

Guidance on best practice in the management of research data

UPDATES

In July 2018 this guidance was updated to reflect changes in data protection legislation only, the following changes were made:

- Page 6: "*complying with Data Protection Act legislation*" was changed to "*complying with Data Protection legislation*"
- Annex A, section 4: The following sentence was deleted, "*The Data Protection Act 1998 contains eight enforceable principles, applying to anyone processing personal data.*"

INTRODUCTION

"By academic freedom I understand the right to search for truth and to publish and teach what one holds to be true. This right implies also a duty; one must not conceal any part of what one has recognized to be true." [Albert Einstein, Source: Letter on his seventy-fifth birthday, 1954]

The RCUK Common Principles on Data Policy¹, published in 2011, set expectations for the systematic and routine management and sharing of research data. This document provides a guide to the interpretation of these common principles, independent of the specific data policy of the Council(s) that may be funding a particular project. Individual Research Councils' data policies and good research practice guidance elaborate further, taking account of differences in approach required for the varying nature of the data within particular domains.

Background

The RCUK data principles build on the Organisation for Economic Co-operation and Development (OECD) publication 'Principles and Guidelines for Access to Research Data from Public Funding'² and are aligned to more recent statements from the EC³, US⁴, the Royal Society⁵ and others. These recognise the growing importance of a strategic approach to the management of research data and are informed by the view that widespread sharing of data will enable researchers, empower citizens and convey academic, economic and social benefits.

They are underpinned⁶ by the maxims that:

- Publicly-funded research data are a public good, produced in the public interest

¹ <http://www.rcuk.ac.uk/research/datapolicy/>

² <http://www.oecd.org/sti/sci-tech/38500813.pdf>

³ EC COMMISSION RECOMMENDATION of 17.7.2012 on access to and preservation of scientific information http://ec.europa.eu/research/science-society/document_library/pdf_06/recommendation-access-and-preservation-scientific-information_en.pdf

⁴ Increasing Access to the Results of Federally Funded Scientific Research, Office of Science and Technology Policy,

Washington Feb 22 2013

http://www.whitehouse.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf

⁵ Science as an open enterprise, The Royal Society, <http://royalsociety.org/policy/projects/science-public-enterprise/>

⁶ <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.92.9016&rep=rep1&type=pdf>

- Publicly-funded research data should be openly available to the maximum extent possible

The G8 definition of Open Data⁷ states that data should be easily discoverable, accessible, assessable, intelligible, useable, and wherever possible interoperable to specific quality standards, while at the same time respecting concerns in relation to privacy, safety, security and commercial interests. The Research Councils wish to promote open, transparent and robust research by encouraging wider data analysis, more data re-use, and the combination of datasets from multiple sources. We believe that an increased emphasis on sharing of data has the potential to stimulate new approaches to the collection, analysis, validation and management of data whilst, at the same time, recognising that the risks to the privacy of individuals must always be considered where data arise from, or are derived from, personal identifiable data.

Expanding access to research outputs, both publications and data, is a key aspect of the Government's transparency agenda⁸ as increased transparency in research can improve the sharing of new ideas raising the prestige of UK research, encouraging innovation and creating new growth opportunities.

⁷ G8 Science Ministers' Statement , <https://www.gov.uk/government/news/g8-science-ministers-statement> , 13 June 2013

⁸ For example, see the BIS Open Data Strategy 2014 to 2016 (<https://www.gov.uk/government/publications/bis-open-datastrategy-2014-to-2016>, 15 July 2014) and The Research Sector Transparency Board (<https://www.gov.uk/government/groups/research-sector-transparency-board>) which has been established to *consider ways in which transparency in research can be a driver for innovation and discovery while furthering the UK's recognised excellence in science*

Principle 1.

Publicly funded research data are a public good, produced in the public interest, which should be made openly available with as few restrictions as possible in a timely and responsible manner.

