

2011

How to Develop a Data Management and Sharing Plan

Sarah Jones
Digital Curation Centre

Follow this and additional works at: <https://arrow.tudublin.ie/dataguide>

Recommended Citation

Jones, S. (2011). 'How to Develop a Data Management and Sharing Plan'. DCC How-to Guides. Edinburgh: Digital Curation Centre. Available online: <http://www.dcc.ac.uk/resources/how-guides>

This Article is brought to you for free and open access by the Data@TU Dublin at ARROW@TU Dublin. It has been accepted for inclusion in Guides by an authorized administrator of ARROW@TU Dublin. For more information, please contact arrow.admin@tudublin.ie, aisling.coyne@tudublin.ie.



This work is licensed under a [Creative Commons Attribution-Noncommercial-Share Alike 4.0 License](https://creativecommons.org/licenses/by-nc-sa/4.0/)

A Digital Curation Centre
'working level' guide



How to Develop a Data Management and Sharing Plan

Sarah Jones (DCC)



Digital Curation Centre, 2011.
Licensed under Creative Commons Attribution 2.5 Scotland:
<http://creativecommons.org/licenses/by/2.5/scotland/>

How to Develop a Data Management and Sharing Plan

Introduction

This guide outlines the process of developing a data management and sharing plan. Planning for the effective creation, management and sharing of your data enables you to get the most out of your research. The guide outlines UK funder expectations for data management and sharing plans and provides practical guidance on how to meet these. It should be of relevance to researchers and those supporting them within higher education.

Why develop a data plan?

There are many benefits to managing and sharing your data:

- you can find and understand your data when you need to use it
- there is continuity if project staff leave or new researchers join
- you can avoid unnecessary duplication e.g. re-collecting or re-working data
- the data underlying publications are maintained, allowing for validation of results
- data sharing leads to more collaboration and advances research
- your research is more visible and has greater impact
- other researchers can cite your data so you gain credit

Planning helps you to achieve these benefits; it is ultimately most useful to you¹. Making a plan helps you to save time and effort and makes the research process easier. By considering what data will be created and how, you can check you have the necessary support in place. Planning also enables you to make sound decisions, bearing in mind the wider context and consequences of different options.

Publishers and research funders may require that you share your data so it is worth investing time to plan for effective data management. Several funders ask for data plans as part of grant proposals. The DCC views plans submitted in grant proposals as preliminary outlines, which should then be developed into more coherent processes and procedures at the outset of your research. For the purposes of this guide we will focus on the application stage requirements. A further guide - *How to put the data management and sharing plan into practice* - will address data management during the research process.

What do research funders want?

Many UK funders have released data policies which advocate curation and data sharing². Several of these require that data management and sharing plans are submitted as part of grant applications³. Funders expect data plans to outline how data will be created, managed, shared and preserved, justifying any restrictions that need to be applied. The plans are an opportunity to demonstrate your awareness of good practice and reassure funders that your proposal is in line with their data policy.

The DCC has collated UK funders' guidelines on what to cover in data management and sharing plans⁴. Funders typically propose broad themes or questions for you to consider as appropriate to your proposal, though the Arts and Humanities Research Council (AHRC) and Economic and Social Research Council (ESRC) ask set questions.

Strict word counts or page limits may be imposed, so you need to be clear and concise. Funders typically expect a succinct summary submitted as part of the 'Case for Support' or in an allocated section of the application form. Avoid repeating content from elsewhere in your application or using this space to provide details unrelated to data management and sharing.

¹ Murray-Rust, Peter. (2011, August). *Why YOU need a data management plan*. From <http://blogs.ch.cam.ac.uk/pmr/2011/08/01/why-you-need-a-data-management-plan/> and *Data repositories for long-tail science: setting the scene* from <http://blogs.ch.cam.ac.uk/pmr/2011/08/15/data-repositories-for-long-tail-science-setting-the-scene/> Both retrieved 22 August 2011.

