UNIVERSITY OF
BATH

Publication date:
2010

Link to publication

University of Bath

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

THE DRAFT IdMRC PROJECTS DATA MANAGEMENT PLAN

MANSUR DARLINGTON, ALEX BALL, TOM HOWARD, STEVE CULLEY AND CHRIS MCMAHON

erim6rep101015mjd10.pdf

ISSUE DATE: 23 December 2010

Catalogue Entry

Title	The Draft IdMRC Projects Data Management Plan
Creator	Mansur Darlington, Alex Ball, Tom Howard (authors)
Contributor	Steve Culley, Chris McMahon
Subject	information management; data management; research data
Description	The need for better management of research data is increasingly becoming understood not only by those who collect and generate these data, but also by those who have a research funding or governance rôle. As a result it is now becoming more usual for research funders to expect evidence of formal data management planning in the initial bid for funds, with detailed plans being required early in the management of a funded project. The IdMRC Projects Data Management Plan (DMP) is both a specification for the DMPs to be written for IdMRC projects, and a summary of the support structures needed to implement them. It is presented in draft form as data management support within the centre, and the University of Bath generally, is still evolving; for this reason it contains elements of both prescription and guidance.
Publisher	University of Bath
Date	15th October 2010 (creation)
Version	1.0
Type	Text
Format	Portable Document Format (PDF/A-1b:2005)
Resource Identifier	erim6rep101015mjd10
Language	English
Rights	© 2010 University of Bath

Citation Guidelines

Mansur Darlington, Alex Ball, Tom Howard, Steve Culley and Chris McMahon. (2010). *The Draft IdMRC Projects Data Management Plan* (version 1.0). ERIM Project Document erim6rep101015mjd10. Bath, UK: University of Bath.

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 maximizing its sharing and re-use. Better data management aimed at promoting enhanced availability of and easier access to existing data 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.

The interest in increasing the opportunity for re-using existing research data that has grown in recent years can be seen in research and other activities supported by such organization as JISC, DCC and UKOLN, encouraged and supported in turn by the research funding councils. At the same time many funding bodies are expressly expecting that those they fund put in place data management plans and there is an expectation that data be shared (Jones, 2009). This phenomenon is not limited to the UK. For example, the NSF in the US now expect evidence to be shown of formal data management planning in all future research projects. Thus a general trend toward better research data management can be identified.

The following paper presents the draft project data management plan for projects within the Innovative Design & Manufacturing Research Centre (IdMRC) at the University of Bath. This research centre is typical of many in that it undertakes a spectrum of research projects all of which, of course, generate research data, but for which the practice of formal management of data is in its infancy. In order for 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 centre-level data management plan from which a specific management plan can be drafted as appropriate to the character of each individual project and it research data. The data management plan presented here is an early draft of just such a centre-level data management plan. Thus this centre-level plan (when adopted) can be seen as a specification for each project-level plan.

This draft is also a deliverable for the Engineering Research Information Management Project (ERIM) funded by JISC as part of a wider exploration into managing research data resulting in the provision of methods, procedures and tools in support of such management. The document remains in draft form at present because much of the detail cannot be confirmed until decisions have been made about, for example, the precise data management support facilities that will be provided by the IdMRC for their projects and the exact form of procedures that are yet to be put in place. It is for this reason, too, that this document contains elements of both prescription and guidance.

This plan is based on the requirement specification for engineering research data management plans in general developed during the ERIM Project and presented in Ball, et al (2010). That specification was informed by three other key documents, these being the template for DMPs drawn up by the Digital Curation Centre (DCC) (Donnelly & Jones, 2010), the Principles for Engineering Data Management (Darlington, et al., 2010) and a thematic analysis of DMP tools and exemplars (Ball, 2010).