Systematic management and sharing of research data has many benefits for the research community and the public. These include reinforcing open, transparent and robust academic enquiry, enabling re-use, and combination of datasets from multiple sources. Further, an increased emphasis on sharing of data has the potential to stimulate new approaches to collection, analysis, validation and management of data. On the other hand, responsible datasharing must ensure that there is no violation of privacy or harm to intellectual property.

These guidelines refer to:

- a) data collected or generated as a result of Research Council funding, including through access to research facilities, and
- b) data underpinning papers published as a result of Research Council funding, whatever the source of that data.

Research data includes: digital information created directly from research activities such as experiments, analysis, surveys, measurements, instrumentation and observations; data resulting from automated or manual data reduction and analysis including the inputs and outputs of simulations and models.

Principle 2.

Institutional and project specific data management policies and plans should be in accordance with relevant standards and community best practice. Data with acknowledged long-term value should be preserved and remain accessible and usable for future research.

Research data should be managed throughout the research life-cycle, from creation to preservation or disposal, to enable the value of the data to be maximised.

There are data management roles and responsibilities for both the institution in the provision of infrastructure and procedures, and within specific projects for the data generated or used. Data Management Plans⁹ (DMPs) are an important tool in the management of data. Preparation of a DMP should normally be integral to the planning of a project that is wholly or partially funded by any of the Research Councils, whether or not it is required as part of the formal research grant application¹⁰. Where a full or outline DMP is required by a Research Council as part of a research grant application it will be subject to review. Guidance on the assessment of DMPs is provided to those involved in Peer Review and is available on request from individual Councils.

⁹ These may be also referred to as 'Data Sharing Plans' and 'Technical plans' in individual Research Council Data policies. ¹⁰ Not all Research Councils require a DMP to be submitted with a grant application; some require an outline plan with the application which is then converted to a full plan in the event that a grant is awarded. A summary of requirements is available on the DCC website at <http://www.dcc.ac.uk/resources/data-management-plans/funders-requirements>

DMPs should follow relevant national and international recommendations for best practice and should be prepared in consultation with relevant institutional and disciplinary stakeholders. They should anticipate requirements throughout the research activity, and should be subject to regular review and amendment as part of normal research project management. Annex A gives more detail on topics which a DMP should address.

Research data with potential reuse value, irrespective of whether it has been used for publication, should be subjected to a risk- and value-proportionate assessment of whether to discard or retain the data. For example, data that by their nature cannot be re-measured or recreated such as earth observations or people-based data may often warrant indefinite storage and preservation. On the other hand, there may be cases in which it may not be possible or cost effective to preserve research data. This will depend on the type and scale of the data, their role in validating published results, and their predicted long-term usefulness for further research. For example, in the case of simulated data or outputs of models, it may be more effective to preserve the means to recreate the data by preserving the generating code and environment, rather than preserving the data themselves. Provided that the ability to validate published research findings is not fundamentally compromised, a deliberate decision to dispose of research data at an appropriate time is acceptable in these cases.

In general, the Research Councils expect data that underpins findings in publications should be accessible for at least ten years after publication. Data related to potential future legal liabilities may need to be stored for substantially longer. Individual Research Councils' data policies and good research practice guidance provide additional requirements and should be consulted.

Principle 3.

To enable research data to be discoverable and effectively re-used by others, sufficient metadata should be recorded and made openly available to enable other researchers to understand the research and re-use potential of the data. Published results should always include information on how to access the supporting data.

Discoverability

Research data should be supported by sufficient contextual information to enable others to find what research data exist, why, when and how they were generated, and how to access them (i.e. metadata for discovery). This information should be made available online with terms that permit its full re-use.

Metadata and clear documentation should be made available with research data, and whenever possible include a Persistent Identifier that refers to the research data¹⁰. Where communities have established common or minimum metadata standards they should be adhered to.

Understandability

The normal expected level of data-sharing is to make the original data available, together with their associated metadata and any other supporting documentation, according to the standard accepted procedures within the research field.