² DCC Funders' data policies, URL: <http://www.dcc.ac.uk/resources/policy-and-legal/funders-data-policies>

³ DCC, Funders' data plan requirements, URL: <http://www.dcc.ac.uk/resources/data-management-plans/funders-requirements>

⁴ Jones, Sarah. (2011). *Summary of UK research funders' expectations for the content of data management and sharing plans v4.0*. Retrieved 8 August 2011, from <http://www.dcc.ac.uk/resources/data-management-plans/funders-requirements>

The DCC also provides DMP Online, a web-based tool to help researchers create data management and sharing plans according to the requirements of major UK funders⁵. The structure of DMP Online is based around the DCC Checklist for a Data Management Plan, which defines elements to include. The main sections are used as headings to structure the guidance in Section 5 of this document. They are:

1. Data Types, Formats, Standards and Capture Methods
2. Ethics and Intellectual Property
3. Access, Data Sharing and Reuse
4. Short-Term Storage and Data Management
5. Deposit and Long-Term Preservation
6. Resourcing

Advice on how to plan

Consult and collaborate: Think through the different options and seek advice to determine what is best for your context. It is particularly useful to consult on technical aspects as these affect how your project is scheduled, the expertise and applications required, and the methods used to acquire and analyse data. Ask for advice from your colleagues, the library, local IT support, legal advisors, ethics boards, data repositories and more. If needed, also build in specialist support.

Use existing support: Others have addressed these challenges before you, so build on existing models. Data management infrastructure is increasing, particularly within institutions⁶. IT services run established back-up regimes and your research group or department may have local policies and procedures that you can refer to. Support is also available through a variety of disciplinary data centres, repositories and structured databases⁷.

Justify your decisions: Funders tend not to specify particular file formats, standards or methodologies that you are expected to use. You need to choose and demonstrate that the selections made are the most appropriate for your context, that of your discipline and future users. Similarly, you need to present a convincing case for any restrictions on data sharing.

Be prepared to implement your plan: Funders want to see that you understand their requirements and have realistic plans in place to meet these. The description of planned work should be clear and achievable so markers can feel confident that you understand the options and will be able to deliver what is proposed. Clearly defined roles and responsibilities add weight to your plans so be explicit about who will do what, how and when.

The content of a data management and sharing plan

The guidance in this section is structured around the six core themes of the DCC Checklist for a Data Management Plan⁸. Next to each of the subsections you will find a summary of the type of questions asked by UK research funders and some pointers and tips on how to answer.

Data Types, Formats, Standards and Capture Methods

Outline and justify your choices: You should detail what data you will create and explain why you have opted for particular formats, standards and methodologies. Bear in mind that the choices you make may make it easier or harder to share and preserve your data.

It can be useful to capture your data in (or convert it to) community-accepted **data formats**. Using standard or widely-adopted formats will make your data interoperable. Open or non-proprietary formats are preferable, as you and others will have less trouble processing these later. If your data are to be deposited into an archive, particular formats may be preferred⁹.

⁵DCC, DMP Online, URL: <http://dmponline.dcc.ac.uk/>

⁶Whyte, Angus and Tedds, Jonathan (2011). 'Making the Case for Research Data Management'. *DCC Briefing Papers*. Edinburgh: Digital Curation Centre, URL: <http://www.dcc.ac.uk/resources/briefing-papers>

⁷DataCite, Repositories list, URL: <http://datacite.org/repolist>

⁸Donnelly, Martin & Jones, Sarah. (2011) *DCC Checklist for a Data Management Plan* v3.0. Retrieved 8 August 2011, from http://www.dcc.ac.uk/webfm_send/431

