Citation for published version:
Darlington, M & Ball, A 2012, *A Research Data Management Plan for Engineering Research*. University of Bath, Bath, UK.

Publication date:
2012

Link to publication

Publisher Rights
CC BY

University of Bath

Alternative formats
If you require this document in an alternative format, please contact:
openaccess@bath.ac.uk

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 23. Jan. 2025

A Research Data Management Plan for Engineering Research

Mansur Darlington & Alex Ball

redm3rep120521mjd10.pdf

ISSUE DATE: 31 May 2012

Catalogue Entry

Title	A Research Data Management Plan for Engineering Research
Creator	Mansur Darlington
Contributor	Alex Ball
Subject	research data management; data management plans
Description	This department-level Research Data Management Plan is offered as a starting point for engineering departments establishing their own policies and procedures for data management at the project level. It recommends how projects should use data management plans: where they should be stored, how open they should be, expectations for review and revision, and so on. It provides a template for project data management plans. Lastly, it recommends tools that can be used for data management, and additional guidance that may be consulted.
Publisher	University of Bath
Date	21 May 2012 (creation)
Version	1.0
Type	Text
Format	Portable Document Format version 1.4
Resource Identifier	redm3rep120521mjd10.pdf
Language	English
Rights	© 2012 University of Bath

Citation Guidelines

Mansur Darlington & Alex Ball. (2012). *A Research Data Management Plan for Engineering Research* (version 1.0). REDm-MED Project Document redm3rep120521mjd10. Bath, UK: University of Bath.

TABLE OF CONTENTS

1. INTRODUCTION 4

2. MANAGING PROJECT-BASED RESEARCH DATA 5

2.1 The purpose and scope of the ERDMP 5
2.1.1 Research life-cycle support 5
2.1.2 Data preparation support 6
2.1.3 Project-level research data management planning 6
2.2 Tools and guidance for the execution of project data management 7

3. A DATA MANAGEMENT PLAN FOR ENGINEERING RESEARCH 7

3.1 Infrastructure and implementation issues 7
3.1.1 Relating the DMP to other documentation 8
3.1.2 Relating other documentation to the DMP 10
3.1.3 Understanding the DMP 10
3.1.4 Rôles and responsibilities 10
3.1.5 Creation and development of the Project Data Management Plan 10
3.1.6 Review of the DMP 11
3.1.7 Revision of the DMP 11
3.1.8 Budgeting for research data management 12
3.1.9 Data Security 12
3.1.10 Identifying Contractual and legal obligations (RS5, RS6) 14
3.1.11 End-of-Life 14
3.2 DMP contents 14
3.2.1 Summary of Research Activity 15
3.2.2 Existing Data re-use 15
3.2.3 Relating new data to existing data 15
3.2.4 Future use of the data 16
3.2.5 The Project Record Manifest 16
3.2.6 Data generation and manipulation 17
3.2.7 Data organization 17
3.2.8 Data quality 18
3.2.9 Data structures and formats 18
3.2.10 Data semantics 19

4. REFERENCES 20

APPENDIX A: DATA MANAGEMENT TERMINOLOGY 21

1. INTRODUCTION

The need for better management of research data is increasingly becoming understood not only by those who collect and generate data for research first use, but also by those who have a research funding or governance rôle.

The motivation for better management comes from the recognition that the best return on the investment made in data acquisition can only be realized through using it in the most effective way and by maximizing its sharing and re-use. Better data management aimed at promoting enhanced availability of and easier access to existing data – the open-data philosophy - can result in a number of benefits. These include reducing the need for redoing work, making research work more transparent, providing a basis for validation of work and making the repetition of collection or generation unnecessary.

As a result it is now becoming more usual for research funders to expect evidence of formal data management planning as early as in the initial bid for funds, with detailed plans being required early in the management of a funded project. At the same time, those who are engaged in the management of research in commercial environments are become more aware of the diverse merits of better data management.

In the following pages is presented a Research Data Management Plan suitable for those engaged in general research in engineering topics. The plan will be referred to henceforth in this document as the ERDMP.

This plan is based is based on the ME-RDMP (Darlington and Ball, 2012a) which was developed during the REDm-MED Project, funded by JISC as part of the second phase of the MRD Programme, an exploration into managing research data resulting in the provision of policies, methods, procedures and tools in support of such management.

The ME-RDMP was, in turn, based on some of the precursor work and deliverables from the ERIM Project and a number of precursor documents (e.g. McMahon, et al. 2009; DCC, n.d.). In addition, the detail of that plan is a response to the requirements specifications contained in the *Research Data Management Plan Requirements Specification for the Department of Mechanical Engineering (RDMPRS)* at the University of Bath (Darlington & Ball, 2012b) a deliverable of the REDm-MED Project.

The ERDMP is a department-level plan which should be seen as a specification for each project-level plan which will be required to be prepared for execution based on the specific requirements of the project and the nature of the data that are expected to be collected and generated during the project.

A number of assumptions have been made in the creation of this data management plan. The first is that there is some expectation that governance at the departmental level and at institution level will embrace matters relating to data management and that, therefore, support in the form of higher-level mandates and infrastructure provision will be made. The second assumption is that there will be in place, at department, faculty or institution level, an individual whose rôle is to support the management of research data; the rôle might be a senior data manager of data scientist, with appropriate credentials and expertise. The same would apply in a commercial research environment.

