

An Introduction to Managing Research Data

FOR RESEARCHERS AND STUDENTS



Author	Andrew Burnham
Date	15 July 2013
Version	1-0
URL	http://www.le.ac.uk/researchdata
File	UoL_DMP_IntroGuide_v1-0
Reviewed by	Research Computing Management Group (including College representatives, IT Services, Library, Research Support Office and Academic Practice Unit)
IPR	This work is a modified version of 'An Introduction to Managing Research Data' (data.bris-Introduction-to-managing-research-data-v21.docx) made with the written permission of the copyright holder (© University of Bristol 2013).
Classification	May be viewed by anyone, anywhere in the world
Acknowledgement	<i>This document is an adaptation of a document produced by the data.bris team at the University of Bristol, whose permission to adapt the document is gratefully acknowledged.</i>

Document History

Version	Date	Author	Detail/Reason for Change
Draft 2 \UoL_DMP_IntroGuide_v0-1.docx	04.06.2013	A. Burnham/University of Bristol	First draft.
Draft 2 \UoL_DMP_IntroGuide_v0-2.docx	10.06.2013	A. Burnham	Second draft for review by RCMG.
Draft 3 \UoL_DMP_IntroGuide_v0-3.docx	10.06.2013	A. Burnham	Third draft for review by RCMG. Approved for release by PVC Research
Draft 4 \UoL_DMP_IntroGuide_v0-4.docx	08.07.2013	A. Burnham	Final draft with minor edits.
First published version \UoL_DMP_IntroGuide_v1-0.docx (.pdf)	15.07.2013	A. Burnham	First published version.



Contents

1. Introduction
2. What is Research Data Management?
3. What is research data?
4. Why this affects you
5. Why manage research data?
6. Drivers for managing research data
7. Benefits of good data management
8. Where to get help and information

Managing research data - an Introduction to practical issues

9. Data Management Planning
10. Ensuring the quality of data
11. Research data format
12. Describing research data - Metadata
13. Data storage, selection, preservation and archive
14. Sharing research and providing long term access to data
15. Ethical considerations
16. Personal information
17. Copyright and IRP
18. Re-using existing data
19. Citing research data in research outputs
20. Roles and responsibilities
21. The cost of managing research data

What would you do if you lost your research data tomorrow?

Research Data Management isn't principally about complying with policy - at heart it means helping you to complete your research, share the results, and allow you to get credit for what you have done.¹

The Research Councils UK (RCUK) Common Principles on Data Policy² states that “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 that does not harm intellectual property” and within its statement on Expectations for Societal and Economic Impact³, there is a clear directive that those who receive funding are expected to “take responsibility for the curation, management and exploitation of data for future use”.

This guide is intended as an introduction to managing research data for all researchers, and students, whether or not they receive research body funding. This document is one in a series of managing research data guides which provide detailed guidance relating to specific research funders.

1. Introduction

Managing data is an essential area of responsible research conduct for staff and students. As members of the University, all researchers have individual responsibility for appropriately managing the research data that they create. This document introduces you to the concept of research data and research data management, explains what constitutes research data and how it differs from other types of information. The document is particularly appropriate for postgraduate students and early career researchers who would like to learn more about managing their research data prior to submitting a funding application.

Many funding bodies now require that recipients of funding create plans at the bidding stage for managing their research data, storing and preserving it in the long term and sharing some, or all of that data once the research is completed.

Academic publishers too, are increasingly calling for scientific claims to be underpinned by publically accessible data which can be checked by anyone.

¹ Professor Kevin Schurer, Pro-Vice Chancellor (Research and Enterprise), <http://www2.le.ac.uk/services/research-data>

² RCUK Common Principles on Data Policy, Research Councils UK (RCUK), <http://www.rcuk.ac.uk/research/Pages/DataPolicy.aspx>

