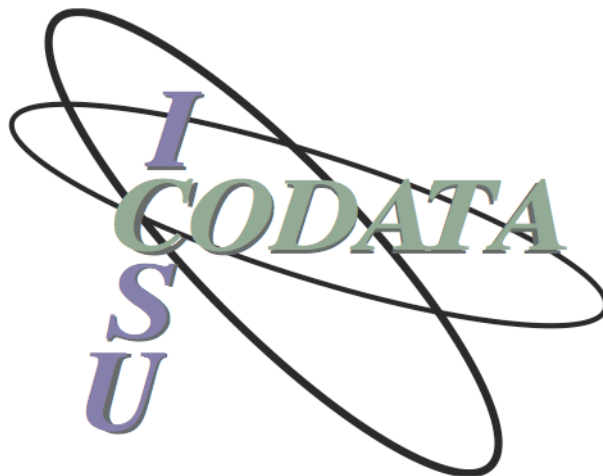


# Current Best Practice for Research Data Management Policies

*A Memo for the Danish e-Infrastructure Cooperation and the Danish  
Digital Library*

Simon Hodson and Laura Molloy

May 2014



## Executive Summary

This memo responds to the request by the Danish e-Infrastructure Cooperation (DeIC), in partnership with Denmark's Electronic Research Library (DEFF), for a document 'to provide an overview of current best practices for research data management (RDM) policies within a number of subject areas, and as such inspire the development of a Danish national strategy on the area of RDM policies.'

In accordance with the instruction, the memo considers current best practices in an international context, highlights what are considered the 'pivotal points' or key elements of the policies mentioned and explores particular variations and distinctive features of those policies relating to particular subject areas (specifically: humanities, social sciences, health sciences, natural sciences and technical sciences).

## Policy Drivers and General Principles

It is important for a survey of good practice to start by considering the drivers and principles underlying data policies which are currently mostly published by funders of research. Institutional research data management policy-making at the current time is mostly in response to these funder policies. The OECD's *Principles and Guidelines for Access to Research Data from Public Funding* (2007) have, since their publication, been of particular influence on research funders across countries and research disciplines. Very recently, there has been great emphasis on the idea of Open Data, the principle of 'open by default' and the criteria for data reuse (particularly as expressed in the concept of 'intelligent openness'). All funder research data policies underline the necessary limits on openness and particularly those relating to personal information and commercial considerations.

Accordingly, good practice in research data policies might be said to start with the following considerations:

1. **An account of the general drivers and principles:** these include the validation of research results, research opportunities for data reuse, the principle of open access by default to the outputs of publicly-funded research, and broader societal and economic benefits.
2. **A discussion of the requirements for effective data sharing:** e.g. 'intelligent openness' and the need for data to be 'discoverable, accessible, assessable, intelligible, useable, and wherever possible interoperable to specific quality standards.'<sup>1</sup>
3. **A statement of the necessary limits of openness:** these are imposed, in particular, by the need to protect personal information, by the requirement to respect commercial considerations and by security concerns.

These general principles outlined, the policy can then move to more specific elements required for implementation.

## Key Policy Elements

We provide a summary of what emerge from our survey as the key elements of current good practice in research data policies. This has the potential to be used as a framework for developing and assessing research data policies. In addition to the principles above, data policies generally contain the following key elements:

4. **A definition of research data:** many policies provide a definition of research data.
5. **An overview of the data within the scope of the policy:** which generally includes two definitions:
  - a. The data that directly underpin or substantiate published research findings (i.e. those that are required for validation). Such data should be made available concurrently with the research publication.

---

<sup>1</sup> G8 Science Ministers Statement, 13 June 2013 <https://www.gov.uk/government/news/g8-science-ministers-statement>

- b. The data assets that are created by the research project, but which may not directly underpin the published research findings. Such data should generally be made available within a specified time after creation or project end.
- 6. **An indication of general criteria for the selection of research data:** it is helpful for policies to indicate which data are likely to be the most important to select for sharing. Policies should also underline the need to make software and tools available.
- 7. **A summary of responsibilities:** in Table 1, we provide an overview of stakeholder responsibilities, as expressed in the research data policies surveyed.