By way of clarification, Figure 1 shows the relationship between these documents which will culminate in an instance of a project-level data management plan.
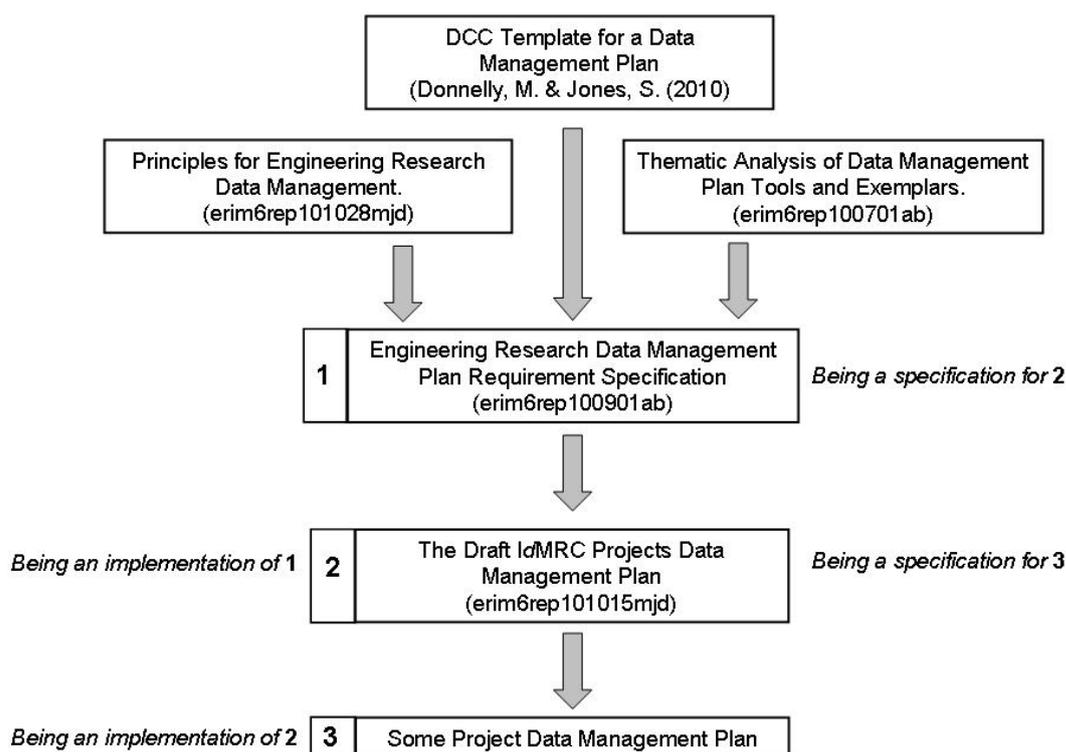

Figure 1. A progression of data management planning instruments

The purpose of this data management plan (DMP) is to provide support to three types of data preparation activity, identified and defined by the authors:

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*).

The DMPs produced under this specification have two functions. In the first instance, they 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, their 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. DMPs under this specification are not intended for use in the preservation stage, though they (along with other documentation) should provide sufficient information to allow **Data Creator**s/**Data Manager**s to construct a suitable DMP for the long-term care of the data.

The DMP presented here makes provision to:

- identify any future known Research Activities that may make use of the research data;
- describe how researchers will make, are making or have made the **Data Case** available and fit for these Research Activities, if applicable;
- describe how Data Creators will manage, are managing or have managed the Data Case to make it amenable for use in a future unknown Research Activity;
- provide additional information, where needed, to allow a
- provide additional information, where needed, to allow a **Data Librarian**/Manager to continue to manage the Data Case, enabling its use in the identified known Research Activities and in future unknown Research Activities generally.

In addition to this, the IdMRC DMP is itself interpretable as the specification to be used for development of the individual data management plan required for each research project carried out in the IdMRC.

This plan should not be seen as a specification for the management of IdMRC management documentation of a more general nature. However, the plan could be generalized, if thought appropriate, to fulfil that rôle too.

It should be understood that the goal of data management for supporting the three data preparation activities is narrower in scope than a more general goal which would include management, for example, for data archiving, preservation and curation. Whilst many of the strategies implemented within the proposed data management plan may serve to support these it is not the express intention that they should do so. Thus actions which are specifically intended for the support of archiving, preservation and curation per se are not within the scope of this data management plan.

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.

Swan and Brown (2008) define four main roles responsible for the management of research data, including those of Data Manager and Data Creator. In the IdMRC, there is currently no data manager per se; this rôle as well as that of data creator is carried out in general by the researcher, to the extent that any formal data management is carried out at all. Thus, the data management plan presented here for implementation in the IdMRC is essentially *researcher-centric*. There are, however, data management activities that are outwith the proper responsibility of the researcher which are referred

to and identified as being the concern of a data manager. It implies therefore that the rôle of a data manager proper be identified and introduced within the IdMRC. The rôle of data manager could be carried out equally well at the centre, department or faculty level. However, whatever the purview it is quite clear that for proper data management to be carried out sufficient resources will have to be found and allocated. It may be that requests for data management funding, including that to support a data manager, at the project level should become anticipated by the funding body in every project bid and where block funding is applied for the rôle of data manager be explicitly identified.