¹⁰ For example using a DOI is available through the DataCite organisation - <http://datacite.org> ¹² This is in accordance with RCUK Policy on Open Access (<http://www.rcuk.ac.uk/RCUKprod/assets/documents/documents/RCUKOpenAccessPolicy.pdf>)

Researchers should ensure that metadata created to support retained research datasets is sufficient to allow other researchers a reasonable understanding of those datasets and thereby minimise unintentional misuse, misinterpretation or confusion. For example, the metadata may need to describe the origin, processing, analysis and/or the researcher's management of a dataset.

Accessibility

Research publications arising from Research Council funding should include a statement on how the supporting data and any other relevant research materials can be accessed¹². For guidance on how to cite research data see Principle 6. Data that are not generated in digital format should be stored and referenced in a manner to facilitate sharing in the event of a valid access request being received.

Where access to the data is constrained, for example due to legal, ethical, commercial or security issues, the published metadata should, if permitted, give the reason and summarise any conditions which may need to be satisfied for access to be granted.

Principle 4:

RCUK recognises that there are legal, ethical and commercial constraints on release of research data. To ensure that the research process is not damaged by inappropriate release of data, research organisation policies and practices should ensure that these are considered at all stages in the research process.

There are many measures researchers can take during the research process to facilitate datasharing. Research Organisation policies and practices should maximise opportunities for data-sharing whilst ensuring that neither the research process nor the individual participants' rights are damaged by inappropriate release of data.

Regulatory and legal requirements and, where appropriate, issues of rights, ethics and privacy should be considered during the planning of the research. Appropriate publication plans and data usage agreements are essential elements of research management strategy. Where research data are shared with third parties the terms and conditions should be clear.

If access to the data needs to be restricted, the published metadata should, if permitted, give the reason and summarise any access conditions which may need to be satisfied. Any appropriate permissions (e.g. ethical, legal, institutional) should be in place before sensitive data are shared, and any restrictions should be managed through a verifiable and transparent process.

Legal

There are sometimes legitimate reasons to restrict access to data. Research Organisations must ensure that the publication of data conforms to relevant legislation. Research Organisations should ensure they are aware of relevant legislation in particular regarding data protection, freedom of information and environmental information regulations and disseminate this to their researchers as appropriate.

Ethical

Privacy, confidentiality and consent need to be considered whenever research data include, or are derived from, personal data. This is important not only to satisfy legal requirements but also, crucially, to maintain and build public trust in the use of personal data for research.

Data which are personal, sensitive or confidential may be shared provided attention is paid from the outset to the legal and regulatory requirements and any professional standards are met. These may include:

- complying with Data Protection legislation;
- gaining appropriate consent, including broad consent to maximise data-sharing;
- protection of participants' identities;
- obtaining appropriate ethical approval(s).

Any restrictions to data access should be considered in the data management plan before commencing the research.

Where research depends on combining data from separate datasets that can only be linked through use of personal identifiers, a trusted environment for population-based research and statistical analysis in which the risk of identifying individuals is minimised may be used.

In every case where personal data are involved, data-sharing agreements should be issued and signed by appropriate authorities before data are released or any analyses requested by third parties are performed. These data-sharing agreements should prohibit

- (a) use of the released data to identify participants or otherwise breach confidentiality
- (b) unapproved contact with participants

There may be other ethical reasons to restrict access to data when it is not in the public interest to release data, for example, where the data identify the location of protected species.

Commercial

Research Councils encourage researchers to work in productive, equitable partnerships, for example with charities and industry. Where Research Councils fund academic research jointly with a commercial partner, data provided by the commercial partner may be restricted for commercial reasons. However, publications arising from such research should still include a statement on how and under what conditions the supporting data and any other relevant research materials can be accessed. Researchers engaged in collaborative research should therefore ensure that the governance of who will have access to data is set out in advance in the collaboration agreement. Likewise, researchers should obtain any necessary clearances from collaborators before submitting work for publication. Research Councils expect commercial partners not to restrict publication of research results unless arrangements to do so have been previously agreed with the funders.