2. MANAGING PROJECT-BASED RESEARCH DATA

In order for co-ordinated data management to be carried out on a project-by-project basis it is necessary to have management procedures in place to support this process – together with a suitable infrastructure – including a higher-level data management plan from which a specific management plan can be drafted as appropriate to the character of each individual project and its research data. The ERDMP is just such a higher-level plan.

This data management is based on the analysis of a typical higher education department with a diverse spectrum of research in mechanical engineering topics. As such the department can be considered representative of many groups of researchers in terms of its generation, development, use and archiving of research data, and the attendant problems of research data management. As a result, many of the details of its data management plan will be generalizable directly to other research centres the research data characteristics of which are similar. This will include not only other departments of mechanical engineering but, more broadly, the departments of engineering with other research specialities. Further, given the very broad nature of the research data encountered in engineering research it may be that parts of the DMP would be usefully adopted outside the sphere of engineering research. The term 'department' can be interpreted as any organizational group within which research projects are carried out and managed, within both academia and industry. The term 'institution' likewise may be interpreted as any establishment or enterprise engaged in research projects.

2.1 The purpose and scope of the ERDMP

The purpose of the ERDMP is to provide data management planning support to those such as principal investigators, project, data and research managers and researchers and others, such as service providers, who have a need to develop, manage, execute, or to provide support for, the individual project-level data management plans which will be derived from this department-level plan and which in many cases will be a mandatory governance requirement for a funded project.

The ERDMP provides the support necessary to manage data throughout the research life-cycle, including data preparation and development activities; and creation, development and execution of **Project Data Management Plan**s. In doing so it will be necessary to identify not only the processes and procedures necessary for data management, but also the infrastructure to support such processes and procedures. Such infrastructure will include physical elements, services and tools.

In the following, examples of such tools will be taken from those recommended for use in the ME-RDMP. In addition, an analysis of infrastructure elements that follow can be found in Infrastructure Supporting a Research Data Management Plan for the Department of Mechanical Engineering, University of Bath which, amongst other things, identifies the sorts of general information and resource requirements necessary for implementing a research DMP.

2.1.1 Research life-cycle support

It is intended that the ERDMP and project-level DMPs specialized from it will fulfil the needs of:

- data management planning,
- data management execution,
- internal and external bid submission,

- project planning,
- doing the research,
- collaboration with colleagues and partners in academia and industry and others,
- supporting data re-use and re-purposing,
- preparation for long-term preservation,
- end-of-life.

The items identified above constitute the research project life-cycle tasks, some of these being continuous and overlapping during the life of a project.

2.1.2 Data preparation support

The plan will include provision for the management of both physical (paper documents) and digital **Record**s, and of research artefacts which themselves constitute or contain data.

The plan will support three types of data preparation activity, identified in earlier work:

1. making existing research data available and fit for a future known research activity (*data re-purposing*);
2. managing existing research data such that it will be available for a future unknown research activity (*supporting data re-use*);
3. using research data for a research purpose or activity other than that for which it was intended (*data re-use*).

To these activities will be added any that are required additionally expressly for making data ready for long-term preservation through archiving.

2.1.3 Project-level research data management planning

The ERDMP should be seen as a broad specification to be used for development of the individual data management plan required for each research project carried out in the department.

In the first instance the development of such a DMP will be to the extent necessary that it conforms to the requirements of and supports the PI or project manager in bid planning and submission.

Then, each complete project level-DMP produced as a specialization of this department-level DMP will have two functions. In the first instance, it will act as a guide to researchers on re-using existing data, re-purposing their own data and supporting data re-use throughout the **Research Activity**. At the end of the Research Activity, its purpose is to act as a record of how the data have been re-used and re-purposed, where applicable, and how data re-use has been supported. Further, the DMP will provide information about the location, accessibility and ownership of the data associated with a project as a means of supporting the principal motivation of research data management which is to promote second use.

It can be seen that much of the data management necessary to fulfil the two functions referred to above must be made during the research activity, since it is during this activity that most is known, and can be recorded about, the data, and the most effective actions can be taken for supporting the three types of activity outlined above.

Many of the terms used in this document have been defined in Howard, et al (2010) in a terminology for research data management; the key definitions in their entirety can be accessed in the on-line terminology and those used within this DMP can be found in a glossary, in Appendix A to which they are hyperlinked at first use here.

2.2 Tools and guidance for the execution of project data management

While every effort has been made within the ERDMP to recommend guidance and tools to be used to support and execute data management at the project level, gaps remain in some areas. The primary reason for this is that mature, intuitive tools and practice-based guidance documents do not exist for all the tasks necessary for good data management, though much work is in progress to remedy this situation. Furthermore, the integration of such tools as there are is in its early stages.

It is anticipated that PIs, Data Managers and Researchers will have access, to some extent, to specialist local assistance in relation to data management (for example, in the person of the data manager or data scientist referred to earlier). General information and expertise may be found on the website of the Digital Curation Centre (DCC) where, amongst other things, a helpdesk service for answering specific queries is provided. Another possible source of expertise is the Research Data Management discussion list which is provided by JISC.