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 on-line,[1] and those used within this DMP can be found in a glossary, in Appendix A to which they are hyperlinked at first use here.

1.1 The IdMRC

The IdMRC is a research centre within the Department of Mechanical Engineering of the University of Bath. It has a research body of about sixty people engaged in a wide variety of research activities covering a spectrum of mechanical-engineering-related topics including constraint-based design, advanced machining systems, large-volume metrology and information and knowledge for engineering design and manufacture. It 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 it data management plan will be generalizable directly to other research centres. It may be said, however, that some of the implementational recommendations will be unique to the IdMRC given its local research support infrastructure and research activity organization and culture.

2. THE IdMRC PROJECTS DATA MANAGEMENT PLAN

The **IdMRC Projects Data Management Plan** which follows is composed of two parts. The first part – Section 2.1 Infrastructure and Implementational Issues – concerns matters which must be in place in order 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 2.2 DMP Contents – provides details of the plan itself which will provide guidance and prescription for the implementation at the time for each Research Activity and associated Data Case. Thus, it is assumed that for each research project in the IdMRC a data management plan will be produced which will be an implementation of this generic DMP, 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 management of both research data and management and other information associated with the project.

The IdMRC Projects DMP should be used in its entirety when developing a DMP for a specific Research Activity (project). Such a plan is referred to henceforth in this document as the **Project Data Management Plan** (Project DMP). Provision of a draft

---

[1] ERIM Data Management Terminology, URL: https://wiki.bath.ac.uk/display/ERIMterminology

project DMP, which fulfils the requirements outline will be a requirement of bids for funding.

2.1 Infrastructure and implementation issues

Delivering and executing a Project DMP is dependent on certain supporting mechanisms either being in place as part of the IdMRC'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.

2.1.1 Relating the DMP to other documentation

The IdMRC Projects Data Management Plan, and any Project DMPs that are an implementation of it, is informed by a number of key documents, each of which is available for inspection:

The University of Bath Good Practice Guide for Research

RCUK Policy and Code of Conduct on the Governance of Good Research Conduct

Documentation specific to a research project or research activity should be made available from a single unrestricted-access on-line source as indentified in that research activity's data management plan. For all IdMRC research projects the on-line source will be a University of Bath wiki page that has been created for this purpose. It is recommended that this page is one contained within the research project wiki space (see below for wiki access details). Records may be stored within the wiki space, where appropriate. Redirection may be provided here also 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 BUCS file space or within a wiki or other web space.

The sorts of information that will be found here will include high-level project documentation relating to the research activity – such as the project proposal document, and the detailed project plan(s) – the Project DMP itself, the Project Record Manifest, confidentiality agreements, IPR statements and other documents that affect how the research data may be used. Care should be taken to ensure that sensitive and confidential information is protected in an appropriate manner, and where necessary redacted versions of documents are used.

Uniquely important amongst these documents will be the **Project Record Manifest**. This template-based document will list all key project documents and other records and their associations, locations, authors and access particulars.

In addition to this the locations should 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.

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 Bath Wiki Space

The principal means of making persistent links to IdMRC project documentation is through the use of Bath Wiki pages. For each project a project wiki space should be created by the principal researcher or designated project manager. This space can be used for activities during the course of the project 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 wiki 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 default when a new 'space' is created the parent node is designated as 'Home' and all other pages are then children of the home page. As children they inherit any access characteristics of the home page, so if the home page is restricted to a user base then so too will all other pages. It is possible to make unrestricted access to the home page in which case, as a default, all child pages will be accessible unless local access restrictions are placed on individual pages. It is a matter of choice how access is arranged for a project wiki space. However, for each project there must always be one page entitled 'Project Management Documentation' that is accessible universally to all.

Assuming that a project space is set up the permissions for which allow access only to members of the research group, it will be necessary to make a 'Project Management Documentation' page which has unlimited access. This will be the default page in which will be found information leading to project information of all sorts.

To achieve universal access it is necessary to create a page which is higher in the hierarchy than the home page, and then to set *anonymous log-in access* to that page. Do this as follows:

1) Go to the home page and create a child page with 'Project Management Documentation' as the title.

2) From the page view options at the base of the home page, select 'View in Hierarchy' or 'Tree'. The newly created page can then be dragged and dropped to a position above that of the home page.