Data-sharing with commercial or non-UK based organisations should where possible conform to the same principles and standards required when sharing data within the UK academic community and, in particular, must conform to the requirements of relevant UK and EU legislation. All reasonable steps should be taken to ensure that research data are not held in any jurisdiction where the available legal safeguards provide lower levels of protection than are available in the UK

Researchers should ensure that intellectual property arising from their work is suitably protected and managed, in line with RCUK Knowledge Exchange Principles¹¹. If potentially valuable intellectual property is identified, the Research Councils recognise that it may be necessary to delay data-sharing for a reasonable period to allow time for measures to be taken to protect that intellectual property, for example, for the drafting and filing of patent or

¹¹ <http://www.rcuk.ac.uk/kei/expectation/Pages/kePrinciples.aspx>

licensing applications. However, delays or restrictions on sharing due to managing IP should be minimised as far as possible.

Principle 5.

To ensure that research teams get appropriate recognition for the effort involved in collecting and analysing data, those who undertake Research Council funded work may be entitled to a limited period of privileged use of the data they have collected to enable them to publish the results of their research. The length of this period varies by research discipline and, where appropriate, is discussed further in the published policies of individual Research Councils.

In order to maximise the benefit that can be derived from research data resulting from Research Council funded research, the default position should be that such data are made publicly available in a timely manner, notwithstanding that there may be legal, ethical or commercial reasons to restrict public access to all or part of a specific research dataset (as discussed with reference to Principle 4, above).

A limited, defined period of exclusive use of data may be reasonable in recognition of researchers' intellectual contribution and in order for the research team to have a first opportunity to publish or otherwise exploit the results of their research. Furthermore, sharing should take account of enhancing the long-term value of the data and the potential value of future datasets should not be compromised by premature sharing and analysis of partially complete information.

Timescales for making research data publicly available will depend on the nature of the data, the research methodology being employed and disciplinary best practice, and should generally be no later than the publication of the main findings. Where an individual Research Council policy has specific requirements, or where there is established best practice in a particular discipline, researchers are expected to make research data available accordingly.

The specific plans for sharing of data, and in particular the length of any such period of exclusive use, should be considered from the earliest stages of project planning and set out in the Data Management Plan. Any period of exclusive use should be balanced against the public interest in release and may be tested if someone makes a formal request for the data. Possible valid reasons for proposing a period of exclusive use might include:

Pending publication

In some fields, restricting access to research data is necessary to protect the research process by allowing researchers to gain credit for their work through publication. A reasonable amount of time to analyse their research data and publish their findings is therefore justifiable. In these cases, and in the absence of any overriding legal, ethical or commercial constraints, it is expected that the research data relied on to support the published findings will normally be made publicly available no later than the publication of those findings.

Effort required to produce data

It is recognised that the creation of some research data may represent the culmination of an enormous effort over many years spanning the design, constructing and operation of

specialised research infrastructure. In such cases, it may be reasonable for the researchers involved to benefit from longer periods of exclusive access to the infrastructure and to the research data it creates in recognition of the integrated effort required to produce it. However, where possible, such exclusivity should not obstruct the advance of knowledge by others in relation to questions outside the scope of the original research.

Data Preparation

There is little point in releasing research data that cannot be understood and potentially reused by others due to a lack of adequate annotation or metadata. Therefore, while it is reasonable not to release research data until such time as sufficient preparation has taken place, it is expected that research management plans will, as a matter of course, allocate sufficient resource to ensure that research data are suitably prepared within a reasonable time. Such preparation should not be used spuriously to unreasonably delay the release of data that are of interest to others.

On-going Data Collection

Data from research spanning a long time scale may be released in batches as they become available or as updated results are published. In these cases, care should be taken that ongoing research contributing to the completion of datasets is not compromised by premature or opportunistic sharing and analysis. Sharing should always take account of enhancing the long-term value of the data.