Funder questions

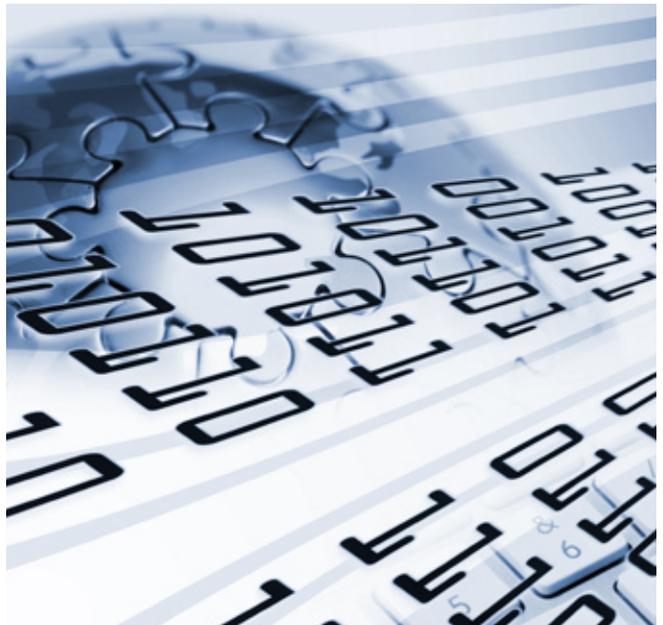
- What data outputs will your research generate?
 - outline volume, type, content, quality and format of the final dataset
- Outline the metadata, documentation or other supporting material that should accompany the data for it to be interpreted correctly
- What standards and methodologies will be utilised for data collection and management?
- State the relationship to other data available in public repositories e.g.
 - existing data sources that will be used by the research project
 - gaps between available data and that required for the research
 - the added value that the new data would provide in relation to existing data

⁹For example see guidance from the UK Data Archive. File formats table, URL: <http://www.data-archive.ac.uk/create-manage/format/formats-table>

Documentation and metadata allow your data to be understood and discovered by others. It is fundamental to capture contextual details about how and why the data were created. Metadata is a subset of this broad documentation, describing the data in detail. There are various metadata standards which can help you to describe your data in a consistent way. Librarians, data repositories or your colleagues may be able to advise on relevant standards.

Make informed decision based on review:

It can help to show your awareness of good practice or that you have sought advice to develop your plans. Some funders also expect you to demonstrate that existing data are not sufficient for your needs, so you may need to show that you have reviewed repository and data centre holdings or consulted with similar projects.



Ethics and Intellectual Property

Present a strong case for any restrictions on sharing:

Explain any constraints, such as embargo periods or restricted access, and ensure these are properly justified as there is a common expectation that publicly funded research data will be openly available as soon as possible. These justifications may also be of use in the event of a Freedom of Information request for your research data¹⁰.

All research involving human data or material is subject to formal **ethical review**. Where appropriate, you should outline the steps you will take to protect research participants, e.g. anonymising data. It helps to show that you've balanced concerns with the desire to share e.g. by negotiating informed consent for data sharing. Many University Ethics Committees provide sample consent forms and services such as the UK Data Archive provide excellent guidance in this area¹¹. You should also demonstrate awareness of relevant legislation such as the Data Protection Act.

Data ownership should be clarified and, where necessary, plans should be in place to negotiate licences at the start of the research process. If you agree/purchase licences to reuse third party data, be aware of any restrictions this places on subsequent deposit and data sharing. JISC Legal¹² provides lots of advice on copyright, IPR and relevant legislation such as the Data Protection Act and Freedom of Information. Institutional support is also available from experts in university libraries, records management and research offices.

Funder questions

- Demonstrate that you have sought advice on and addressed all copyright and rights management issues that apply to the resource
- Make explicit mention of consent, confidentiality, anonymisation and other ethical considerations, where appropriate
- Are any restrictions on data sharing required – for example to safeguard research participants or to gain appropriate intellectual property protection?