3. A DATA MANAGEMENT PLAN FOR ENGINEERING RESEARCH

It is recommended that the ERDMP be used in its entirety – together with such tools as are available – to guide the process of developing a DMP for a specific Research Activity. Such a plan is referred to henceforth in this document as a **Project Data Management Plan** (Project DMP).

Provision of a draft Project DMP, which fulfils the requirements outlined, may be a requirement of the bid for project funding and then subsequently, in completed form, as part of the requirement for project governance. Without such a plan, it is unlikely that the research data will be managed in such a way as to maximise its current and future usefulness and, thus, the maximum value of the data will not be achieved.

The ERDMP which follows is composed of two parts. The first part – Section 3.1 Infrastructure and Implementational Issues – concerns matters which must be in place in order for development and provision of data management plans and for researcher-based data management (carried out during the research activity and specified in the second part) to be supported and carried out satisfactorily. Details of the implementation of these form part of this DMP. The second part – Section 3.2 DMP Contents – provides details of the plan itself which will provide guidance and prescription for the implementation at the time for each Research Activity (project) and associated **Data Case**. Thus, it is assumed that for each research project undertaken a data management plan will be produced using the guidance found here, taking into account the special needs which will spring from the particulars of the research activity and the data being managed. The DMP will take into account both management of research data and management of other information associated with the project.

3.1 Infrastructure and implementation issues

Delivering and executing a Project DMP is dependent on certain supporting mechanisms either being in place as part of the Department's management infrastructure or being put in

place as part of execution of the Project DMP. These items are enumerated below in terms of the functions they will support.

3.1.1 Relating the DMP to other documentation

The Department data management plan, and any Project DMPs that are an implementation of it may be informed by a number of key documents relating to the governance of research in the department or establishment. One generic document that provides such guidance is the RCUK Policy and Code of Conduct on the Governance of Good Research Conduct.

In addition, of particular importance – because of the EPSRC's important position as funder of engineering research – is the on-line research data management expectations guide published by the EPSRC.

Documentation specific to a research project or research activity should be identified from a single source as indentified in that research activity's data management plan. For all research projects a default on-line or other local source should be recommended or mandated for this purpose. For example, for the Department of Mechanical Engineering at the University of Bath (UoBDofME) a departmental RDM wiki has been created within which pages may be created for dealing with the research data management for individual projects. In particular, it is here details would be given of key documents and where they are located, even though some of these documents may not be universally available.

Records may be stored within such a wiki space, where appropriate. As a default, in the interests of an open-data philosophy, access to such documentation should be unrestricted. However, where the interests of the protection of intellectual property and commercial interests are paramount, especially where adherence to contractual agreement is required, then this information should be made access-restricted as appropriate to the case. If necessary, the manager responsible (perhaps the principle investigator or project manager) should take advice on this from the establishment's legal department. Redirection may be provided to documentation located elsewhere including that having access restrictions and for which access will be available only to those who have the appropriate access authority at that location. Documentation that is to be made available under some restriction should be placed in some password-protected location; this could be in a department-level RDM wiki or within a dedicated Project wiki, or a password-protected web or other storage space.

It should be noted, however, that to conform to the requirements for perpetuated discovery of and access to project RDM documentation, the following key documentation (at least) should be made available:

- The project proposal document.
- The detailed project plan(s).
- The **Project Data Management Plan** (final version).
- The **Project Record Manifest**.
- Confidentiality agreements (where such agreements are themselves not confidential).
- IPR statements and other documents that affect how the research data may be used.

Uniquely important amongst these documents will be the project record manifest (PRM). This template-based document will list all key project documents and other records and their associations, locations, authors and access particulars. An example of such a template is that created for the UoBDofME.

In addition to this the locations can be given of any other management documents relating to the research activity or protocols, regulations or procedures for carrying out the research activity. These might include requirements and guidance from a receiving repository, if any, in relation to data management, ethics forms, etc.

Ideally, in the interests of making research data discoverable, accessible, interpretable and usable, the items identified above should be access-free. However, care must be taken to ensure that sensitive and confidential information is protected in an appropriate manner and where necessary redacted versions of documents are used. If necessary elements of the documentation should not available until the end of the Project, or until such time as is appropriate to the case. The responsibility of ensuring appropriate sharing or protection of documentation, and the timing thereof, rests with the PI.

The location of electronic records should in general be identified using, where possible, an embedded URL (i.e. a hyperlink). For physical records (including physical specimens or objects, which represent 'data') a description of the physical location should be given, together with the name and contact details of the owner of the records.

To ensure persistence of the information a description of the physical object, with information about its ownership and location should be placed on a secure, persistent wiki page (see the following sub-section below) dedicated to this function. All references to the location and ownership should be made on this page, a reference to which should be placed in the Project Record Manifest in the section related to project management documentation.

Providing Access-free and Access-restricted Collaboration Space

There are a number of ways of making access-controlled and persistent links to documentation. These include using conventional web space, using wikis and using dedicated collaboration tools such as MS Sharepoint and Sakai. All these methods provide a means of controlling the security and confidentiality of documentation through the use of password-controlled access. The recommended means of making persistent links to Department research project documentation in the UoBDofME is through the use of Bath Wiki pages. For each project a project wiki space can be created by the principal researcher or designated project manager. This space can then be used for activities during the course of the project (e.g. project management and communications and deliverable and paper co-creation, etc.) and to provide access to and redirection to project management documentation and research data both during and after the project.