Protection of Intellectual Property

Research results should be evaluated for possible commercial exploitation. As stated in Principle 4, when potentially valuable intellectual property is identified, it may be appropriate to delay data to allow time for measures to protect the intellectual property to be taken. However, commercialisation of research results does not preclude data-sharing and should not unduly delay it. Potential commercialisation opportunities should, as far as possible, be anticipated in the data management plan and data-sharing managed accordingly to maximise research benefits without jeopardising such opportunities.

Recognition of Intellectual Contribution

In some research, a novel approach to data collection, for example the introduction of a new research methodology, represents a significant part of the overall intellectual contribution. Such novel approaches may also result in the creation of datasets that permit a much wider range of research questions to be considered than can be addressed by the researchers who developed the methodology.

Although delaying the release of novel research methodologies and datasets may be justified, for example in order to allow researchers to gain credit for their work through the analysis and publication of research findings as described above, and this may indeed be the norm in some fields, a balance needs to be struck as such delays can unnecessarily restrict the opportunities of others to use the same novel methodologies and/or datasets for other purposes. Such delays may also unnecessarily delay the recognition due to those responsible for the innovation.

Researchers are therefore encouraged to publish details of novel methodologies and datasets at the earliest opportunity, if possible, even in advance of publishing their findings. In this way they can facilitate other research and sustain the creation of new knowledge by

July 2015

enabling others to build on their work and be given early credit for their contribution of the new data and/or methodology.

Principle 6.

In order to recognise the intellectual contributions of researchers who generate, preserve and share key research datasets, all users of research data should acknowledge the sources of their data and abide by the terms and conditions under which they are accessed.

The researchers who create primary research data, and those who actively enable the release and sharing of that research data, make a valuable contribution to the research process. They should receive full and appropriate recognition by funders, their Research Organisations, their peers and any other users of the primary data.

Proper citation of data sources is extremely important because it:

- directly acknowledges the contribution of others;
- makes identifying and locating the research data easier;
- promotes the reproduction of research results;
- allows the impact of specific research data to be tracked;
- provides a structure through which data creators can be recognised and rewarded.

To facilitate this recognition, published research findings should cite the data on which those findings rely. To enable robust and unambiguous citation, research data sets should be deposited in appropriate data repositories and assigned Persistent Identifiers¹² at a level of granularity informed by disciplinary norms and the envisaged functional use and versioning of the identified data.

Data citations should acknowledge all those known to have made significant contributions to the creation, preservation and sharing of the data, and may refer to databases which assemble data from multiple contributors. A data citation should include sufficient information to enable the exact version of the data being cited to be located, for example by using a Persistent Identifier to ensure that the cited data can still be found even if its location changes.

As with citation of an academic article, a data citation does not and should not be read as in any way suggesting that the data provider endorses the user or their use of the data.

A general statement giving non-specific credit to sponsors or distributors is not a replacement for a proper citation¹³.

Principle 7

It is appropriate to use public funds to support the management and sharing of publicly funded research data. To maximise the research benefit which can be gained from limited budgets, the mechanisms for these activities should be both efficient and cost-effective in the use of public funds.

Research data management services and e-infrastructure are an essential component of the modern research environment which Research Organisations are responsible for making

¹² For example a DOI registered with the global DataCite network (<http://datacite.org>). Many data repositories assign such DOIs to datasets at the time of deposit. Lists of data repositories are available at <http://databib.org/index.php> and <http://www.re3data.org/>

¹³ More information about the importance of data citation and the role of DOIs is available on the UKSDS website (<http://ukdataservice.ac.uk/use-data/citing-data.aspx>) and in the brochure 'Data Citation: what you need to know' (http://ukdataservice.ac.uk/media/104397/data_citation_online.pdf). While this primarily reflects the perspective of economic and social research it is broadly applicable across all research areas.

available to their researchers. The Research Councils recognise that there are costs associated with the provision of such services and infrastructure, not only related to the production of the data but also to ensuring that it is maintained and is accessible for the requisite period of time to ensure the full value of the resource is delivered.