¹⁰Rusbridge, Chris & Charlesworth, Andrew. (2010). *FOI and Research Data: Researchers' Questions and Answers*. Retrieved 24 August 2011, from: <http://foiresearchdata.jiscpress.org/>

¹¹UKDA, Consent and ethics, URL: <http://www.data-archive.ac.uk/create-manage/consent-ethics>

¹²JISC Legal, URL: <http://www.jisclegal.ac.uk/Home.aspx>

Access, Data Sharing and Reuse

Anticipate and plan for data reuse: It can help to envisage which users your data would be of value to, and address their needs when deciding how to make the data available. Data centres may also ask you to meet minimum quality standards to make sure your data can be understood and reused by other researchers.

Provide specific details on access: Reassure funders by being very clear about where, when and how your data will be made available. The DCC offers guidance on how to licence your data to make clear who can use it and for what purpose¹³. Funders often state expected timeframes for release, such as making data available on publication. If you can't meet these expectations or need to impose any restrictions, try to demonstrate that you have considered various means of overcoming these challenges.

Use existing infrastructure: Where possible select an appropriate disciplinary database, data centre or institutional repository. If you are unsure which services are available to you, check the repository list collated by DataCite, BioMed Central and the DCC¹⁴. If access to your data needs to be restricted, look for secure data services or data enclaves.

Short-Term Storage and Data Management

Define data management support: Outline what provision is available to you within your institution and any additional skills or resources that you need to secure. If local support is available, it helps to demonstrate that you have discussed and agreed requirements. If you need to secure external support, justify the selections made and budget requested. Be clear about who will be responsible for different tasks.

Consider the practicalities: Are the investigators co-located, or will you need infrastructure that accommodates secure remote access? How will data quality be monitored if you are working in a distributed network across several sites? Strong file-naming conventions and versioning applications may be of use to keep track of the development process, particularly when several people are working together.

Funder questions

- What are the further intended and/or foreseeable research uses for the completed dataset(s)?
- How you will make the resource accessible to the potential audience(s) identified.
 - Where will you make the data available?
 - How will other researchers be able to access the data?
 - Will a data sharing agreement be required?
 - What is the timescale for public release of the data?
- State any expected difficulties in data sharing, along with causes and possible measures to overcome these difficulties.
- How will data sharing provide opportunities for coordination or collaboration?

Funder questions

- Describe the planned quality assurance and back-up procedures [security/storage]
- Specify the responsibilities for data management and curation within research teams at all participating institutions

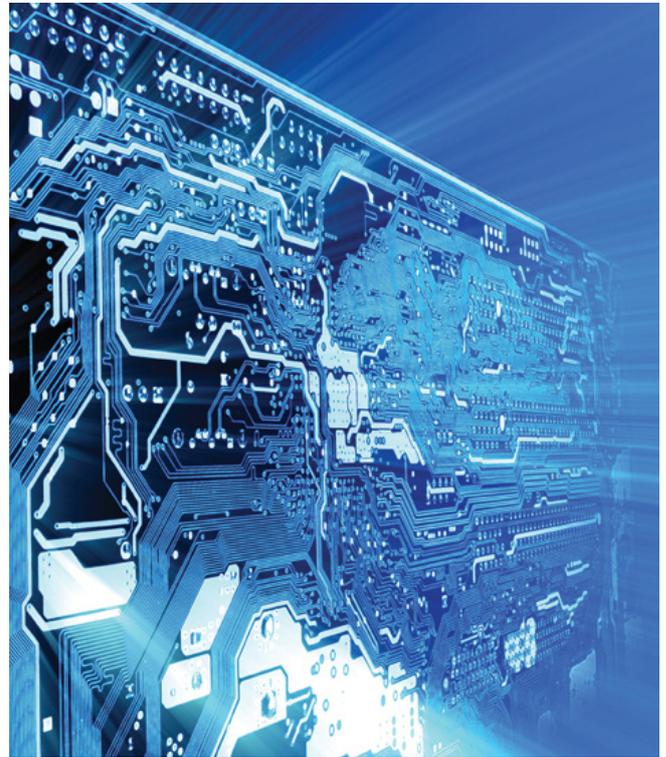
¹³Ball, Alex. (2011). *How to Licence Research Data*. Retrieved 22 August 2011, from: <http://www.dcc.ac.uk/resources/how-guides/license-research-data>