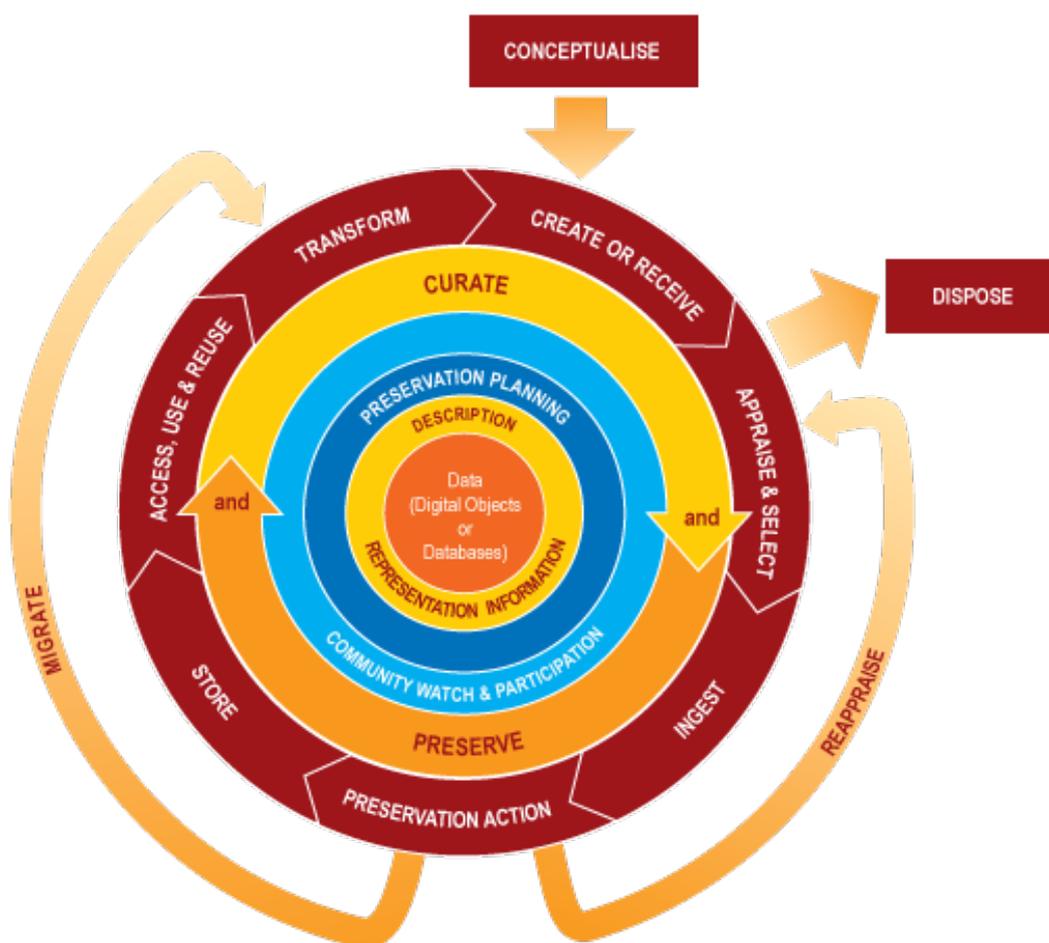
³ RCUK expectations for societal and economic impact. Research Councils UK (RCUK), www.rcuk.ac.uk/documents/innovation/expectationssei.pdf

2. What is Research Data Management?

"Research data management concerns the organisation of data, from its entry to the research cycle through to the dissemination and archiving of valuable results. It aims to ensure reliable verification of results, and permits new and innovative research built on existing information."⁴

Research Data Management is part of the research process, aims to make the research process as efficient as possible, and meet expectations and requirements of the university, research funders, and legislation⁵.

Key elements of the DCC's Digital Curation Lifecycle Model⁶



⁴ Whyte, A., Tedds, J. (2011). 'Making the Case for Research Data Management'. DCC Briefing Papers. Edinburgh: Digital Curation Centre, <http://www.dcc.ac.uk/resources/briefing-papers/making-case-rdm>

⁵ "What is Research Data Management?", <http://www2.le.ac.uk/services/research-data/rdm/what-is-rdm>

⁶ Digital Curation Centre, <http://www.dcc.ac.uk/resources/curation-lifecycle-model>

In simple terms research data management concerns how you:

- **Create** data and plan for its use
- **Organise**, structure, and name data
- **Keep** it, make it secure, provide access, store and back it up
- **Find** information resources, and **share** with collaborators and more broadly, publish and get cited

At the University resources specifically in support of Research Data Management (RDM) are signified by this graphic:



3. What is research data?

Research data is information that is created, manipulated or cited in the course of funded or unfunded research.

Newly created research data is often arranged or formatted in such a way as to make it suitable for communication, interpretation and processing, often by a computer. Advances in technology are transforming research. Although the growth of data has been most apparent in large scale research projects, small projects will also generate important research data.

Defining 'research data' in detail is often viewed as challenging however⁷,

- There is not a consensus
- It varies according to discipline
- It varies according to the research funder

Other definitions for example include, "*Research data, unlike other types of information, is collected, observed, or created, for purposes of analysis to produce original research results.*"⁸,

⁷ A range of definitions can be found here: "Research Data – Definitions", http://www2.le.ac.uk/services/research-data/documents/UoL_ResearchDataDefinitions_20120904.pdf

⁸ University of Edinburgh, <http://www.ed.ac.uk/schools-departments/information-services/services/research-support/data-library/research-data-mgmt>

and from a research funding body:

"Research data is defined as recorded factual material commonly retained by and accepted in the scientific community as necessary to validate research findings; although the majority of such data is created in digital format, all research data is included irrespective of the format in which it is created." Engineering and Physical Sciences Research Council (EPSRC)⁹

For the purposes of this document research data shall be regarded as that which is created in a digital form (born digital), converted to a digital form (digitised) or significantly altered within the digital realm, during the course of research activities¹⁰. Research data does not typically include data generated in the course of personal activities, desktop or mailbox backups, or data produced by non-research activities such as University administration or teaching. The same information may be research data one point in time, but not at another time, depending on whether information is being used for research purposes.

Example: a scanned photographic image of the University Fielding Johnson building in a historical archive is an archived image in an image bank. When used by a researcher to study the history of mental health in Leicestershire, the photographic image becomes research data.