3) Open the Project Management Documentation page and go to 'Browse -> Space Admin'.

4) On the Space Administration page select 'Permissions' from the left-hand column menu.

5) Under the 'Anonymous Access' panel select 'edit permissions' and tick the box for 'View'.

It will now be possible, given the URL, for any anonymous user of the Bath Wiki Space to have view-only access to this page.

The Project Management Documentation page can now be used as a conduit to further information. Any project documentation that is added as an attachment to this page will, as a child, be access free. Links can be made also, from this 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 elsewhere.

2.1.2 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, the default location for project documentation, including the Project DMP, will be in the BUCS-supported research project wiki space created for this purpose as part of the project data management process as described in the preceding section.

2.1.3 Understanding the DMP

A system must be in place to allow readers of the Project DMP to find and read its precursor documents. To this end, the URL of this document, the, IdMRC DMP (erim4rep101015mjd), and the higher-level specification 'Engineering Research Data Management Plan Requirement Specification' (erim6rep100901ab) will be embedded in the Project DMP.

The URL for the IdMRC Projects DMP (i.e. this document) is: http://opus.bath.ac.uk/22200.

The URL for the ERDMPRS is: http://opus.bath.ac.uk/21280

2.1.4 Rôles and responsibilities

The responsibilities for writing, implementing and reviewing the Project DMP are shared between a number of rôles within the IdMRC, as dictated ultimately by The University of Bath Good Practice Guide for Research:

'Research data management and archiving is the responsibility of the principal investigator (PI) for each project and must be considered from the outset of new projects. … The management and archiving of research data generated by postgraduate research students is the responsibility of the lead supervisor, reporting to the project PI where applicable.'

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 the project submission document. It noted earlier is likely 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, with assistance from the Centre data 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.

2.1.5 Review of the DMP

The Project DMP acts as guidance and as a record of activity. Conforming to the Project DMP will promote good data management practice, while providing an accurate record which will assist both Data Librarians/Managers (when reworking the DMP for long-term archiving and preservation) and future Data Creators (enabling more realistic DMPs to be drafted for future projects).

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.

2.1.6 Revision of the DMP

The DMP is not a static document, but should be updated throughout the project to reflect what actually happened. Versioning should conform to the IdMRC guidelines for project documentation file naming and version control.. The Project DMP will uniquely have the document type 'dmp' as part of the file name.

Change dates and authorship should be recorded within the Project DMP, together with author contact details.

2.1.7 Budget

It is the expectation of funding bodies that funds commensurate with the details of the data management requirements for the project in question are identified in the project proposal. Therefore 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.

2.1.8 Storage, back-up and security

It is important that steps are taken to ensure that research data (in the form of research **Data Record**s, 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. During the course of a project it is likely that data will be stored, temporarily, in a number of places. Measures should be put in place to secure data, according to the inherent security of the storage space.

Full details of the facilities and methods that are available and that should are to be used will be found in the IdMRC guidance documentation.

The location and the identities of all storage spaces must be identified in general in the Project DMP. This 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.

2.2 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.

2.2.1 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.

2.2.2 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 is 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; unclear data licensing. Typical reasons for generating new data might include performing a comparison over time; extending existing research to cover new areas.

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

2.2.3 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 2.2.8).
- Method of recording provenance (Section 2.2.5).
- Mechanisms for ensuring trustworthiness of data (Sections 2.2.6, 2.2.8).
- Choice of standard formats, ontologies, conventions, etc. for the data (Section 2.2.9).
- Choice of standard formats, ontologies, conventions, etc. for the metadata (Sections 2.2.9, 2.2.10).

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).

2.2.4 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 Librarians/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: full details should be given in corresponding sections later in the DMP.

- Forms of data organization (Section 2.2.7).

- Choice of standard formats, ontologies, conventions, etc. for the data (Section 2.2.9).
- Choice of standard formats, ontologies, conventions, etc. for the metadata (Sections 2.2.9, 2.2.10).

2.2.5 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 IdMRC is template based, and is an instance of the 'man' type of document. A new PRM should be created for each research project and should be reciprocally associated with the Project DMP.

For recording 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 RAID modelling method and specifications, alongside notes on how these will be/have been implemented in this case.

An acceptable alternative – where it has not be possible to use a RAID association method, would 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.

Versioning of the PRM should conform to the IdMRC guidelines for project documentation file naming and version control.. The PRM will uniquely have the document type 'man' as part of the file name.

