

Research data are important...

Research data are an increasingly important and expensive output of the scholarly research process, across all disciplines. They are an essential part of the evidence necessary to evaluate research results, and to reconstruct the events and processes leading to them. Their value increases as they are aggregated into collections and as they become more available for re-use to address new and challenging research questions. But we shall realise the value of data only if we move beyond research policies, practices and support systems developed in a different era. We need new approaches to managing and providing access to research data.

In order to address these issues, the RIN established a group to produce a framework of key principles and guidelines, and we consulted on a draft document in 2007. The framework is founded on the fundamental policy objective that ideas and knowledge, including data, derived from publicly-funded research should be made available for public use, interrogation, and scrutiny, as widely, rapidly and effectively as practicable.

The framework, of which this document is a summary, is structured around five broad principles which provide a guide to the development of policy and practice for a range of key players: universities, research institutions, libraries and other information providers, publishers, and research funders as well as researchers themselves. The full framework document, which is available at www.rin.ac.uk/data-principles, then introduces under each of the principles, a series of questions which serve a practical purpose by pointing to how the various players might address the challenges of effective data stewardship.

In seeking to develop the framework further, all parties need to work collaboratively and to ensure that it is sensitive to the needs of researchers and the different contexts in which they work. All parties must also take account of relevant technical and policy-making developments in the UK and overseas. Some key recent developments initiated by, among others, JISC and HEFCE in the UK, by the NSF in the US, and by the Department of Education, Science and Training in Australia, are listed at www.rin.ac.uk/data-principles.

The principles

Principle 1- roles and responsibilities

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.

Without clarity as to these roles and responsibilities, 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 informal arrangements alone are no longer adequate to ensure the effective stewardship of data, and need to be complemented by more formal codes. Such codes of practice – light in touch and based on mutual trust between researchers, their institutions and their funders – should be seen as an essential part of research governance.

Principle 2 – standards and quality assurance

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.

As research data become more widely available, it is the more important that users should be offered some assurance as to their nature and quality. 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. The adoption of common standards and formats wherever possible is of critical importance in ensuring usability.

Principle 3 – access, usage and credit

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.

The fundamental policy objective of making data available and accessible requires that users should have timely, user-friendly access to relevant data, and at the lowest possible cost. There is a need to balance conflicting interests and rights, but wherever possible, free and open access should be the default option. There are particular implications here for:

- *Data creators and collectors:* it is important to make clear how the data were created or collected, who owns them, what rights are associated with that ownership, who else may have rights associated with the data, who should be given access, and on what terms.
- *Data services, publishers and users:* services are needed through which users can easily authenticate themselves and secure access; increasingly, users wish to gain access to digital research data alongside publications, and to be able to link the two.
- *Credit, citation and evaluation:* 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.

Principle 4 – benefits and cost-effectiveness

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.

Managing, preserving, and providing access to digital research data involves significant costs which will increase as the volumes of data increase. This presents significant challenges to research institutions and funders. Effective and efficient management of data requires investment in infrastructure and 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. The costs and benefits of such investment must be regularly and systematically reviewed – but it is clear that without investment in effective data management and access regimes, the overall efficiency of research will fall.

Principle 5 – preservation and sustainability

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

Effective procedures are required to determine which data are of continuing value, and sustainable arrangements for their curation and preservation. Some datasets are of such strategic significance that they should be preserved and made accessible over the very long-term. As digital research data are increasingly updated, amended and annotated over their lifecycle, provenance protocols and audit trails are needed to indicate clearly who has annotated or amended data, how and when.