Research data comes in an endless variety of formats and may include any of the following:

- Documents (text, MS Word), spread sheets
- Scanned laboratory notebooks, field notebooks, diaries
- Online questionnaires, transcripts or surveys
- Digital audio or video recordings
- Transcribed test responses
- Database contents
- Digital models, algorithms or scripts
- Contents of an application (input, output, log files for analysis software, simulation software, schemas)
- Documented methodologies and workflows
- Records of standard operating procedures and protocols

The following *research records* may also need to be managed during and after the life of a project but are not generally considered to be research data:

- Correspondence (electronic mail and paper correspondence)
- Grant applications
- Ethics applications
- Research progress reports
- Research publications
- Master lists
- Internal social media communications such as blogs, wikis etc.

⁹ EPSRC, <http://www.epsrc.ac.uk/about/standards/researchdata/Pages/scope.aspx>

¹⁰ Note that the EPSRC Policy Framework on Research Data (<http://www.epsrc.ac.uk/about/standards/researchdata/Pages/expectations.aspx>) includes an expectation that "research data that is not generated in digital format will be stored in a manner to facilitate it being shared".

4. Why this affects you

With a range of drivers and benefits there are many reasons to pay attention to RDM – it (and this document) is likely to have relevance to all researchers and students, whether they wish to, or are currently carrying research which is externally funded or not, by Research Councils UK (RCUK)¹¹ or non-RCUK bodies.

Researchers and students can personally benefit from good practice in RDM whether it be helping to navigate through required processes, protecting their intellectual property, being able to locate and accurately distinguish between files/datasets, keep them secure and share them with collaborators, or improve the opportunities to collaborate, be published and cited, and to be given the opportunity to carry out more research.

Equally pressures are growing on researchers and institutions with greater oversight of the research process and demands for evidence of research integrity, the principle of data as a public good being a driver. Legislative and regulatory demands in the area of both disclosure (Freedom of information) and confidentiality (the Data Protection Act) present significant demands, as do the range of funding body data policies.

5. Why manage research data?

Research data management concerns the organisation and curation of data, from its entry into the research cycle through to the dissemination and archiving of valuable results and the checking of those results by third parties. It also includes activities that ensure research data is ‘fit for purpose’. In the course of your research you are likely to create a significant amount of data and this can quickly become disorganised, out-of-date, meaningless, or lost.

By actively managing your research data in an appropriate way you will ensure that funding and regulatory requirements are met; transparency and accountability are maintained, data remains accurate, reliable and complete; research data keeps its integrity and research results may be replicated; duplication of effort is kept to a minimum; data security is enhanced and the risk of data loss is minimised. Above all good research data management will enhance the long-term value of your research, allow it to be better shared amongst a wider research community, increasing its visibility and impact.

Your approach to managing research data will be influenced by many issues:

- The risk of data loss
- Non-repeatability of research e.g. weather observational measurements
- The need to repeat work if you can't make sense of it if it is not documented effectively
- Institutional reputational risk – can you demonstrate research verification/validation/integrity?
- Inclusion of sensitive/confidential data and the demands of the Data Protection Act

¹¹ RCUK bodies are the AHRC, BBSRC, EPSRC, ESRC, MRC, NERC and STFC.

- 'Big data' – storage and access demands, enabling re-use
- Giving access to data and/or results to other researchers/the public
- Routine good practice – the ability to describe, share, cite, re-use
- Funder requirements – gaining new and continuation funding
- Institutional reputational and funding risk if there is no infrastructure and/or poor practice
- Being able to find data and combine with other's data
- Identifying versions of data
- To enable sharing and collaboration within a research group and beyond
- Citation impact if made available – getting credit for your work
- Demonstrating value for funding and likelihood of further funding

6. Drivers for managing research data

*"There has been a decisive shift towards greater oversight of the research process motivated by the driving principle of data as a public good. This shift is seen in the concerns of policy-makers, and in changes in legislation and its implementation. The needs are being addressed through coordinated action by funders including the UK Research Councils, charities and JISC, with significant responsibilities falling to HEIs and individual researchers."*¹²

Research integrity

Data management is a means to assure research integrity. The UK Research Integrity Office (UKRIO) states in its Code of Practice that *"Researchers should consider how data will be gathered, analysed and managed, and how and in what form relevant data will eventually be made available to others, at an early stage of the design of the project"*¹³.

The House of Commons Select Committee on Science and Technology concluded in 2011 *"...employers must take responsibility for the integrity of their employees' research"*, and called for regulatory oversight to ensure funders and institutions fulfill their responsibilities¹⁴.

Legislative change and regulatory compliance

Effective data management can mitigate risks to institutional reputation. Commonly researchers need to balance requirements for disclosure and confidentiality, complying with both the **Freedom** of Information Act and the Data **Protection** Act.

¹² Whyte, A., Tedds, J. (2011). 'Making the Case for Research Data Management'. DCC Briefing Papers. Edinburgh: Digital Curation Centre. This section is based on this document.

¹³ UK Research Integrity Office 'Code of Practice for Research', <http://www.ukrio.org/what-we-do/code-of-practice-for-research/>

¹⁴ House of Commons Select Committee on Science and Technology. Eighth Report: Peer Review in Scientific Publications, 2011, <http://www.publications.parliament.uk/pa/cm201012/cmselect/cmsctech/856/...>

Funders' data policies

Research Councils UK 'Common Principles on Data Policy'¹⁵ provides an overarching framework for individual research council policies, focusing on:

- Public good
- Planning for preservation
- Discovery
- Confidentiality
- First use
- Recognition
- Public funding

Across Europe, USA and Australia policies have been developed, and this series of guidance documents highlights the requirements of the UK research bodies.