Research Organisation Costs

Research Organisations are encouraged to seek to recover costs that will be incurred in relation to research data arising from Research Council grants. Such costs could be associated with the production, curation and sharing of the research data according to the guidelines (which may be data-specific) published by the Research Councils, and may for example include (but is not limited to) staff time, software, hardware and third-party storage services.

There are three areas of Research Council grants where these costs could be sought.

- Directly Incurred Costs
- Directly Allocated Costs
- Indirect Costs

Effective use of Data Management Plans can help Research Organisations ensure appropriate use is made of each of these cost-recovery options on a project-by-project basis. The following sections outline how each cost recovery mechanism may be used. Additional guidance, previously set out as responses to questions addressed to research funders by the Research Data Managers Forum on 25th April 2013, is available online¹⁴.

Directly Incurred Costs

As long as adequate justification is provided within the grant application, any element of research data management may be included as a directly incurred cost. The justification should demonstrate the value for money, both in relation to method chosen and also to the value of the data.

If the grant is awarded, all expenditure on directly incurred costs must take place before the actual end date of the project and must be fully auditable. Subject to the use of an auditable charging model, it is permissible for a Research Organisation RDM service to recover operational costs from active research grants by invoicing for defined services. The business model may include charges to cover storage for a defined or an indefinite period. However, at a Research Organisation level, no charging model should allow grants to be invoiced for services as direct incurred costs that are also recovered through the Research Organisation's indirect cost rate.

Where Research Council guidelines indicate data arising from a research grant should be deposited and/or made accessible via a subject-specific data repository directly supported by one or more Research Councils, the allowable expenditure on the research grant is limited to the cost of preparing the data for deposit and ingestion to the standard specified by that repository.

¹⁴ Previously issued at: <http://blogs.rcuk.ac.uk/files/2013/07/RCUK-Responses-to-DCC-RDMF-Funder-Questions-.pdf>

Directly Allocated Costs

It is permissible to use Directly Allocated costs for activity relating to the management of data.

Directly Allocated costs are the costs of resources used by a project which are shared by other activities. They are charged to projects on the basis of estimates rather than actual costs and do not represent actual costs on a project-by-project basis.

Such costs can only be included in relation to activity taking place within the dates of the grant project. The proposal needs to provide a justification for the level of resource needed to undertake the activity.

Indirect Costs

While a Research Organisation cannot vary its Indirect Costs on a grant-by-grant basis to reflect the variation of data management related costs between different projects, it can choose to include the cost of data management provision across the Research Organisation as one of its drivers in calculating their Indirect Cost rate.

Factoring in these costs would allow the Research Organisation to cover the more long-term costs associated with the management of its data, including the infrastructural and administrative costs of supporting research projects. However, where such costs are factored in, they may not then also be applied for as direct costs on research grants.

Annex A: Guidelines on data management plans

The table below outlines some topics that a data management plan¹⁵ should consider addressing. Specific Research Council Guidance¹⁶ should also be consulted. This information can be found:

- AHRC: <http://www.ahrc.ac.uk/Funding-Opportunities/Research-funding/RFG/Applicationguidance/Pages/Technical-Plan.aspx>
- BBSRC: <http://www.bbsrc.ac.uk/datasharing>
- EPSRC: <http://www.epsrc.ac.uk/about/standards/researchdata/>
- ESRC: <http://www.esrc.ac.uk/about-esrc/information/data-policy.aspx>
- MRC: <http://www.mrc.ac.uk/research/research-policy-ethics/data-sharing/datamanagement-plans/>
- NERC: <http://www.nerc.ac.uk/research/sites/data/dmp/>
- STFC: <https://www.stfc.ac.uk/1930.aspx>

Additional guidance is also available from Digital Curation Centre¹⁷ and the UK Data Service¹⁸

A data management plan is not a static document and will evolve over the lifetime of the project. However, certain areas should be considered from the start of project planning and these are highlighted in the template below.