¹⁴DataCite, Repositories list, URL: <http://www.datacite.org/repolist>

Apply appropriate levels of data management:

Funders want to be reassured that the day-to-day data management is fit for purpose. You may apply differing levels of service or adopt a combination of approaches:

- Security may be more robust for any sensitive data you collect than for secondary data you hold under licence. Think about how you will transfer data securely e.g. encrypting data or using secure online storage. If using online services, you should know where your data are hosted and be certain that this is legally permissible.
- Back-up of your unique data is more critical than copies of secondary data. The more important the data and the more often it is used, the more regularly it needs to be backed up. Fully managed file services with automated back-up, such as those offered by university IT services, are very robust and save you the time and effort of implementing your own system. Such services could be used in combination with portable storage or cloud computing to meet particular needs.



Deposit and Long-Term Preservation

Select data of long-term value: Data sharing and preservation may not be applicable in every case. The DCC provides a 'How to ...' guide on appraisal, which offers practical strategies to help you select important data¹⁵. Deciding what has long-term value and preparing those data to expected standards for deposit are time-consuming processes, for which you should allocate significant resources.

Safeguard the data behind the graph: It is a common expectation among RCUK funders that published results will include information on how to access the supporting data¹⁶. Even if there is no obvious home for the majority of your data, the data which underpin publications should be extracted, captured in machine-readable form and deposited somewhere so they remain accessible.

Assure that your data will remain accessible: Whatever approach you adopt, focus on making a convincing case that your data will remain accessible. If you plan to deposit in a data centre, it helps to speak with their staff early on as they can advise what is appropriate and feasible in terms of preservation. Universities are increasingly providing infrastructure to support data management and there are some disciplinary services which may be of use.

Funder questions

- Identify which of the data sets produced are considered to be of long-term value
- Outline the plans for preparing and documenting data for preservation and sharing
- Explain your archiving/preservation plan to ensure the long-term value of key datasets

¹⁵Whyte, Angus & Wilson, Andrew. (2010). *How to Appraise and Select Research Data for Curation*. Retrieved 8 August 2011, from: <http://www.dcc.ac.uk/resources/how-guides/appraise-select-research-data>

¹⁶RCUK. (2011). *Common Principles on Data Policy* Retrieved 19 August 2011, from: <http://www.rcuk.ac.uk/research/Pages/DataPolicy.aspx>

Resourcing

Outline and justify costs: If you need to purchase storage, outsource services such as back-up and preservation, or plan to pay for data management support, these costs should be outlined and justified in your proposal. Where institutional provision is available, show that the support you require has been discussed and agreed. It also helps to link resources with roles and responsibilities to demonstrate how the plan will be implemented.

Don't underestimate the human effort required:

Creating documentation and making your data understandable to others is very time consuming, so be realistic about how much effort is needed to prepare your data for sharing and preservation. The UKDA offers a toolkit to help researchers cost activities related to managing and sharing social science data¹⁷.

Show efficient use of public funds: The RCUK *Common Principles on Data Policy* state that it is appropriate to use public funds to support

Funder questions

- What resources will you require to deliver your plan?
- Outline additional hardware, software and technical expertise, support and training that is likely to be required and how it will be acquired

the management and sharing of publicly-funded research data, but this is expected to be efficient and cost-effective¹⁸. A summary of individual funder's views on meeting associated costs is available via the DCC policy pages¹⁹.

Practical steps to get started

Examples

ICPSR's *Framework for Creating a Data Management Plan* is of particular use. It proposes elements to include, advises why each is important, and gives a wealth of example texts.