Research is global and more data intensive

Funders expect UK research to be international in scope and be based on international collaboration. Research and data management practice need to reflect this, and have access to tools, services, support and appropriate IT infrastructure.

Institutional responses

Many UK universities now have Research Data Management policies¹⁶. Responding to drivers including the emphasis on institutional responsibilities e.g. within EPSRC research data expectations¹⁷, and often within the context of Jisc funded Managing Research Data projects, there are developing 'RDM Support Services' which address sustainability of RDM improvement.

7. Benefits of good data management

The benefits of good data management encompass the researcher, the broader research community, the research institution, funders, and the public.

They include, for example, according to the UK Data Archive¹⁸:

- **Efficiency** – makes research easier
- **Safety** – protect valuable data
- **Quality** – better research data
- **Reputation** – enhances research visibility
- **Compliance** – with ethical codes, data protection laws, journal requirements, funder policies

¹⁵ RCUK Common Principles on Data Policy, <http://www.rcuk.ac.uk/research/Pages/DataPolicy.aspx>

¹⁶ Example RDM policies, <http://www2.le.ac.uk/services/research-data/rdm/rdmguidance-leicester/RDM-policy>

¹⁷ EPSRC Policy Framework on Research Data – Expectations, <http://www.epsrc.ac.uk/about/standards/researchdata/Pages/expectations.aspx>

¹⁸ UK Data Archive, 'Data Management: what does it mean for researchers', <http://www.slideshare.net/catharineward/data-management-what-does-it-mean-for-researchers>

By comparison the EPSRC highlight benefits for the research community and public of systematic management and sharing of research data¹⁹:

- Increasing the visibility of research and generating citations, leading to growth of scientific reputation of individual researchers, their research teams and their institutions.
- Reinforcing open scientific inquiry.
- Protecting against use of faulty data by allowing published results to be independently verified, refuted or refined, thus improving the overall quality of research and encouraging diversity of analysis and opinion and helping to resolve scientific disputes.
- Stimulating new approaches to data collection and methods of analysis.
- Increasing awareness of research in related areas leading to more opportunities for collaboration.
- Allowing re-use of data for research not foreseen by the initial investigators – this increases the efficiency of use of public funding by avoiding unnecessary duplication of data collection.
- Permitting the creation of more highly powered data analysis by combining data from multiple sources.
- Facilitating education of new researchers and the wider public.

8. Where to get help and information

Whatever your area of study or research, and however it is to be funded there is support and advice available at the University.

As a signpost to a variety of internal and external resources refer to the University research data website www.le.ac.uk/researchdata. Here specific funder and discipline related information and data management advice is developing alongside general information. It also highlights the range of contacts within the University who offer support relating to the many and varied aspects of data management:

- IT Services
- Library
- Research Support Office
- Academic Practice Unit
- Information Assurance Services
- Enterprise and Business Development

¹⁹ EPSRC, <http://www.epsrc.ac.uk/about/standards/researchdata/Pages/scope.aspx>

A single point of contact is also available: contact researchdata@le.ac.uk at any time and as early as possible in your work and in the bid process. This means specific queries or general request for assistance can be directed to the right place(s). You can also request assistance with development of a data management plan via this email address.

In terms of IT Services support, in 2011 an IT Strategy²⁰ was approved giving a clear mandate to provide comprehensive information management services. This aims to underpin all aspects of the University's mission, including its research. The document responds to funding bodies' announcements regarding principles that must be adopted for managing research data.

Specific research IT services available include Research File Storage²¹, high performance computing, Wiki, 'LAMP' stack (a general purpose, Linux, relational database and web hosting service, based around open source software - Linux, Apache, MySQL and PHP), file transfer and source code control²².

Internal and external RDM training resources are also included within the University RDM website²³.

In addition specific advice regarding data protection and freedom of information can be accessed via Information Assurance Services²⁴. Reference here should also be made to the University Information Security Policy and compulsory staff information security awareness online training.

²⁰ University of Leicester, IT Strategy 2011, <http://www2.le.ac.uk/offices/itservices/about/strategy>

²¹ Research File Storage, <http://www2.le.ac.uk/offices/ithelp/services/rfs>

²² IT Services, <http://www2.le.ac.uk/offices/ithelp/>

²³ University of Leicester RDM website, <http://www2.le.ac.uk/services/research-data/training>

²⁴ University of Leicester Information Assurance Services, <http://www2.le.ac.uk/offices/ias>