¹⁵ The description of each topic is drawn from Research Council Guidance. The structure is organised using the DCC DMP

Checklist <http://www.dcc.ac.uk/resources/data-management-plans/checklist>

¹⁶ Note that: individual Research Councils may require the use of a specific DMP template when submitting funding applications.

¹⁷ <http://www.dcc.ac.uk/resources/data-management-plans/>

¹⁸ <http://www.ukdataservice.ac.uk/manage-data/plan/planning.aspx>

1. Scope

Plans should cover all research data expected to be produced as a result of a project or activity, from 'raw' to 'published'. They may include details of:

- an analysis of the gaps identified between the currently available and required data for the research
- anticipated data volume
- anticipated data type and formats including the format of the final data
- measures to assure data quality
- standards (including metadata standards) and methodologies that will be adopted for data collection and management, and why these have been selected
- relationship to data available from other sources
- anticipated further/secondary use(s) for the completed dataset(s)

The Research Councils recognise that different approaches to data sharing are appropriate in different situations and that the usefulness of data beyond its original purpose will vary according to the type of data collected. The data management plan should make an evaluation as to which data and representations of that data can be preserved in practice, and subsequently which data and

representations are of potential value to others and should also be made accessible. The criteria to be considered are the cost of preservation (including the cost of ensuring the data are usable), the likelihood of re-use, the feasibility of reuse by a third party, the potential value, and the cost of re-creating the data.

2. Principles, Standards and Methodologies

Plans should specify the principles, standards and technical processes for data management, retention and preservation that will be adopted. These may be determined by the area of research and/or funder requirements. Processes should be supported by appropriate standards addressing confidentiality and information security, legal compliance, monitoring and quality assurance, data recovery and data management reviews where suitable.

In order to maximise the potential for re-use of data, where possible, researchers should generate and manage data using existing widely accepted formats and methodologies.

Plans should provide suitable quality assurance concerning the extent to which 'raw' data may be modified. Where 'raw' data are not to be retained, the processes for obtaining 'derived' data should be specified and conform to the accepted procedures within the research field.

3. Documentation and Metadata

Researchers should ensure that appropriately structured metadata, using a recognised or de facto standard schema where these exist, describing their research data are created and recorded in a timely manner. The metadata should include information about regulatory and ethical requirements relating to access and use.

Protocols for the use, calibration and maintenance of equipment, together with associated risk assessments, should be clearly documented to ensure optimal performance and research data quality. Where protocols change, they should be version controlled and the current version should be available and readily accessible.

Documentation may include: technical descriptions, code commenting; project build guidelines; audit trail supporting technical decisions; resource metadata. Not all types of documentation will be relevant to all projects and the quantity of documentation proposed should be proportionate to the anticipated value of the data.

Plans should consider dependencies on software and whether specific software packages and/or bespoke code will need to be made available to enable the data to be read and interpreted and results reproduced or reanalysed.

4. Collaborative Research, Ethics and Legal Compliance

There should be clarity on the ownership and custodianship of research data used or created in the course of the research, including provision for the event that researchers move from one organisation to another. Projects that involve collaboration between separate legal entities should have formal agreements to clarify responsibilities and arrangements for access to data and managing permissions. Such agreements should be developed at an early stage and should take account of any applicable funding terms and conditions.

Where sensitive data (e.g. personal data and/or arising from biological samples) are involved, the terms of the consent should be considered and the data management plan should ensure that appropriate security measures (e.g. in line with Data Protection legislation) are adopted when handling and storing data. Researchers have a duty to understand and comply with the applicable laws, and/or to seek advice.

Where research data are subject to restricted access a data management plan should refer to and, if necessary, describe the security and quality controls that will be implemented, and provide assurance that these are specifically assigned responsibilities.

See also related guidance under Principle 4.

5. Storage and Backup

Data management plans should provide for all retained data and related materials to be securely preserved in such a way as to allow them to be accessed, understood and used by any others having appropriate authorisation in future.