To ensure persistent access to information about project documentation and data it is necessary to arrange for space that is controlled in different ways. Usually it is necessary to have working space for the project the access to which is limited to those in the research group.

By its nature, however, some project management documentation and a large proportion of, specifically, data management documentation must be universally accessible. To this end, a Project Data Management wiki page should be created in an appropriate place, linked reciprocally from the project wiki (and, perhaps, project web site page(s)).

Links can be made from this wiki page, to documentation that is located elsewhere and is password protected. The location of these documents may be in an access-restricted part of the wiki or some other secure environment.

3.1.2 Relating other documentation to the DMP

A system must be in place to allow readers of high-level project documentation and users of the Data Case itself to find the corresponding DMP. To satisfy this requirement, a default location should be specified for project documentation.

3.1.3 Understanding the DMP

Because of the unfamiliarity that many researchers have with the notion and practice of research data management, it may be judicious to provide readers of the Project DMP with precursor documents which place in context the DMP and the need for its creation and execution. Such documents might include higher-level rationale and mandates both from within the establishment and, where, appropriate from external partners. So for example, it would be worth including this document's location in the preliminaries of each Project RDMP.

The URL for the ERDMP (i.e. this document) is: http://opus.bath.ac.uk/30104

The URL for the ME-RDMP (the precursor to this document) is: http://opus.bath.ac.uk/30099/

3.1.4 Rôles and responsibilities

The responsibilities for writing, implementing and reviewing the Project DMP may be shared between a number of rôles within the Department. These rôles and the individuals associated with them should be identified in the Project DMP. Additionally, where documentation clarifies the responsibility it is suggested that it should be identified and made available to the local research community

It is implicit in this that the project data management requirements be considered, according to the perceived needs of the project research, at the time of development of case for funding the research. It is becoming increasingly the case – certainly in academia – that delivery of a complete Project DMP, either as part of the submission document or as an early deliverable, will be required by the project funding body. The task of origination of the Project DMP will fall, accordingly, either to the author of the project proposal or another person, probably the project researcher or, in larger projects, to a project manager. The name of the plan's authors and their contact details should be appended to the Project DMP and be kept up to date.

3.1.5 Creation and development of the Project Data Management Plan

This document – the ERDMP – provides the foundational information for the creation and development of the Project Data Management Plan. Section 2.2 below identifies the key components of the plan itself.

For initial creation, and subsequent through-project development of the Project DMP, it is strongly recommended that use be made of the DMP Online tool, which is provided by DCC. This tool has been constructed to allow customized templates to be uploaded that reflects acknowledgement of the fact that 'one size doesn't necessarily fit all' and that the local research management may vary. By way of example, this tool already contains a template that reflects the local needs of the UoBDofME (visible under 'Institutional templates' when

creating a new DMP) and provides DMP creation and versioning support. The outputs of the tool, that is, the initial DMP and any subsequent versions, are stored on the DMP Online site where they can be accessed and revised as necessary. Users of the service are strongly advised, however, to export copies of these outputs and store them locally.

It is recommended that the file name and versioning for these exported DMPs conform to some specified syntax. An example of this used at the UoBDofME is Naming Documents for Better Organization, which provides one method of achieving file naming and version control, in the absence of an over-arching electronic document management system to handle such things. At the same time it extols the virtues of augmenting the use of formal file names with that of document titles.

There may well be a good reason for using other tools to create a Project DMP. However, for ease of administration, and to help later interpretation of management documents – at a time perhaps well into the future – it is suggested that a preferred means of developing data management plans is identified. There will then be an expectation that DMPs will conform to that format. If the choice is made not to use the preferred means, the rationale for the choice should be included in the project documentation.

The Project DMP can be created either as a wiki page within the project documentation or as a stand-alone document attached to it. In the latter case, the same recommendations apply for file naming and versioning as for DMPs exported from DMP Online. In the interests of open-data provision all such documents should be given the most liberal access permissions appropriate to the confidentiality or sensitivity of the documents.

The wiki has been proposed as a good means of recording and publishing data management documentation at the project level. The same mechanism might be suitable for publishing similar, department or institution level documentation. An alternative to this, where an interactive and collaborative working space is not required, is the use of web pages to publish static documentation and up-to-date information and links.

3.1.6 Review of the DMP

The Project DMP acts as guidance and as a record of activity; its provision may fulfil the governance requirements of the University or management requirements of other establishments – it will provide the *potential* for good data management. Implementing and conforming to the Project DMP will promote good data management practice and result in better-managed data, making its use and re-use more effective.

To ensure conformance with and accuracy of the Project DMP, and to ensure that the data management arrangements best support the research data as the research activity unfolds, reviews of the Project DMP will be required during the project. The timings of these reviews, perhaps being included in regular project meetings, will be recorded in the Project DMP itself, under this heading.

3.1.7 Revision of the DMP

The DMP is not a static document, but should be updated throughout the project to reflect what actually happened. It is therefore important that the versioning and amendment history of the DMP be created and maintained.

