amended data, how and when. Otherwise there is the danger that data will be misinterpreted or used inappropriately.

     ■ Have service providers established protocols and audit trails to show who has had access to each dataset, and who has enhanced or annotated it (and how and when)? *(repositories and data centres)*

Annex: list of organisations that responded to the RIN consultation on the draft of this document

- Aberystwyth University
- Aston University
- University of Bath
- University of Bedfordshire
- University of Brighton
- University of Cambridge
- University of Central Lancashire
- Coventry University
- Data Information Specialists Committee (DISC-UK)
- University of Exeter
- University of Glasgow
- Heriot-Watt University
- Institute of Education
- Joint Information Systems Committee (JISC)
- Keele University
- Lancaster University
- University of Leeds
- University of Leicester
- London School of Hygiene and Tropical Medicine
- Loughborough University
- Manchester Metropolitan University
- National Cancer Research Institute
- Northumbria University
- University of Nottingham
- Open University
- Queen Mary University of London
- Queen's University Belfast
- Roehampton University
- Sheffield Hallam University
- University of Stirling
- University of Sunderland
- University of Sussex
- Swansea University
- University College London
- University of Warwick
- University of Wolverhampton