Some funders offer detailed guidance to help you develop your plan. The Wellcome Trust provides an FAQ on developing a plan²⁰, and the BBSRC Data Sharing Policy offers detailed guidance notes and illustrative examples²¹.

There are example data plans online that can give you a sense of what to write:

- The UK cross-council Rural Economy and Land Use (RELU) programme provides example data management plans²²
- Detailed data management plans provided by NERC data centres for its thematic programmes are online²³
- Psychology specific guidance and example texts for completing DMP Online have emerged from the JISC DMTPsych project²⁴
- ICPSR, a social science data archive in the USA, has collated example plans from the natural sciences²⁵
- The University of California San Diego provides plans submitted by its researchers as part of National Science Foundation grant proposals²⁶
- Yale University provides examples of data management plans²⁷

¹⁷UKDA. (2011). *Activity-based data management costing tool for researchers*. Retrieved 8 August 2011, from: http://www.data-archive.ac.uk/media/257647/ukda_jiscdmcosting.pdf

¹⁸RCUK. (2011). *Common Principles on Data Policy* Retrieved 19 August 2011, from: <http://www.rcuk.ac.uk/research/Pages/DataPolicy.aspx>

¹⁹DCC, Overview of Funders' Data Policies, URL: <http://www.dcc.ac.uk/resources/policy-and-legal/overview-funders-data-policies>

²⁰Wellcome Trust. (n.d.). Guidance for Researchers: *Developing a Data Management and Sharing Plan*. Retrieved 8 August 2011, from <http://www.wellcome.ac.uk/About-us/Policy/Spotlight-issues/Data-sharing/Guidance-for-researchers/index.htm>

²¹BBSRC. (2010). *Data Sharing Policy v1.1*. Retrieved 25 August 2011, from <http://www.bbsrc.ac.uk/organisation/policies/position/policy/data-sharing-policy.aspx>

²²RELU, Example Data Management Plans, URL: <http://relu.data-archive.ac.uk/DMPexample.asp>

²³See for example RAPID Climate Change programme Data Management Plan, URL: http://www.noc.soton.ac.uk/rapid/sci/documents/rapid_data_plan.pdf and Micro to Macro µ2M Data Management Plan, URL: <http://www.bgs.ac.uk/micromacro/docusearch.cfm?dtype=d>

²⁴Guidance Notes for Completing a Data Management Plan, URL: <http://www.dmtpsych.york.ac.uk/s.php?p=12>

²⁵ICPSR, Data Management Plan Resources and Examples, URL: <http://www.icpsr.umich.edu/icpsrweb/ICPSR/dmp/resources.jsp>

²⁶UCSD, Example Data Management Plans, URL: <http://rci.ucsd.edu/dmp/examples.html>

²⁷Yale University, Data Management Plan Examples, URL <http://odai.research.yale.edu/data-management-plan-examples>

Support and advice

The DCC provides a number of resources to help you understand funders' requirements and develop data management and sharing plans²⁸. Guidance is available on our website and we run a helpdesk service to provide tailored advice²⁹. Of particular interest is DMP Online, the DCC's web-based tool to help you develop a plan customised to your funder's expectations³⁰.

There are lots of people within your institution and subject community who can offer advice and may be able to help develop your plan. Ask for advice from your colleagues, the library, local IT support, legal advisors, ethics boards, data repositories and more. It is useful to speak with them early on so you can build their contributions into your proposal.

Remember:

Data plans are an integral part of grant applications – not an afterthought

Reviewers will look for evidence that data management is embedded throughout your proposal and forms an integral part of your research process. Include high-level data management and sharing details in the Case for Support and briefly explain associated costs in the Justification for Resources.

Data plans are enhanced through collaboration

Few people have all of the skills required to manage and share data throughout its lifecycle, so seek input from others with relevant expertise and use tools provided by your community. Don't go it alone – getting support will strengthen your proposal.