For stand-alone documents, such as those created using the DMP Online tool (see Section 3.1.5), two practices will together satisfy the requirement. First, versioning should conform to

the local guidelines for project documentation file naming and version control. Second, change dates and authorship should be recorded within the Project DMP, together with author contact details.

A Project DMP created as a wiki page will have its revision history tracked automatically by the system.

3.1.8 Budgeting for research data management

It is recognized by academic funding bodies that costs for data management will be accrued as a result of expenditure of management time and for storage. A number of these bodies (e.g. ESRC, BBRC, NERC and Wellcome Trust) expect that these costs be identified in the project submission budget and that they will be met by them as part of the full economic costing of the project. Other bodies may well follow suit.

Where the funder expects to meet such costs, the cost of implementing the Project DMP and, if appropriate, future preservation activity, should be estimated and factored into the project budget. It follows that a data management plan of detail sufficient to identify the data management budgetary requirements will be a required part of any funding submission.

It continues to be a widespread belief that the cost of research data management consists essentially in the cost of the provision and use of data storage space. It should be realized that this is only one – perhaps the least expensive and most easily predicted – part of the costs that will be expended. Other costs will accrue from the management of the management of research data, and the activities by researchers and data managers before, during and after the research is carried out necessary to ensuring that the value of the data can be maximized by its re-use. Guidance on how these costs are predicted accurately has yet to materialize.

Some general guidance on the topic of costing research data management activities is provided by the UK Data Archive and a costing tool is available. Likewise some guidance can be had from the Keeping Research Data Safe cost/benefit analysis. Currently, however, there is little practical experience of research data management costing and – with the exception of the assessment of storage costs – no ready information that may be relied on by which accurate predictions can be made about project-related RDM costs so that appropriate funds may be requested during grant submission. Until such time as personal expertise can be acquired or reliable advice is published, an educated guess necessarily must be made by the manager responsible.

3.1.9 Data Security

In making data management provision it is important that steps are taken to ensure that research data (in the form of **Data Records** both physical and electronic, and physical specimens which represent data) are not lost, and are made accessible only to those who are entitled to see them. Both loss prevention and the maintenance of confidentiality rely on appropriate storage and access provisions being made.

Planning to Prevent Loss

The secure storage of data requires, minimally, that a robust back-up policy be in force for the storage space being used. In general – for academic research – these policies must conform with the legal and governance requirements mandated by government and by funders. Where other less secure storage facilities must be used (for example, when using laptops, external hard-drives, etc.) appropriate back-up procedures must be recorded in the Project DMP and

implemented during the project. It is recommended that file and data storage preferred-practice guidance is provided, which should be conformed with unless special provisions are necessary.

As well as ensuring that digital material is made secure similar measures will be necessary for research data held on physical media, and for research artefacts which contain or constitute data. Data storage guidance should be made available for physical data entities. It may be necessary to make special provision for secure storage to prevent gross loss (e.g. a fire safe, lockable cabinet, etc.) or to prevent damage (specimen mounts, cabinets, etc.) over an appropriate time frame.

The location and the identities of all storage spaces must be identified in general in the Project DMP. This information should be updated as necessary during the course of and on completion of the research project, when it is likely that distributed data associated with a project will be brought together and their locations recorded as a post-project activity in the final version of the Project Record Manifest.

Planning to Maintain Confidentiality

The need to keep data confidential stems from such things as a desire to protect intellectual property or because of commercial or state sensitivity. The Project DMP should identify any areas of sensitivity and make provision for data use where access is constrained or made available for sharing with appropriate limitations.

Where external collaboration will be carried out it is likely that a collaboration agreement will already be mandated to formalize such things as the agreement of duties, responsibilities and IP rights. Where appropriate, consideration should be given – and if necessary, advice taken – to a form of words clarifying the data access, sharing and security requirements agreed by research partners.

Necessarily, to achieve appropriate levels of confidentiality, it must be possible to control access to material, through the use of passwords (for electronic resources) and through some form of ‘gate-keeping’ for physical entities.

Planning for long-term preservation

Long-term preservation implies the use of data archives, for which local data preparation requirements will apply. These should be specified clearly. In general, however, data should be organized and contextualized in an appropriate way by ensuring that the information described in Section 3.2 is provided as comprehensively as possible. In particular advantage should be taken of appropriate metadata standards for describing research data and its organization and location.

Consideration should be given to the file formats used in connection with data. Standard formats and formats with widespread software support are likely to remain understandable for longer than closed, software-specific formats. In cases where the latter formats are unavoidable, ‘preservation copies’ in standard formats should be made and kept alongside them.

For important or valuable data records, it may be worth generating and recording checksums for them using an algorithm such as MD5 or SHA1; information on checksum-generation tools may be found readily using an internet search engine. The algorithm used should be

recorded. This information can be used at a later stage to check the integrity of the Data Record, and may help support staff to locate uncorrupted backup copies.

Where specialist software is required to understand or re-use a Data Record, it is recommended that a copy of the software be kept in a location recorded in the PDMP alongside a copy of the licence terms and a record of known system requirements. Alternatively, the software could be installed on a virtual machine, and the virtual machine stored in a specified location alongside the licence information. Instructions on running the virtual machine should then be provided in the Project DMP (see Section 3.2.9). The practicability and advisability of retaining software will be dependent on the software licence conditions and the sort of resources required, judgements being made on a case-by-case by the manager responsible with advice elicited as appropriate.