Data held electronically should be backed up regularly and duplicate copies held in alternative locations in a secure and accessible format where appropriate.

6. Deposit, Retention Periods and Long-Term Preservation

Some guidance on retention periods is given in the section on Principle 2.

Unless there are compelling reasons not to do so, Research Councils expect research data to be managed through an established repository such as one provided directly or indirectly by a Research Council, a University, a Laboratory or an independently managed subject specific database. The chosen repository should be the one most likely to maximise the research value obtained from aggregation of related data. It may be appropriate to choose different repositories for data from different stages of a project, for example raw data might be deposited in a facility repository and the derived/analysed data deposited in a subject specific database.

Plans should specify which data are to be deposited in which repository, where and for how long, with appropriate justification. They should include specific provision for any data preparation that may be required prior to each deposit, including the preparation of sufficient appropriate metadata to enable reuse.

Plans may reference the general policy(ies) for the chosen repository(ies) and, if appropriate, include further details related to the specific project. It is the responsibility of the person preparing the data management plan to ensure that the repository policy is appropriate. Where data are not to be managed through an established repository, the data management plan should demonstrate reasonable assurance of the stability and sustainability of the proposed approach.

All reasonable steps should be taken to ensure that research data are not held in any jurisdiction where the available legal safeguards provide lower levels of protection than are available in the UK

7. Data Sharing

Planning for data sharing should begin at the earliest stages of project design and well in advance of beginning the research. Any potential issues which could limit data sharing should be identified and mitigated from the outset. Data management plans should therefore address how the research data will be shared. Any reason for not eventually sharing data should be explained with a justification citing for example legal, ethical, privacy or security considerations.

While a defined limited period of exclusive use of research data may be reasonable in recognition of researchers' intellectual contribution and in order for them to have a first opportunity to exploit the results of their research, data resulting from funding provided by any of the Research Councils should normally be made openly available in a cost effective manner after any such period has elapsed.

The Research Councils recognise that different approaches to data sharing are appropriate in different situations and that data management plans may vary according to the type of data collected. In considering where, how, and to whom research data should be made available, researchers should take account of the potential for further research and wider public benefit. Where possible, research data should be shared using established standards and existing resources. Where the data are likely to be in great demand by others it may be appropriate to request specific resources for a more proactive approach to maximise opportunities for cross linkage with other sectors.

Planning for data sharing should, as far as possible, anticipate and identify which data may be of value to others and should include provision for appropriate data quality assurance. In general, the Research Councils expect that all data relied on to support published research findings will be made freely and publicly available. Other retained research data should be made available wherever it is appropriate

and cost-effective to do so, taking into account the cost of curation and access compared with the cost or feasibility of re-creation, the potential long-term demand for the data and the feasibility of their reuse by others.

Research data may be shared subject to appropriate terms and conditions set out in a data licence and the data owner may require registration to enable tracking of data use and to allow for notification and acceptance of any terms and conditions that may apply to subsequent use of the data.

Data released for sharing should be validated and verified in line with accepted best practice and be of high quality. Any deliberate attempt to compromise that integrity, e.g. by the modification of data or the provision of incorrect metadata, will be considered as a serious breach of research integrity¹⁹.

¹⁹ See RCUK Policy and Guidelines on Governance of Good Research Conduct, Section 3. <http://www.rcuk.ac.uk/Publications/researchers/Pages/grc.aspx>

8. Responsibilities and Resourcing

Researchers and the Research Organisations receiving funding both play a role in ensuring research data are well managed.

Research Organisations should establish appropriate research data management policies and procedures, as well as ensuring that their researchers have access to suitable training and resources, and that they comply with the legal requirements and abide by established policies and procedures.

All researchers should have a clear understanding of their role to ensure that

- research data are of the highest quality and are well documented, so that other researchers can access, understand, use and add value to them independently of the original investigators
- sensitive research data are managed legally and to the highest standards of ethical practice
- where appropriate, data with long-term validity retains value over decades.