University and external research data management information

University Research Data Management Website	www.le.ac.uk/researchdata
University Research Data Management Support	researchdata@le.ac.uk
Research Data Management Training	http://www2.le.ac.uk/services/research-data/training
IT Services for researchers	http://www2.le.ac.uk/offices/ithelp/it4/researchers
University Information Assurance Services – Information Security Policy	http://www2.le.ac.uk/offices/ias/resources/policies/ispolicy
University staff information security awareness online training	http://www2.le.ac.uk/offices/ias/resources/training/online
Whyte, A., Tedds, J. (2011). 'Making the Case for Research Data Management'. DCC Briefing Papers. Edinburgh: Digital Curation Centre	http://www.dcc.ac.uk/resources/briefing-papers/making-case-rdm
Jisc Digital Media	http://www.jiscdigitalmedia.ac.uk/
UK Data Archive – Managing and Sharing Data	http://www.data-archive.ac.uk/media/2894/managingsharing.pdf
UK Data Archive Data Management Checklist	http://data-archive.ac.uk/create-manage/planning-for-sharing/data-management-checklist
Digital Curation Centre Funder's Data resource	http://www.dcc.ac.uk/resources/policy-and-legal/overview-funders-data-policies
Digital Curation Centre 'DMP Online' tool	https://dmponline.dcc.ac.uk/
Digital Curation Centre DMP Checklist	<i>Being reviewed currently</i>
RCUK Common Principles on Data Policy	http://www.rcuk.ac.uk/research/Pages/DataPolicy.aspx
RCUK Joint Electronic Submission System (Je-S)	https://je-s.rcuk.ac.uk/JeS2WebLoginSite/Login.aspx

Research Funder Policies - Research Data Management etc.

RCUK Common Principles on Data Policy	http://www.rcuk.ac.uk/research/Pages/DataPolicy.aspx
AHRC Technical Plan	http://www.ahrc.ac.uk/Funding-Opportunities/Research-funding/RFG/Application-guidance/Pages/Technical-Plan.aspx
AHRC Research Funding Guide	http://www.ahrc.ac.uk/SiteCollectionDocuments/Research-Funding-Guide.pdf
BBSRC Data Sharing Policy (with FAQs and Data Sharing leaflet)	http://www.bbsrc.ac.uk/datasharing
EPSRC Policy Framework on Research Data	www.epsrc.ac.uk/about/standards/researchdata
ESRC Research Data Policy	http://www.esrc.ac.uk/about-esrc/information/data-policy.aspx
MRC guidance - data management plans	www.mrc.ac.uk/Ourresearch/Ethicsresearchguidance/datasharing/DMPs/index.htm
MRC Policy on Research Data-sharing	www.mrc.ac.uk/Ourresearch/Ethicsresearchguidance/datasharing/Policy/index.htm
NERC Data Policy	www.nerc.ac.uk/research/sites/data/policy2011.asp
STFC Scientific Data Policy	http://www.stfc.ac.uk/Resources/pdf/STFC_Scientific_Data_Policy.pdf
STFC Data Management Plan	http://www.stfc.ac.uk/1930.aspx
CRUK Data Sharing Policy	http://www.cancerresearchuk.org/prod_consump/groups/cr_common/@fre/@gen/documents/generalcontent/cr_016308.pdf
NC3Rs website	http://www.nc3rs.org.uk/
NIHR Website	http://www.nihr.ac.uk/research/Pages/default.aspx
NIHR - Integrated Research Application Service (IRAS)	http://www.nres.nhs.uk/applications/integrated-research-application-system/
Wellcome Trust Position on data management and sharing	http://www.wellcome.ac.uk/About-us/Policy/Spotlight-issues/Data-sharing/Data-management-and-sharing/index.htm
Wellcome Trust Guidance for researchers: Developing a data management and sharing plan	http://www.wellcome.ac.uk/About-us/Policy/Spotlight-issues/Data-sharing/Guidance-for-researchers/index.htm

Managing research data - an Introduction to practical issues

9. Data Management Planning

Data Management Planning is increasingly being seen as an integral part of the research process, and an expectation, leading to the implementation of good data practices. A Data Management Plan (DMP) is a formal document that outlines how you will handle your data both during your research, and after the project is completed.

A number of research funders now require submission of either an outline or detailed data management plan (DMP) as part of the research proposal process.

Whether or not there is an external prompt for a Data Management Plan it is recommended that DMPs should be routinely created as part of the research process as a beneficial tool for research management.

Your DMP should be written with two distinct audiences in mind: fellow scientists in your own field and data managers. Avoid generalisations and be concise. Before you write anything, consider the complexity of the data you intend to create and how much long-term value your data is likely to offer to the wider research community.

If the scale, complexity and cost of managing your data are low, your DMP could be less than 500 words long. In the case of population cohorts, genetic and omic data, bio banks and other datasets that are potentially rich resources for the wider community, your DMP is likely to be considerably longer.

Example: if your research will form part of an established population or patient-based study, your DMP will be significantly longer, as it must indicate how you will meet twenty-one additional Medical Research Council (MRC) requirements²⁵.