3.1.10 Identifying Contractual and legal obligations (RS5, RS6)

There will be legal and, in most cases, contractual obligations with respect to data management which will have to be met as a result of funded or collaborative research. Obligations to funders might include the requirement to develop and submit data management plans, to keep research data for a specified length of time, to submit a research data set to national or discipline-related repositories, and so on. Obligations to collaborators might include ensuring the security and confidentiality of the data that they are providing, constraints on re-use and publication, and end-of-project disposal.

3.1.11 End-of-Life

Efficient use of resources dictates that research data management includes making choices about what data to retain and what might be safely deleted.

Support for making decisions about what data to delete is limited at present. Fundamentally, there are no accepted principles to guide evaluation of the worth of data, and little practical guidance. At the same time predicting whether data, which may seem valueless at present, might be valuable in the future is fraught with difficulty.

Nevertheless, there will be occasions on which it is sensible to delete data; for example, data which are known to have become corrupted, have been found to be collected or created incorrectly, duplicate data, or where security considerations mandate deletion.

Where confidentiality is a concern, data should be deleted using secure methods. Some guidance on reliable deletion – including that for paper documents – is given by the UK Data Archive.

Advice on which research data sets to select for long-term preservation can be found in the DCC's advice on the appraisal and selection of data sets. It should be understood however, that for the usefulness of research data to be maximized, it is not only the data themselves that must be preserved but, together, the contextualizing information which makes the research data findable, accessible, readable and interpretable.

3.2 DMP contents

The items in this section are those which, together with those noted in the preceding section complete the elements that will be found in a properly completed data management plan for a specific research project. This section can be thought of as high-level template for developing a plan.

3.2.1 Summary of Research Activity

A brief summary should be provided of high-level project documentation relating to the Research Activity, and any confidentiality agreements, IPR statements and other documents that affect how the research data may be used.

This summary should include a link to the Project Record Manifest, where full details of documentation are recorded. This summary, therefore, need contain only the most notable facts, such as the project name, project dates, funder(s) and organizations involved.

Access restrictions and ownership of information should be stated here in the simplest, most concise form; where important complexities are glossed, prompts to consult the full documentation should be given.

3.2.2 Existing Data re-use

In the interests of economy, consideration should be given, if possible during formulation of the project proposal, to whether the research activity's data requirements could be met in whole or in part by existing data.

A Research Activity can either mine existing data for new results, add to an existing body of data (to fine tune, generalize or place limits on previous results), or create an entirely new body of data. When planning a new Research Activity, researchers should be able to justify taking one of these three approaches.

If existing data can be used, then such data should be identified together with any potential access problems. If no data are available, briefly indicate how this is known.

Typical reasons for not re-using data include conducting a search and finding that there has been no similar previous research; conducting research on an object that has not previously been studied; operating in an area where all research is covered by strict confidentiality agreements. Typical access issues might include access that is contingent on successful application and unclear data licensing. Typical reasons for generating new data might include performing a comparison over time and extending existing research to cover new areas.

The rationale should be given here for the generation of any new data.

3.2.3 Relating new data to existing data

Firstly, describe how the newly generated data relates to the wider landscape of existing data. This is not concerned so much with existing data that may be used in the Research Activity, but rather with the disciplinary context. A typical answer might identify a body of data with which it would be helpful to harmonize newly generated data, or from which methodologies might be drawn, e.g. ISO standard materials testing data, time/motion studies data.

Second, state the measures that will be/have been taken to ensure integrability between newly generated data and existing data. The following are possible issues to consider. Only brief answers are required here: full details should be given in corresponding sections later in the DMP:

- Method of assuring data quality (Section 3.2.8).
- Method of recording provenance (Section 3.2.5).
- Mechanisms for ensuring trustworthiness of data (Sections 3.2.6, 3.2.8).

- Choice of standard formats, ontologies, conventions, etc. for the data (Section 3.2.9).
- Choice of standard formats, ontologies, conventions, etc. for the metadata (Sections 3.2.9, 3.2.10).

A typical way in which data are re-used is in combination with similar data. This is considerably easier if compatibility issues are addressed in the planning stages of a Research Activity (see Principle of Reusability in The Development of a Set of Principles for the Through-Life Management of Engineering Information).

3.2.4 Future use of the data

If the future uses for research data are known or can be predicted at the outset, special provisions can be made during the research that increase the compatibility of the data with that future use (Principle of Reusability cited above). Explicitly stating where this has been done can help **Data Librarian**s/Managers continue this work in the preservation stage.

To assist future use, firstly list any bodies/groups which might be interested in the data, and the foreseeable contemporary or future uses to which they might put the data.

It is acceptable to define groups based on discipline, research interest or specific research topic. It is acceptable to list bodies or groups without reference to uses, and foreseeable uses without reference to specific groups, if appropriate.

Second, state the measures that will be/have been taken to prepare the data for these bodies/groups/uses.

The following are possible issues to consider. Only brief answers are required here: it is suggested that full details should be given in corresponding sections later in the DMP.