**Table 1: Summary of High-Level Stakeholder Responsibilities**

Stakeholder	Responsibility
Funder	Develop and communicate RDM policy; provide advice directly or through data services; review implementation.
Researcher	Conform to policy in grant proposals, during the lifetime of the project; some responsibilities may remain after the completion of the project.
Research performing organisations	Ensure that the execution of the policy requirements by grant holding lead researchers is adequately supported; do this through institutional policies and provision of support and guidance, particularly for creation and execution of data management plans; depending on national data infrastructure provision the research performing organisation may also need to provide long-term stewardship for some data.
Research data services/centres	Provide long-term stewardship for specific data in accordance with funder policies; provide guidance and support according to role designated by funder.

- 8. **An indication of the availability of infrastructure and responsibility for costs:** the policy should indicate what expectations there are for the provision of research data infrastructure and the associated costs of data management and stewardship. This applies to data management during the lifetime of a project, but is particularly important with regard to the long-term stewardship of the data. International or national data centres serving the research area should be used, if available. Where these are not available, the policy should indicate where responsibility lies. This is generally with the research performing institution. What costs may be covered from the research grant and is it, for example, acceptable for deposit charges to be paid? How is the cost of long-term stewardship of research data in the host institution to be resourced?
- 9. **An overview of data management planning requirements:** it is good practice to require researchers to prepare data management plans (DMPs) and most data policies surveyed make this stipulation. The policy should lay out the procedure around the DMP. Many funder policies require either a brief statement or a more detailed plan to be presented as part of a project proposal. Ideally, the policy should indicate how the DMP will be assessed. It is good practice to provide guidance for those reviewing the DMP.

Some policies require a more detailed DMP to be prepared once the project is underway. Again, the policy should lay out any procedures for reviewing and reporting against the plan.

Above all, the policy should indicate, broadly at least, the issues that should be addressed in a DMP and should link to more detailed guidance and support if available. A number of examples of these guidelines are included in the memo and appendices.

*Inter alia*, the DMP should address who will take responsibility for the long-term stewardship of data created. For cases where this falls to the research performing organisation, the policy may indicate general expectations of how long data should be retained. There is some variation in what the policies surveyed say on this matter and practice is still in the process of evolving.

- 10. **Recommendations on enabling discovery and reuse:** policies generally make a number of specific recommendations or requirements that relate to the objective of enabling discovery and reuse. These typically include the requirement for published research to state how the supporting data may be

accessed as well as recommendations relating to the use of appropriate metadata, permanent identifiers and licenses which enable reuse.

- 11. Stipulations to encourage recognition and reward for data providers:** policies generally acknowledge that moving to an open data regime requires for many research areas a shift in practice and culture. For this reason policies often include, alongside the requirement that publicly funded data should be made open, some statement of the need for appropriate recognition and reward for those researchers who make data open.

There are two notable policy implications of this principle. Firstly, many policies require acknowledgement of data reuse and the citation of data where it underpins further research findings. Secondly, some policies allow periods of privileged access. In relation to the latter, there is considerable diversity. Certain policies, while stressing the time should be limited, uphold it as an important principle. Others reject the principle and require deposit within a limited timescale of project completion or data generation.

- 12. A summary of reporting requirements, compliance monitoring and any possible sanctions:** policies should indicate how compliance will be monitored, what reporting is required and what sanctions may be imposed. The policies surveyed are usually specific on the reporting procedures but do not generally indicate which precise sanctions may be imposed.

## Subject Area Considerations

The third part of this memo provides an overview with examples of variations in research data policies in different subject areas. There are some limitations that should be mentioned here.

Many variations between the policies of funders in different research areas appear to be contingent and procedural rather than based on specific and intrinsic issues to do with data sharing in the given research area.

Many research data policies present broad, high-level principles. Nevertheless, some funders in a given research area highlight particular issues which are characteristic for that discipline. However, it is often only at the level of accompanying or more detailed guidance and support that a lot of such issues are really addressed, rather than in the research data policy itself.

With policy development and - above all - data practice still hardly mature, we would hesitate to draw particularly strong conclusions about subject-specific variations from the range of examples surveyed. Nevertheless, the most significant variations between the policy concerns in given subject areas relate to:

**Legal and ethical requirements specific to the type of research being conducted.** For example, privacy issues are particularly important in the health and social sciences.

**Existence of more established data infrastructure and practice.** For research disciplines where international or national data centres have been established, policies more often provide (in appendices or guidance) lists of appropriate data centres or databases and allude to any existing technical standards or practices widely used in that discipline.

**Accepted technical approaches in a specific research area.** In some areas of the social sciences, life sciences and natural or technical sciences, specific data format or metadata standards have emerged and become common practice in that community. Where this is the case, these standards may be mentioned directly or indirectly in policy documents and recommendations.