²⁵ MRC Policy on Sharing of Research Data From Population and Patient Studies, www.mrc.ac.uk/Ourresearch/Ethicsresearchguidance/datasharing/Policy/PHSPolicy/index.htm

The Digital Curation Centre (DCC) provides the 'DMP Online' resource to enable you to create a DMP. This tool has been developed in collaboration with UK Universities including the University of Leicester, where it has been piloted.²⁶ This, and the related DMP Checklist²⁷ includes a range of common issues faced by and which should be addressed within research:

- **Data Collection**
 - What data will you collect or create?
 - How will the data be created?
- **Documentation and Metadata**
 - What documentation and metadata will accompany the data?
- **Ethics and Legal Compliance**
 - How will you manage any ethical issues?
 - How will you manage copyright and Intellectual Property Rights (IPR) issues?
- **Storage and Backup**
 - How will the data be stored and backed up?
 - How will you manage access and security?
- **Selection and Preservation**
 - Which data are of long-term value and should be shared and/or preserved?
 - What is the long-term preservation plan for the dataset?
- **Data Sharing**
 - How will you share the data?
 - Are any restrictions on data sharing required?
- **Responsibilities and Resources**
 - Who will be responsible for data management?
 - What resources will you require to deliver your plan?
- **Relevant policies and procedures (e.g. institutional RDM Policy)**

Reference should also be made to the UK Data Archive Data Management Checklist²⁸.

²⁶ DCC DMP-Online, <https://dmponline.dcc.ac.uk/>

²⁷ Currently being reviewed by the DCC, <http://www.dcc.ac.uk/news/consultation-data-management-planning-checklist-and-themes>

²⁸ UK Data Archive Data Management Checklist, <http://data-archive.ac.uk/create-manage/planning-for-sharing/data-management-checklist>

10. Ensuring the quality of data

Quality should be considered whenever data is created or altered, for instance at the time of data collection, data entry or digitisation. Funders often ask for information about the procedures you will carry out to ensure data quality is maintained, such as putting time aside to validate data or entering values into pre-prepared databases or templates.

11. Research data format

Some disciplines have well defined guidelines on which file formats to use and how to describe or catalogue a dataset to best support secondary data users. Other disciplines are still developing these. A major barrier to data sharing is the widespread use of non-standard formats or rapidly obsolete commercial formats. Many research funders do not dictate the formats you should use but do ask that you justify your decisions to use certain data formats over others. When selecting a data format to use, your own research needs must come first.

If you find you need to use a non-standard or rarely used format, you should consider converting your data to a more widely re-usable format once your own data analysis is complete.

The re-usable format you select should be as accessible as possible to as many people as possible. In order to make use of any data a number of digital technologies must be available, these are known as technological 'dependencies'. These may be fairly common technologies such as; a desktop PC, the Windows 7 operating system and Adobe Reader 9 software. Alternatively, the technology required to access data might be rare and hard to acquire or even unique. You should address this challenge as far as it is possible, by minimising the number of unnecessary technological dependencies involved in using your data.

You should also favour 'open' technologies rather than proprietary ones whenever possible. Proprietary technologies are owned by a commercial company or group of companies. Commercial pressures may lead to the withdrawal of a particular piece of hardware or software and its replacement with a new version. Open technologies are supported by community of users and so do not have the same commercial vulnerabilities.

If you're unsure which file formats to use the UK Data Archive publish a list of recommended deposit formats²⁹. These formats are appropriate for many non-specialised uses.

12. Describing research data - Metadata

Without an accurate description of what a dataset actually consists of, why it was created or collected and what secondary users are permitted to do with it, the value of research data is greatly reduced. Wherever a dataset is made available it should be accompanied by a useful description.

Metadata is 'data about data' and is information (or 'cataloguing information') that enables data users to find and or use a dataset. A description of a dataset is often kept in a separate, dedicated database or spread sheet, or simply within MS Word document 'Properties'. As with data formats,

²⁹ UK Data Archive File Formats Table, www.data-archive.ac.uk/create-manage/format/formats-table

some disciplines offer precise guideline on how to describe a dataset while, for other disciplines, no such guidelines yet exist.

In your DMP you should outline plans for documenting your research data to meet both your own needs and those of later users. It is generally best to use established and shared metadata standards, rather than create new ones. This helps with consistency and saves effort.

Prominent metadata standards include Dublin Core³⁰ (Dublin Core Metadata Initiative) which is a set of vocabulary terms used to describe resources for the purposes of discovery. The Data Documentation Initiative³¹ (DDI) is an international standard for describing data from the social, behavioural and economic sciences. DDI uses XML to allow metadata to be encoded in a standardised way, simplifying data sharing and subsequent re-use.

If you find you have to take a pragmatic approach to describing data, it may help to imagine another data user attempting to make sense of your data in your absence, after your project has concluded. If presented with only the data itself the other user may well be faced with the difficult task of 'unpicking' it. How will they make sense of your file and folder naming conventions? What extra information would they need to make the maximum use of your data? How were new datasets derived from raw data?

13. Data storage, selection, preservation and archive

You should explain where your data will be stored, how it will be organised in the short term and who will back it up.

You need to decide how you will keep your data safe in the short-term *before* it's deposited in a long-term secure storage facility. This is particularly important if you are conducting field research. As a minimum requirement, try to ensure that at all times at least two copies of the data exist and that every copy can easily be accounted for and located if required. Mobile devices such as laptops, external disc drives and voice recorders should be encrypted. You should plan for data transfer to secure storage as early as possible.

If you do not have existing data storage arrangements, it is recommended that when you create data you store it in the University's Research File Storage facility (RFS), managed by IT Services. All those with research storage needs are able to register for this service in order to be allocated a storage allocation appropriate to the project³².