- Forms of data organization (Section 3.2.7).
- Choice of standard formats, ontologies, conventions, etc. for the data (Section 3.2.9).
- Choice of standard formats, ontologies, conventions, etc. for the metadata (Sections 3.2.9, 3.2.10).

3.2.5 The Project Record Manifest

Providing details of what Data Records are included in a Data Case, how they came about and what relationships exist between them helps future researchers to understand the data, assess their suitability and re-use them for new research; it also satisfies Principle 8 from the Principles for Engineering Research Data Management (Darlington, et al, 2010). In particular, recording the relationships between Data Records (and between data) satisfies some users' requirements for provenance information.

The vehicle for making these records is the Project Record Manifest (PRM). The PRM contains information on two classes of data, this being on the one hand the project management and associated documentation, and on the other, the research Data Case, which consists of the set of data records gathered and generated in the course of the Research Activity.

The PRM for the Department may be created in a number of ways. For example, the PRM suggested for use at the UoBDofME is template based. A new PRM should be created for each research project and should be reciprocally associated with the Project DMP.

An approach which augments the PRM is to use some form of data record contextualization of the sort pioneered in the RAID Modelling technique and implemented in the RAIDmap software tool. Using such an approach it is possible to record the development and association of items in the Data Case. Ideally one, or if necessary a number of, **Research Activity Information Development (RAID)** diagrams should be presented for the Data Case, alongside instructions on accessing any computer-interpretable versions; the PRM should cite the adopted modelling method and specifications, alongside notes on how these will be/have been implemented in this case and information about any associated software.

An acceptable alternative – where it has not be possible to use a RAID association method – might be to present an annotated list of Data Records showing associative information. The procedure for maintaining this list must be given in the Project DMP and the record(s) recorded in the PRM.

The PRM could be presented as a wiki page (to allow collaborative development during the project) a web page or as a stand-alone document. If it is the last, it is recommended that the file name and versioning conforms to the local guidelines for project documentation file naming and version control.

3.2.6 Data generation and manipulation

Give a detailed account of how the data will be/have been generated and manipulated, including the methods, technology, conventions, coding schemes, etc. that were used.

It is expected that the level of detail provided here will be low initially, but will increase as the plans are implemented.

At all stages, Principle 4 of the Principles for Engineering Information Management should be borne in mind which states that the notions of re-usability and re-purposing should be supported by the use, where possible, of generic or standard data generation and manipulation tools. This approach is reinforced further in The Principle of Robustness set out in the Principles for Engineering Information Management which exhorts all information workers to use robust methods to capture, create and manipulate information entities as a means of supporting future use.

When writing a DMP in retrospect, it is acceptable to cite a journal/conference paper containing the information, provided it is detailed enough and that a pre- or post-print is available in case of access difficulties. In the normal course of events, the information should be provided here first and then adapted for use in a journal/conference paper. It may be helpful to provide this information in the form of a commentary on a RAID diagram.

3.2.7 Data organization

Describe how the data will be/have been organized. This refers equally to how data are organized within Data Records, how Data Records are organized within the Data Case, and how project management records are organized. Providing this information makes it easier for others involved in the later management of research data (for example, by a project or **Data Manager**) or re-use of the data to navigate the Data Records and find specific parts. In addition, such information will aid in the efficient response to FOI requests. It can also help you to check that all the Data Records have been included.

The basic method recommended for organization of project documents (both within the Data Case and project management information) is by using a file naming convention of the sort referred to above.

In addition to allowing unique (human-readable) record names to be created, the file naming convention also provides the means for version control. A useful reference document both for version control (in general) and for record organization is the Versions Toolkit developed by LSE; its use is recommended.

It is strongly encouraged for management of Departmental research data and associated records, that full use is made of the metadata recording facilities provided by the electronic records-handling methods that are used. An example of this is the 'properties box' provided for individual files in all Microsoft Office applications in which can be recorded such things as document title, author(s), and so on.

Where possible, additional metadata description of the research data or **Research Data Record**s should be provided. It is recommended at present that the Data Cite metadata schema is adopted for this (Starr et al., 2011).

All research data and other data and document records associated with a research activity can be assembled into a Data Case at completion of the research activity for ease of long-term curation and management. The Data Case should be packaged in a format that will promote its long-term potential for re-use. In particular, where special arrangements for archiving are mandated (e.g. by the funder) the packaging should follow the guidance provided which will ensure it is acceptable to the Data Librarian(s) taking custody of it. In the absence of such mandates, the BagIt packaging format is recommended: at their simplest such packages consist of a directory containing a file identifying the BagIt version, another listing the files in the package along with checksums, and a data directory containing the files and folders themselves. It is the responsibility of the PI that this is done and that all data is collected from independent file storage (independent and removable media) and from the personal (H-drive) storage space of any researcher involved in the project. At this juncture the data case records should be associated by bringing the Project Record Manifest up to date.

3.2.8 Data quality

Using high quality data in research is important to its outcome; equally, for their confident re-use by others data not only must be of high quality but must also be demonstrated to be so.

The quality assurance procedures and standards can be recorded here that will be/have been used for collection, generation and manipulation of the research data. If any data quality issues were encountered, list them and describe what was done to resolve them.