2.2.6 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.

2.2.7 Data organization

Describe how the data will be/have been organized. This refers both 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 a Data Librarian/Manager or re-user of the data to navigate the Data Records and find specific parts. It can also help you and your data manager to check that all the Data Records have been included.

The basic method for organization of project documents (both within the Data Case and project management information) is by using the IdMRC file naming convention 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.

It is a mandatory requirement for management of IdMRC research data and associated records, that full use is made of the meta data recording facilities provided by the electronic records-handling 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.

All research data and other data and document records associated with a research activity should 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 acceptable to the Data Librarian(s) taking custody of it. 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.

2.2.8 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 should 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.

2.2.9 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.

Specify the information, tools or resources that would be needed to manipulate or make your data human records human readable. Provide an explanation of why a particular format has been selected for use. Where possible use off-the-shelf software and software that is supported by BUCS or a more local IT support service.

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. If available/known, cite here format specification documents for all data formats used.

If the choice of formats has been justified elsewhere in the DMP (e.g. Sections 2.2.2 and 2.2.3) readers may be directed to those sections in place of a recapitulation here.

2.2.10 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.

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.

3. REFERENCES

Ball, A. (2010). *Thematic analysis of Data Management Plan tools and exemplars* (ERIM Project Document erim6rep100701ab10). Bath, UK: University of Bath.

Ball, A., Darlington, M., Howard, T., McMahon, C. & Culley, S. (2010). *Engineering Research Data Management Plan Requirement Specification* (ERIM Project Document erim6rep100901ab10). Bath, UK: University of Bath.

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

Donnelly, M. & Jones, S. (2010). *Template for a Data Management Plan.* Digital Curation Centre. Retrieved July 12, 2010 from http://www.dcc.ac.uk/sites/default/files/DMP_template_v1.2_100106.rtf

Freiman, L., Ward, C., Jones, S., Molloy, L., & Snow, K. (2010). *Scoping study and implementation plan: 'A pilot project for supporting research data management'* (Incremental Project Document). Cambridge, UK and Glasgow, UK: University of Cambridge and University of Glasgow. Retrieved October 27, 2010 http://www.lib.cam.ac.uk/preservation/incremental/documents/Incremental_Scoping_Report_17091 0.pdf

Howard, T., Darlington, M., Ball, A., McMahon, C., & Culley, S. (2010). *Understanding and characterizing engineering research data for its better management* (ERIM Project Document erim2rep100420mjd10). Bath, UK: University of Bath. Retrieved September 21, 2010 from http://opus.bath.ac.uk/20896/

Jones, S., A report on the range of policies required for and related to digital curation, version

1.2, (DCC, Glasgow, March 2009)

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. Retrieved October 27, 2010 from http://opus.bath.ac.uk/16132/

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

Yale Law School Roundtable on Data and Code Sharing. (2010). Reproducible research: Addressing the need for data and code sharing in computational science. *Computing in Science and Engineering, 12*(5), 8–12. DOI: 10.1109/MCSE.2010.113. Retrieved October 27, 2010 from http://www.stanford.edu/~vcs/papers/RoundtableDeclaration2010.pdf

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.[2]

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.

Associative Data Record A Record which documents the association between other Data or Data Records. The Data contained within an Associative Data Record is a special case of contextualizing Data and the Data Record a special case of a Context Data Record.

Context Data Record A Record containing Data explicitly intended to place in context other Data or abstract aspects of the Research Activity or subject.

Experimental Apparatus Data Record A Digital Object which is analogous to the physical experimental apparatus familiar in much laboratory-based research.

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 Object Either a Physical Object or a Digital Object. (Swan & Brown, 2008)

Digital Object An object composed of a set of bit sequences. (Swan & Brown, 2008)

Data Creator Researchers with domain expertise who produce data. These people may have a high level of expertise in handling, manipulating and using data. (Swan & Brown, 2008)

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

---

[2] ERIM Data Management Terminology, URL: https://wiki.bath.ac.uk/display/ERIMterminology

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

Data Scientist People who work where the research is carried out – or, in the case of data centre personnel, in close collaboration with the creators of the data – and may be involved in creative enquiry and analysis, enabling others to work with digital data, and developments in data base technology

IdMRC Projects Data Management Plan The high-level data management plan which serves as the specification for instances of each IdMRC Project Data Management Plan.

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 IdMRC 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.