Researchers are not charged for this service (unless requirements are extremely large), it offers peace of mind (for the researcher and research funder) and reduces researcher IT responsibilities, being managed by IT Services as a secure service, backed-up daily.

If you do not make use of the RFS, your storage provider's back-up security standards and procedures should be described instead. If you will be working collaboratively with other

³⁰ Dublin Core Metadata Initiative, <http://dublincore.org/>

³¹ Data Documentation Initiative, www.ddialliance.org/getting-started

³² Research File Storage, <http://www2.le.ac.uk/offices/ithelp/services/rfs>

institutions, make sure that the security and back-up procedures of each data-holding partner are described in your DMP.

Funders commonly also have requirements regarding long-term preservation and archiving, necessitating appropriate selection and discarding of duplicate data etc. There are numerous research data archives, often associated with funders and specific research disciplines e.g. Archaeology Data Service³³.

14. Sharing research and providing long term access to data

Most major funding bodies require non-confidential research data to be not only retained but also actively shared at the end of a funded project. This will either be done via deposit of data into an established national 'data centre' (supported by a research funding body), via a discipline-specific data repository (usually supported by a number of different organisations) or via an institutional research data repository.

The option for researchers to provide access directly is also sometimes acceptable to funders. Data may have to be retrained and shared for many years (potentially even in perpetuity) as a condition of funding and during this period the contact details and responsibilities of individual researchers can be expected to change. These factors usually make this option impractical.

Example: The Wellcome Trust maintains a list of major biomedical data repositories that preserve and provide access to research data. Researchers may choose to share their data by depositing it in a repository such as the UK Data Archive³⁴.

The University requires that all research publications are deposited with the Leicester Research Archive (LRA)³⁵.

Data must be shared in a timely and responsible manner. The MRC for example recognises that on-going research contributing to the completion of datasets must not be compromised by premature sharing and analysis, therefore a limited *and defined* period of exclusive data use is reasonable. Access to data may also be delayed for a short period to allow time for the preparation and filing of patent applications. You should indicate which data cannot be retained and/or shared, and explain why this is necessary.

15. Ethical considerations

All major research funders recognise that some research data cannot be shared as it is particularly sensitive. Most data however, can be shared, at least on part, if certain precautions are taken. Commonly used precautions are data anonymisation, seeking permissions to share data *at the time of data collection* and restricting access via the need for a signed secondary user agreement.

³³ Archaeology Data Service, <http://archaeologydataservice.ac.uk/>

³⁴ UK Data Archive www.data-archive.ac.uk

³⁵ Leicester Research Archive, <http://www2.le.ac.uk/library/find/lra>

The immense benefits of sharing data within areas such as medicine and the biological and social sciences are a powerful argument for doing so. Your plans to share research data should reflect your ethical plan, if you have one.

16. Personal information

Issues regarding the management of personal information are particularly pertinent to areas such as health and medical research e.g. with NIHR or MRC funding.

Example: NIHR

NIHR (National Institute of Health Research) funding applications involve completion of detailed 'IRAS' forms focussing on handling of confidential/sensitive/personal information.

Example: MRC

If your research is part of an established study, data security measures will already be in place and these should be summarised in your DMP. All MRC-funded researchers have a responsibility to ensure that opportunities for data re-use are maximised, within the regulatory requirements of the law. MRC adopts the view that the potential benefits to patients and the public should outweigh identified risks. The MRC's guide to using Personal Information in Medical Research³⁶ states:

“Principal investigators must take personal responsibility for ensuring (as far as is reasonably practical) that training, procedures, supervision, and data security arrangements are sufficient to prevent unauthorised breaches of confidentiality.”

If you are planning to make use of a data storage facility that complies with a recognised information security standard (such as ISO 27001), explain this in your DMP. If not, describe the main risks to the security of any data relating to human participants and how these risks will be addressed, such as by access control or encryption.

17. Copyright and IRP

Research funders also recognise the challenges faced by researchers who wish to commercially exploit the research data they have created. This is broadly encouraged by funders who typically grant a defined period of 'exclusive data access' after a project has ended, during which researchers can exploit any commercial potential the research data may have.

It is generally a requirement of funding that, after this period has ended, research data will be made 'freely available' i.e. without charge. At the University advice in this area can be received from the Enterprise and Business Development Office³⁷.

³⁶ Personal Information in Medical Research, www.mrc.ac.uk/Utilities/Documentrecord/index.htm?d=MRC002452

³⁷ Enterprise and Business Development Office, <http://www2.le.ac.uk/offices/ebd>

18. Re-using existing data

Funders recognise that researchers use existing data in increasingly diverse ways: for instance, by using data linking or meta-analysis. You should identify existing datasets you expect to draw on. If you intend to generate new data, you should also explain why this is necessary.

Research that makes use of existing datasets should meet the same high standards as all research in regards to quality, ethical requirements and value for money. They should also add recognisable value to the original datasets.

If you have previously collected or generated research data of your own which you intend to use as part of new research, you should ensure that your DMP describes procedures for managing both existing and any newly generated data.