Data plans are living documents – they will change

The plan you develop for the grant application is just an initial idea. Once funded, you'll need to expand this outline by developing policies and procedures or implementing guidelines from your research group, department or institution. Processes often evolve over time to respond to new opportunities or changes in the research. Your data plan can help to document this.

Acknowledgements

Thank you to all for the many helpful comments, particularly James Wilson (University of Oxford), Philip Lord (The Digital Archiving Consultancy), Tim Banks (University of Leeds), Laurence Horton (GESIS), John Milner (JISC), Catherine Jones (STFC e-Science), Graeme Cannon & Laura Molloy (University of Glasgow), Louise Corti, Veerle Van den Eynden & Libby Bishop (UKDA), Robin Rice (EDINA), Chris Rusbridge (consultant), Ben Ryan (EPSRC) and Kate McNeill (MIT Libraries).

²⁸DCC, Data Management Plans resources, URL: www.dcc.ac.uk/resources/data-management-plans

²⁹DCC helpdesk, URL: <http://www.dcc.ac.uk/contact-us/help-desk>

³⁰DCC, DMP Online, URL: <http://dmponline.dcc.ac.uk/>

Further Information and Bibliography

The following tools and resources are provided by the DCC:

Data Management Plan resources:

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

Summary of UK funders' expectations for DMPs:

http://www.dcc.ac.uk/webfm_send/358

DCC Checklist for a Data Management Plan:

http://www.dcc.ac.uk/webfm_send/431

DMP Online: <http://dmponline.dcc.ac.uk/>

BBSRC. (2010). *Data Sharing Policy* v1.1. Retrieved 25 August 2011, from <http://www.bbsrc.ac.uk/organisation/policies/position/policy/data-sharing-policy.aspx>

Brunt, James. (2011). *How to Write a Data Management Plan for a National Science Foundation (NSF) Proposal*. Retrieved 8 August 2011, from: <http://intranet2.1tinternet.edu/node/3248>

ICPSR. (n.d.). *Framework for Creating a Data Management Plan*. Retrieved 8 August 2011, from: <http://www.icpsr.umich.edu/icpsrweb/content/ICPSR/dmp/framework.html>

James, Hamish. (2008) *AHDS Notes on Writing the AHRC Technical Appendix*. Retrieved 8 August 2011, from: <http://www.ahds.ac.uk/creating/information-papers/writing-appendix/index.htm>

JISC Legal (2011) *Copyright and Intellectual Property Law*. Retrieved 8 August 2011, from: <http://www.jisclegal.ac.uk/LegalAreas/CopyrightIPR.aspx>

Murray-Rust, Peter. (2011). *Data repositories for long-tail science: setting the scene*. Retrieved 22 August 2011, from: <http://blogs.ch.cam.ac.uk/pmr/2011/08/15/data-repositories-for-long-tail-science-setting-the-scene/>

RCUK. (2011). *Common Principles on Data Policy*. Retrieved 19 August 2011, from: <http://www.rcuk.ac.uk/research/Pages/DataPolicy.aspx>

Rusbridge, Chris & Charlesworth, Andrew. (2010). *FOI and Research Data: Researchers' questions and answers*. Retrieved 24 August 2011, from: <http://foiresearchdata.jiscpress.org/>

UKDA. (2011). *Managing and Sharing Data: best practice for researchers*. Retrieved 8 August 2011, from: <http://www.data-archive.ac.uk/media/2894/managingsharing.pdf>

Wellcome Trust. (n.d.). *Guidance for Researchers: Developing a Data Management and Sharing Plan*. Retrieved 8 August 2011, from <http://www.wellcome.ac.uk/About-us/Policy/Spotlight-issues/Data-sharing/Guidance-for-researchers/index.htm>

Please cite as: Jones, S. (2011). 'How to Develop a Data Management and Sharing Plan'. *DCC How-to Guides*. Edinburgh: Digital Curation Centre. Available online: <http://www.dcc.ac.uk/resources/how-guides>