3.2.9 Data structures and formats

At the project planning stage, the hardware and software environment in which the Research Activity will be conducted should be specified to the extent that it can be known. Where possible use off-the-shelf software and software that is supported by BUCS or a more local IT support service.

Indicate the formats to be used, and explain why these have been selected for use. If the choice of formats has been justified elsewhere in the DMP (e.g. Sections 3.2.2 and 3.2.3) readers may be directed to those sections in place of a recapitulation here.

Also indicate at this stage how this section will be completed during the course of the Research Activity.

Once Data Records have been made, start by specifying the hardware and software environment in which the data were generated or manipulated, and then consider alternative environments, tools and libraries that might support the data. If specialist tools were used, consider installing them on a virtual machine; in which case, provide details here of how to run the virtual machine.

Specify the information, tools or resources that would be needed to manipulate or make your Data Records human readable. If available/known, cite here format specification documents for all data formats used.

3.2.10 Data semantics

Data cannot be re-used if their meaning is not properly understood.

At the project planning stage, provide if possible a general statement about the conventions that will be used to allow interpretation of data (such as schemas, ontologies, and so on), and provide justification for the conventions used. Indicate how this section will be completed during the course of the Research Activity.

Once Data Records have been made, provide any additional information that would be needed by an interested reader to understand the Data Records.

As an example, tabular data can have terse column headings; fuller explanations of what a column represents can be given here. Other examples of information to provide here include data dictionaries, coding schemes and ontologies. The information can be given directly in the DMP, or instructions can be given on how to look up the information for each Data Record.

4. REFERENCES

Mansur Darlington & Alex Ball. (2012a). *A Research Data Management Plan for the Department of Mechanical Engineering, University of Bath* (version 1.2). REDm-MED Project Document redm1rep120207mjd12.Bath, UK: University of Bath. http://opus.bath.ac.uk/30099/

Mansur Darlington & Alex Ball. (2012b). *Research Data Management Plan Requirements Specification for the Department of Mechanical Engineering, University of Bath* (version 1.0). REDm-MED Project Document redm1rep120105mjd10. Bath, UK: University of Bath. http://opus.bath.ac.uk/28040/

Darlington, M., Ball, A., Howard, T, McMahon, C. & Culley, S. (2010). *Principles for Engineering Research Data Management*. (ERIM Project Document, erim6rep101028mjd. Bath, UK: University of Bath. Accessed 8 November 2010 from http://wiki.bath.ac.uk/display/ERIM/Principles+for+Engineering+Research+Data+Management.

Digital Curation Centre. (n.d.). *DCC Charter and Statement of Principles.* Retrieved October 27, 2010 from http://www.dcc.ac.uk/about-us/dcc-charter

Howard, T., Darlington, M., Ball, A., McMahon, C., & Culley, S. (2010). *Understanding and characterizing engineering research data for its better management* (ERIM Project Document rep100420mjd10). Bath, UK: University of Bath. http://opus.bath.ac.uk/20896

McMahon, C., Caldwell, N., Darlington, M., Culley, S., Giess, M., & Clarkson, J. (2009). *The Development of a Set of Principles for the Through-Life Management of Engineering Information* (KIM Project Document kim40rep007mjd10). Bath, UK: University of Bath. http://opus.bath.ac.uk/16132/

Starr, J., Ashton, J., Brase, J., Bracke, P., Gastl, A., … Ziedorn, F. (2011). *DataCite metadata schema for the publication and citation of research data* (version 2.2). Hannover: TIB. doi:10.5438/0005

Swan, A. & Brown, S. (2008). *The skills, role and career structure of data scientists and curators: Assessment of current practice and future needs.* London: JISC. Retrieved September 21, 2010 from http://www.jisc.ac.uk/publications/reports/2008/dataskillscareersfinalreport.aspx

APPENDIX A: DATA MANAGEMENT TERMINOLOGY

This document uses the ERIM Data Management Terminology (presented in draft form Howard, et al. 2010) from which the following terms are extracted for ease of reference. The full released terminology can be accessed on-line.

Research Activity The process through which research Data and context Data are accumulated and developed.

Data Case The set of Data Records associated with some discrete Research Activity (project, task, experiment, etc.).

Record Information in any medium, created, received and maintained as evidence of an activity.

Data Record The Data Object which contains the Data.

Research Data Record A Record containing research Data, i.e. Data that are descriptive of the research object.

Research Object Data Record A Data Object which is itself the object of research interest or which together with Research Object Data Records constitutes the object of research interest.

Data Librarian People originating from the library community, trained and specialising in the curation, preservation and archiving of data. (Swan & Brown, 2008)

Data Manager Computer scientists, information technologists or information scientists and who take responsibility for computing facilities, storage, continuing access and preservation of data

Project Data Management Plan The data management plan which contains details of management of the management documentation and the Data Records (the Data Case) of a discreet I*d*MRC Project or activity.

Project Record Manifest The inventory which lists both the set of project management documents and the set of Data Records which constitute the Data Case(s), and which provides associational, locational, ownership and other metadata about these physical and digital data objects.

Research Activity Information Development (RAID) Diagram A diagram using the RAID modelling formalism which records and visualizes the research and associated data record development in a Data Case resulting from the Research Activity.