19. Citing research data in research outputs

From 1 April 2013 all the UK's Research Funding Councils, as part of RCUK, required research outputs (i.e. journal articles) to provide a means by which third parties can access any underpinning research datasets. This may be a reference (such as a unique URL or DOI) printed in a paper which will lead an enquirer to a specific web page where the data is available. Alternatively the enquirer might be directed to a page which displays the contact details of a custodian of the data and asked to email them in order to gain access to the data.

Given the extended timescales involved in publication, it is strongly recommended that the authors of published academic outputs *do not provide their current contact details* as a means of accessing underpinning research data, as these will change over time. If you plan to use an established data repository service, ask it for a unique reference identifier which could be included in the publication instead.

20. Roles and responsibilities

Data management responsibilities should be clearly assigned to named individuals within your DMP. The Principal Investigator (or Research Director for larger studies) is usually responsible for research data sharing. The person responsible for maintaining and updating the DMP (if this is different) should also be named in your DMP.

Support services are in place at the University to help you manage and share your research data (see [8. Where to get help and information](#)) and any such services you plan to use should be mentioned in your DMP to demonstrate the role played by data specialists.

21. The cost of managing research data

If any costs are involved in meeting funder data management requirements (for example, the cost of dedicated effort, equipment or software tools for managing, storing or providing access to your data), these should be mentioned in your application. If the costs are substantial, you should differentiate between:

- Costs associated with collecting and/or processing new data
 - Your own research on newly acquired and legacy data
 - On-going data curation and preservation
 - Providing access and data sharing
-

The Managing Research Data guide series comprises:

- An Introduction to Managing Research Data – For Researchers and Students
- Data Management Planning – AHRC funding applicants
- Data Management Planning – BBSRC funding applicants
- Data Management Planning – EPSRC funding applicants
- Data Management Planning – ESRC funding applicants
- Data Management Planning – MRC funding applicants
- Data Management Planning – NERC funding applicants
- Data Management Planning – STFC funding applicants
- Data Management Planning – Non-RCUK funding applicants

They are part of a range of RDM material produced by the University, all available via www.le.ac.uk/researchdata.

University of Leicester

What would you do if you lost your research data tomorrow?

Take the research data health check... and find help to secure, share and exploit your valuable research.

Chances are you could use some helpful pointers in all of these!

Create	Organise	Keep	Find & Share
<p>Have you...</p> <ul style="list-style-type: none"> <input type="checkbox"/> fully understood your research funders' data management requirements? Government and funders require that publicly funded research is made available for reuse - do you set in date with their latest policies? Your future funding might depend on it! <input type="checkbox"/> written a data management plan? Your funder may already require this but build it in from the proposal stage to avoid headaches in the future. <input type="checkbox"/> gained ethics approval/comment? Writing a data management plan will aid planning and help you to navigate ethics and governance requirements. <input type="checkbox"/> protected your intellectual property? Leaving intellectual property considerations for a later date could have you appropriate needs, safeguarding future projects and perhaps your financial future too! 	<p>Are your research files and data...</p> <ul style="list-style-type: none"> <input type="checkbox"/> clearly described, in terms of content being standard nomenclature? An unclear filename is in danger of data being lost! Will you be able to remember how you generated your data, and will you or anyone else be able to find it in the future when you wish to reuse and share? <input type="checkbox"/> clearly labelled with versions and dates? How will you remember which was the definitive version and which others are each produced by your research interest? <input type="checkbox"/> logically structured and named? Once you've remembered how you generated data, can you do that for the research too? <input type="checkbox"/> having provided against broken links, using persistent identifiers? The persistent identification of digital resources can play a vital role in ensuring that accessibility and availability over time using recommended data standards. 	<p>Do you know...</p> <ul style="list-style-type: none"> <input type="checkbox"/> how to restrict access to your research data to the right people? Have you consulted with university or data centre experts so that only the right people have access to your research? <input type="checkbox"/> which data to keep and which data to discard? Managing research data effectively means being selective; which data to discard and when is vital to keep and for how long? <input type="checkbox"/> how securely your data is stored? What happens if your storage media fail? How robust is it? Could it get left on the train and could somebody else misuse it? <input type="checkbox"/> how your data is backed-up? Have you made use of university and/or external resources to back up data so that you have multiple copies in case of loss or theft? 	<p>Do you know how to...</p> <ul style="list-style-type: none"> <input type="checkbox"/> find existing information resources related to your research? Where can you find research data that you can re-use or combine with your work to produce new research? <input type="checkbox"/> share data with your collaborators securely and effectively? Whether building a collaborative project, generating results for others to comment on or sharing the final product of your research, how will you fit with your colleagues' needs or other things? <input type="checkbox"/> deposit your research data and outputs in an open repository? Is there an appropriate discipline or institutional repository and what do you need to do to deposit your research output? That should be used effectively. <input type="checkbox"/> publish your research, and get it cited as well? Institutions and data centres must make research data available to others with providing credit to the researchers who did the work. Your future career could depend on it!

To find information, support, advice and training, as well as links to external resources, go to www.le.ac.uk/researchdata or email: researchdata@le.ac.uk



© University of Leicester
Leicester LE1 7RH
UK

www.le.ac.uk