

Clarifications of EPSRC expectations on research data management.

Expectation I

Research organisations will promote internal awareness of these principles and expectations and ensure that their researchers and research students have a general awareness of the regulatory environment and of the available exemptions which may be used, should the need arise, to justify the withholding of research data;

Clarification

This is expected to be an on-going part of the normal business of organisation-based training and awareness for staff and students in all research organisations funded by EPSRC. For example, research data management may be addressed as part of the normal induction process for new staff/research students, and awareness of the issues may be subsequently refreshed through regular online and/or facilitated training.

Expectation II

Published research papers should include a short statement describing how and on what terms any supporting research data may be accessed.

Clarification

This expectation is consistent with the RCUK policy on Open Access, and applies to all papers which acknowledge EPSRC funding with a publication date after 1st May 2015. It is also consistent with the Principle that EPSRC-funded research data should be made freely and openly available with as few restrictions as possible in a timely and responsible manner. We acknowledge that not all research papers are supported by research data, and will therefore rely on researchers making informed judgements about when it is appropriate to include such a statement.

The expectation is that data relied on in published research findings will, by default, be available for scrutiny by others. The expectation could be satisfied by citing such data in the published research and including in such citations direct links to the data or to supporting documentation that describes the data in detail, how it may be accessed and any constraints that may apply. It is important that any such links are *persistent* URLs such as DOIs.

If compelling legal or ethical reasons exist to protect access to the data these should be noted in the statement included in the published research paper. A simple direction to interested parties to 'contact the author' would not normally be considered sufficient.

It is not recommended to publish findings that cannot be validated by others. If published research findings rely on data from third parties every effort should be made to ensure that this is also available for scrutiny. It is accepted that ethical or legal constraints may apply – see Expectation VI.

Expectation III

Each research organisation will have specific policies and associated processes to maintain effective internal awareness of their publicly-funded research data holdings and of requests by third parties to access such data; all of their researchers or research students funded by EPSRC will be required to comply with research organisation policies in this area or, in exceptional circumstances, to provide justification of why this is not possible.

Clarification

The expectation that requests to access research data will be logged is pragmatic and applies primarily to data held by the research organisation; it will help build evidence of impact and inform future decisions that may need to be taken about whether to keep specific datasets (data that has not attracted any interest for more than 10 years is not expected to be retained – see discussion of expectation VII below).

EPSRC does not expect research organisations to log access requests to research data originated by their researchers but deposited in an external repository (e.g. subject specific or journal/publisher specific); however, we believe research organisations will find such information a valuable indicator of impact.

Expectation IV

Publicly-funded research data that is not generated in digital format will be stored in a manner to facilitate it being shared in the event of a valid request for access to the data being received (this expectation could be satisfied by implementing a policy to convert and store such data in digital format in a timely manner);

Clarification

We do not expect research organisations to digitise all research data, but we also do not expect requests for access to research data to be rejected on the basis that it is physically too awkward to share. See also Clarification under Expectation IX

Expectation V

Research organisations will ensure that appropriately structured metadata describing the research data they hold is published (normally within 12 months of the data being generated) and made freely accessible on the internet; in each case the metadata must be sufficient to allow others to understand what research data exists, why, when and how it was generated, and how to access it. Where the research data referred to in the metadata is a digital object it is expected that the metadata will include use of a robust digital object identifier (For example as available through the DataCite organisation - <http://datacite.org>).

Clarification

The emphasis in this expectation is on publishing 'metadata for discovery' (rather than publishing the actual data itself) with a further expectation that this metadata will be published 'normally within 12 months of the data being generated'. However, it can be difficult to be precise about the exact 'date of generation', particularly for data recorded over a period of time or which requires significant processing before it can be interpreted. Given the uncertainties that can arise, the following guidance may be helpful and is intended to help research organisations to prioritise their activity and to establish effective data management as part of the normal research process:

1. If the data directly supports research findings published after 1st May 2015 the expectation will be met if the published findings contain appropriate data citations (see clarification under Expectation II). In such cases the cited data/supporting documentation is expected to be accessible online no later than the date of first online publication of the article. This should not be onerous: RCUK Policy on Open Access, which took effect from 1st April 2013, already states that published research papers must include 'if applicable, a statement on how the underlying research materials... ..can be accessed'.
2. In the case of other retained research data with a clear 'date of generation' being on/after 1st May 2015, and where there is no intention to publish findings which rely on that data, metadata for discovery should be exposed online within one year of the 'date of generation'.
3. If retained research data has no clear 'date of generation' and the relevant EPSRC grant ends on/after 1st May 2015, and if there is no intention to publish findings which rely on that data, the end date of the EPSRC grant may be used instead, i.e. the metadata for discovery should be exposed online no later than one year after the end date of the EPSRC grant.

Where an intention to publish is relied on to justify withholding research data attention is drawn to the law on Freedom of Information and the importance of ensuring a publication plan is in place.

The expectation that published metadata 'must be sufficient to allow others to understand what research data exists, why, when and how it was generated, and how to access it' may be satisfied by a combination of information made directly accessible online and more detailed information contained in documentation that can be identified and freely downloaded.

Over time, research organisations are expected to use their best endeavours to make older research data discoverable and accessible in line with our policy framework.

The expectation that research organisations will ensure ‘appropriately structured metadata’ is published does not refer to any specific standard. This is intentional: making research data normally ‘open’ is a relatively new undertaking for all concerned, and robust standards for metadata that facilitates discovery of research data have yet to emerge. EPSRC will continue to monitor developments but does not anticipate being able to mandate use of specific standards in this area for several years. In implementing the policy at this time, we will focus on the desired outcome of ensuring metadata is available and freely accessible rather than assessing how an institution has chosen to achieve that.

Expectation VI

Where access to the data is restricted the published metadata should also give the reason and summarise the conditions which must be satisfied for access to be granted. For example ‘commercially confidential’ data, in which a business organisation has a legitimate interest, might be made available to others subject to a suitable legally enforceable non-disclosure agreement.

Clarification

- Research data in which third parties have a legitimate interest should be managed in a way that does not damage the research process. Personal information should not be put into the public domain without the explicit consent of the person to which it relates. Sensitive information should not be placed in the public domain. Sensitive information is information the release of which would compromise unprotected intellectual property or which, in the judgement of the security services, would result in unacceptable risk to the citizens of the UK or its allies.
- Research organisations and researchers have a responsibility to ensure that publicly funded research involving third parties is planned and executed in such a way that published findings can be scrutinised and if necessary validated by others.
- A delay in the publication of research findings and access to supporting research data is acceptable if necessary to protect intellectual property that would otherwise be compromised.
- Third parties who collaborate in publicly funded research should be made aware of the importance of ensuring that published findings can be validated by others.

Expectation VII

Research organisations will ensure that EPSRC-funded research data is securely preserved for a minimum of 10-years from the date that any researcher 'privileged access' period expires or, if others have accessed the data, from last date on which access to the data was requested by a third party; all reasonable steps will be taken to ensure that publicly-funded data is not held in any jurisdiction where the available legal safeguards provide lower levels of protection than are available in the UK.

Clarification

Data that has not attracted any interest for more than 10 years is not expected to be retained.

It is accepted that 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.

It is reasonable to use data citations, or any other metric based on reliable sources of evidence and widely accepted at the time, to evaluate if interest has been shown in a dataset. It may be reasonable for research organisations to require persons requesting access to specific research datasets to do any of the following before granting access: register; create an account; prove identity; accept terms and conditions of use of the data.

In the event of any transfer of assets (arising for example through merger, acquisition or closure) Research Organisations should ensure that appropriate arrangements are made for the preservation and continued accessibility of retained research data in accordance with these expectations.

Research Organisations are not expected to assume responsibility for the preservation and management of third party research data not generated within their own organisation; however, in the case of important research results relying on such data it may be prudent to take reasonable steps to assure the continued availability and accessibility of the data to allow others to validate the research results.

The wording 'legal safeguards' in the expectation refers to legislation governing access to, or otherwise affecting, the security of information held in digital or electronic form.

Expectation VIII

Research organisations will ensure that effective data curation is provided throughout the full data lifecycle, with 'data curation' and 'data lifecycle' being as defined by the Digital Curation Centre. The full range of responsibilities associated with data curation over the data lifecycle will be clearly allocated within the research organisation, and where research data is subject to restricted access the research organisation will implement and manage appropriate security controls; research organisations will particularly ensure that the quality assurance of their data curation processes is a specifically assigned responsibility.

Clarification

The Digital Curation Centre (www.dcc.ac.uk) provides 'how to' guidance on the scope of services that institutions should be able to provide their researchers, whether by establishing in-house capabilities, collaborative arrangements with other institutions, or through outsourcing to external service providers.

Expectation IX

Research organisations will ensure adequate resources are provided to support the curation of publicly-funded research data; these resources will be allocated from within their existing public funding streams, whether received from Research Councils as direct or indirect support for specific projects or from higher education Funding Councils as block grants.

Clarification

Provided two simple rules are adhered to, all costs associated with research data management are eligible expenditure of research grant funds. The rules are:

1. no expenditure can be 'double funded' (a service that is centrally supported by the indirect costs paid on all research grants cannot then also be included as a direct cost on a grant)
2. all directly incurred expenditure of a grant must be incurred before the end date of the grant. RCUK has already clarified that all elements of research data management are eligible for research grant funding

It is suggested that research offices ensure appropriate provision for research data management is included in a research proposal before it is submitted to EPSRC. In particular:

- a) does a data management plan (DMP) exist? (EPSRC does not require DMPs with research grant applications, but our research data principles include that '...project specific data management policies and plans... ..should exist for all data');
- b) does the project plan reflect the time and resource needed for data management activity, such as preparation of datasets for deposition in an appropriate repository?
- c) if the project is likely to need more resource for research data management than may be provided centrally by the institution, has this been budgeted for in the proposal?

While research organisations have clear responsibilities to ensure, and allocate resource for, effective stewardship of the research data they hold, there is a balance to be struck in terms of meeting the costs of individual sharing requests. Provided the research data is suitably *discoverable* and (subject to any legitimate constraints) *accessible*, Research Organisations are free to choose the extent to which they support the *sharing* of specific research datasets; they may for example consider that enhanced reputational benefits justify greater investment in support of datasets in wide demand. In general, however, while researchers are clearly responsible for the quality of data and supporting documentation made available for sharing, EPSRC considers it reasonable that those requesting access to data be held are responsible for:

- a) determining the suitability for intended use of accessed data; and
- b) ensuring they have the necessary resources to be able to make use of it, if necessary including:
 - re-formatting it to suit their own research needs
 - access to proprietary third party software that may be necessary to process the data
 - travelling to the place where the data is held if it is in physical form and it is impractical to digitise it.